



Enfermería Clínica

www.elsevier.es/enfermeriaclinica



ARTÍCULO ESPECIAL

Aplicación del análisis Rasch en la investigación enfermera: una introducción metodológica

José Verdú-Soriano^{a,*} y Héctor González-de la Torre^b

^a Departamento de Enfermería Comunitaria, Medicina Preventiva y Salud Pública e Historia de la Ciencia, Facultad de Ciencias de la Salud, Universidad de Alicante, Alicante, España

^b Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Las Palmas, España

Recibido el 7 de marzo de 2024; aceptado el 23 de octubre de 2024

PALABRAS CLAVE

Análisis Rasch;
Investigación
en enfermería;
Investigación
en evaluación
de enfermería;
Estudios
de validación;
Teoría respuesta
al ítem

Resumen El diseño y la validación de herramientas constituye un pilar fundamental de la investigación enfermera, donde es imprescindible incorporar las metodologías y los análisis más actuales para poder garantizar la validez y la fiabilidad en su aplicación clínica.

El objetivo de este trabajo es describir las características del análisis Rasch y la metodología para su desarrollo, así como discutir su relevancia y aplicabilidad en la investigación enfermera, destacando su potencial contribución a la mejora de la calidad y precisión de los instrumentos de medición en la disciplina.

A través de una síntesis narrativa se describen los fundamentos teóricos del análisis Rasch, y se presentan las características y los supuestos que se deben cumplir para llevar a cabo este tipo de análisis. Posteriormente, se presenta la metodología de implementación en 11 pasos: definición de objetivos y preparación, diseño de instrumento, recolección de datos, análisis inicial y de la unidimensionalidad, evaluación de la bondad de ajuste (*infit* y *outfit*, fiabilidad y separación), evaluación de la independencia local de los ítems (coeficiente Q3 de Yen), calibración de ítems y estimación de las habilidades, análisis de la invarianza de la medida (análisis DIF), revisión y modificación del instrumento, análisis final y validación, interpretación de los resultados. Se presentan ejemplos de uso, así como las ventajas y las limitaciones del método. Como conclusión, el análisis Rasch proporciona una metodología valiosa para la evaluación de competencias y las habilidades clínicas, así como para el desarrollo y la validación de instrumentos de medición de gran utilidad para la investigación en cuidados, aunque sería necesario potenciar la formación y estandarización en su uso.

© 2024 Los Autores. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

* Autor para correspondencia.

Correo electrónico: pepe.verdu@ua.es (J. Verdú-Soriano).

<https://doi.org/10.1016/j.enfcli.2024.10.005>

1130-8621/© 2024 Los Autores. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

Cómo citar este artículo: J. Verdú-Soriano and H. González-de la Torre, Aplicación del análisis Rasch en la investigación enfermera: una introducción metodológica, *Enfermería Clínica*, <https://doi.org/10.1016/j.enfcli.2024.10.005>

KEYWORDS

Rasch analysis;
Nursing research;
Nursing evaluation
research;
Validation studies as
topic;
Item response theory

Rasch analysis implementation in nursing research: A methodological approach

Abstract The design and validation of tools constitutes a fundamental pillar of nursing research, where it is essential to incorporate the most current methodologies and analyses in order to guarantee validity and reliability in their clinical application.

The aim of this paper is to describe the characteristics of the Rasch analysis and the methodology for its development as well as to discuss its relevance and applicability in nursing research, highlighting its potential contribution to the improvement of the quality and accuracy of measurement instruments in the discipline.

Through a narrative synthesis, the theoretical foundations of Rasch analysis are described and the characteristics and assumptions that must be fulfilled to carry out this type of analysis are presented. Subsequently, the implementation methodology is presented in 11 steps: definition of objectives and preparation, instrument design, data collection, initial and unidimensionality analysis, goodness-of-fit assessment (infit-outfit, reliability and separation), item local independence assessment (Yen's Q3 coefficient), item calibration and estimation of skills, measurement invariance analysis (DIF analysis), review and modification of the instrument, final analysis and validation, interpretation of results. Examples of use are presented, as well as the advantages and limitations of the method. In conclusion, the Rasch analysis provides a valuable methodology for the evaluation of clinical competencies and skills, as well as for the development and validation of measurement instruments of great utility for research in care, although it would be necessary to promote training and standardization in its use.

© 2024 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La enfermería, como disciplina centrada en el cuidado de la salud, requiere de herramientas de evaluación precisas para garantizar la calidad y la eficacia de los cuidados proporcionados. Así pues, juega un papel crucial en el desarrollo de instrumentos de medida, especialmente en áreas relacionadas con la salud, el bienestar, la calidad de vida y la evaluación de cuidados. Las enfermeras contribuyen de forma activa junto a otras disciplinas de la salud y las ciencias sociales en la consecución de avances en la creación y validación de estos instrumentos.

El análisis Rasch, dentro del ámbito de la investigación en enfermería, se destaca como una herramienta estadística poderosa y refinada para la evaluación y la mejora de instrumentos de medición, especialmente en lo que respecta a encuestas y evaluaciones de competencias. Originario de la teoría de respuesta al ítem (TRI)¹, el análisis Rasch permite a los investigadores superar limitaciones inherentes a los métodos de escala tradicionales, proporcionando un marco más sólido y detallado para la interpretación de datos en estudios de salud. Ofrece una metodología sólida para la construcción y la validación de instrumentos de medición, permitiendo a los investigadores y profesionales clínicos desarrollar escalas y pruebas que reflejan con precisión las competencias y habilidades clínicas².

Así pues, en el contexto de la investigación enfermera, el análisis Rasch se utiliza para validar y refinar instrumentos de medición, asegurando que estos sean tanto fiables como válidos para medir los fenómenos específicos dentro de la disciplina³. Esto es de suma importancia, ya que los instrumentos de alta calidad son esenciales para la recopilación de datos precisos y útiles, los cuales, a su vez, informan de las prácticas basadas en evidencia y políticas de salud pública⁴.

En este contexto, los objetivos de este artículo son describir el análisis Rasch y discutir su relevancia y aplicabilidad en la investigación enfermera, destacando cómo esta metodología puede contribuir significativamente a la mejora de la calidad y precisión de los instrumentos de medición. Al hacerlo, se busca iluminar la importancia del análisis Rasch, no solo como una herramienta estadística, sino también como un enfoque que puede facilitar la investigación significativa y relevante en enfermería, la cual es crítica para el avance de la práctica y la formulación de políticas basadas en evidencia.

La comprensión y aplicación del análisis Rasch ofrece varias ventajas. No solo mejora la capacidad de desarrollar y validar instrumentos de medición robustos, sino que también proporciona una base sólida para la interpretación de los resultados de la investigación, asegurando que las decisiones basadas en estos datos sean lo más informadas y precisas posible. Además, el énfasis del análisis Rasch en la calidad de los datos refuerza la importancia de la rigurosidad metodológica en la investigación en enfermería, contribuyendo así al cuerpo general de conocimiento en la disciplina.

Fundamentos teóricos del análisis Rasch

El análisis Rasch, nombrado así por el matemático danés Georg Rasch⁵, es un enfoque dentro de la TRI que proporciona una base sólida para la medición objetiva en las ciencias sociales y de la salud. A diferencia de otros modelos estadísticos que pueden basarse en suposiciones menos estrictas sobre los datos, el modelo Rasch se caracteriza por su capacidad para transformar respuestas cualitativas (p. ej., correcto/incorrecto, de acuerdo/en desacuerdo) en medidas cuantitativas a lo largo de un continuo de habilidad o actitud, de modo que valores bajos en este continuo de la

variable latente representan características de la variable latente que ocurren frecuentemente o que son fácilmente observables. Por otro lado, valores altos en este continuo indican características que son raramente observadas o bien que son más difíciles de conseguir⁶.

Descripción del modelo Rasch

El modelo Rasch es, fundamentalmente, un modelo matemático, con función de probabilidad logística, que asume que la probabilidad de una respuesta correcta a un ítem (o la elección de una respuesta particular en una escala) depende, únicamente, de la diferencia entre el nivel de habilidad del individuo y la dificultad del ítem⁷. Esta característica singular permite que el modelo Rasch proporcione mediciones independientes de las personas y los ítems, una propiedad conocida como «invarianza de la medida»⁸, lo que significa que las mediciones no dependen del grupo específico de ítems presentados o de la población de respondientes en la muestra.

Otro componente crucial de este modelo es la «jerarquía o calibración de ítems»⁹, un proceso mediante el cual se estima la dificultad de cada ítem dentro de un conjunto de datos. Es decir, se ordenan a los ítems y a las personas en el mismo continuo, en el que los ítems más probables se sitúan en la parte inferior del continuo¹⁰. La jerarquía de los ítems a lo largo del continuo determina el orden de las ubicaciones de los ítems en relación con la distribución de las personas⁹. Esta calibración es fundamental para asegurar que los ítems sean adecuadamente comparables, y para facilitar la construcción de pruebas equitativas y justas. Por último, la «estimación de habilidades» de los individuos es igualmente importante, permitiendo la evaluación precisa del nivel de habilidad o competencia de una persona en relación con el constructo medido.

También hay que tener en cuenta que el modelo Rasch se basa en 2 suposiciones clave que se deben cumplir^{6,11}. Si no es posible asumir estas suposiciones, se deberían aplicar otros métodos de validación de instrumentos. Estas suposiciones son:

- **Asunción de unidimensionalidad:** Es la asunción básica del análisis Rasch¹². Este concepto implica que todas las preguntas o ítems en la prueba están diseñados para medir una única dimensión subyacente o constructo. En otras palabras, un conjunto de ítems se considera unidimensional si todas las variaciones en las respuestas de los individuos pueden atribuirse a diferencias en el nivel de esa única dimensión que se está evaluando, y no a otras dimensiones o factores irrelevantes.
Por ejemplo, si estamos evaluando la habilidad de comprensión lectora en una prueba, queremos que todos los ítems midan esta habilidad y no estén influenciados por otras habilidades como la memoria o el conocimiento enciclopédico. Si la prueba es unidimensional, podemos confiar en que las diferencias en las respuestas de los individuos reflejan verdaderamente diferencias en su comprensión lectora y no están distorsionadas por otros factores.
- **Independencia local de los ítems**^{6,11}: Esta propiedad implica que las respuestas de los individuos a cada ítem

son independientes entre sí, una vez que se ha tenido en cuenta su habilidad en la dimensión que se está evaluando. En otras palabras, la independencia local significa que la probabilidad de que un individuo responda correctamente a un ítem determinado no está influenciada por sus respuestas a otros ítems, una vez que se ha considerado su nivel de habilidad en la dimensión que se está evaluando.

Esta propiedad es fundamental en el modelo de Rasch porque permite que cada ítem contribuya de manera única a la medición de la habilidad del individuo en la dimensión específica. Si no se cumple esta independencia local, puede haber sesgos en la interpretación de las puntuaciones de la prueba y se podría afectar la validez y la fiabilidad de la medida.

A diferencia de otros modelos de la TRI, el análisis Rasch no requiere que los datos se ajusten a una distribución específica, lo que lo hace particularmente versátil y robusto para diferentes tipos de evaluaciones¹. Esta característica subraya la importancia del modelo Rasch en la creación de instrumentos de medición fiables y válidos, facilitando su aplicación en un amplio espectro de contextos.

Así pues, a diferencia de otros modelos de la TRI o de los métodos de la teoría clásica de los tests (TCT), el modelo Rasch ofrece varias ventajas únicas:

- **Medición conjunta**¹³: Significa que los parámetros de las personas y de los ítems se expresan en las mismas unidades y se localizan en el mismo continuo. En consecuencia, la interpretación de las puntuaciones no se fundamenta necesariamente en normas de grupo, sino en la identificación de los ítems que la persona tiene una alta o baja probabilidad de resolver correctamente. Esta característica dota al modelo de Rasch de una gran riqueza diagnóstica.
- **Invarianza de medida:** Como se mencionó anteriormente, una de las contribuciones más significativas del análisis Rasch es su capacidad para facilitar comparaciones válidas entre ítems y personas, independientemente de la muestra de ítems o la población de personas a las que se aplica el test⁴.
- **Objetividad específica**¹⁴: Una medida solo puede ser considerada válida y generalizable si no depende de las condiciones específicas con que ha sido obtenida. Es decir, la diferencia entre 2 personas en un atributo no debe depender de los ítems específicos con los que sea estimada. Igualmente, la diferencia entre 2 ítems no debe depender de las personas específicas que se utilicen para cuantificarla. En consecuencia, si los datos se ajustan al modelo, las comparaciones entre personas son independientes de los ítems administrados y las estimaciones de los parámetros de los ítems no estarán influenciadas por la distribución de la muestra que se usa para la calibración. En la propiedad de objetividad específica se fundamentan aplicaciones psicométricas muy importantes como la equiparación de puntuaciones obtenidas con distintos tests, la construcción de bancos de ítems y los tests adaptados al sujeto.
- **Validación de los ítems:** El análisis Rasch permite una revisión detallada de cada ítem en un test, identificando aquellos que no se ajustan al modelo y que, por tanto,

podrían ser ambiguos o mal interpretados por diferentes subgrupos¹⁵.

- *Escala de intervalo*: Otro aspecto distintivo es que el modelo Rasch convierte las respuestas en una escala de intervalo, lo que facilita la interpretación de las diferencias entre puntuaciones y mejora la precisión de las mediciones¹⁶.

El modelo Rasch original y sus extensiones a modelos de datos politómicos

En el contexto del desarrollo de instrumentos y escalas de medida, los ítems pueden tener diferentes opciones o tipos de respuesta. Así, los datos dicotómicos se refieren a ítems que tienen solo 2 posibles categorías de respuestas: correcto/incorrecto, verdadero/falso, etc. En contraposición, los datos politómicos se refieren a ítems que pueden tener 3 o más categorías de respuesta, algo común en muchos instrumentos de medida desarrollados para medir diferentes constructos, como los instrumentos de calidad de vida, de actitud o de satisfacción. Es frecuente que estos instrumentos utilicen categorías de respuesta ordinales que pueden ir, por ejemplo, desde «totalmente de acuerdo» hasta «totalmente en desacuerdo».

En el caso de datos dicotómicos nos referimos al modelo Rasch original, también conocido como el modelo de Rasch para medidas logísticas uniparamétricas^{1,7}. Sus características principales, además de las ya enunciadas, son la *simplicidad*, puesto que su análisis es relativamente sencillo, ya que cada ítem contribuye con un punto a la puntuación total, facilitando su interpretación y análisis; y ser un *modelo uniparamétrico*, donde solo se estima un parámetro por ítem, la dificultad del ítem, manteniendo la modelación y la interpretación más simple. Este modelo es ideal para pruebas que determinan si una respuesta es simplemente correcta o incorrecta, como en el caso de pruebas de conocimientos o habilidades de enfermería.

En el caso de los datos politómicos, donde estamos utilizando escalas tipo Likert, se utilizan extensiones del modelo Rasch como el «modelo de crédito parcial de Masters» (*Credit Partial Model*)¹⁷ o el «modelo de clasificación de Andrich» (*Rating Scale Model*)¹⁸. En este caso sus características, en contraposición al modelo original, son la *complejidad*, dado que puede manejar la complejidad de los ítems con múltiples categorías de respuesta, permitiendo una evaluación más matizada de las actitudes o habilidades; y el uso de *modelos multiparamétricos* que permiten, además de la estimación de la dificultad del ítem, estimar parámetros adicionales como el umbral de categorías de respuesta.

El modelo de escala de clasificación *Rating Scale Model*

El *Rating Scale Model* (RSM) de Andrich se utiliza para analizar datos de respuesta politómica donde todas las preguntas o ítems comparten el mismo conjunto de categorías de respuesta. Por ejemplo, en una encuesta donde las respuestas varían de 1 a 5, donde 1 es «Totalmente en desacuerdo» y 5 es «Totalmente de acuerdo», todas las preguntas utilizan esta misma escala de 5 puntos, de modo que:

- Las categorías de respuesta son comunes a todos los ítems.
- Se asume que los umbrales entre las categorías de respuesta (la transición de una categoría a la siguiente) son los mismos a través de todos los ítems.
- El modelo estima un conjunto de parámetros de umbral que son aplicables a todos los ítems.

El modelo de crédito parcial *Partial Credit Model*

El *Partial Credit Model* (PCM), desarrollado por Geoff Masters, es más flexible que el RSM y se utiliza cuando las preguntas o ítems tienen diferentes números de categorías de respuesta o cuando no se espera que los umbrales entre categorías sean los mismos a través de todos los ítems. Por ejemplo, algunas preguntas pueden tener una escala de respuesta de 1 a 4, mientras que otras pueden tener respuestas de 1 a 5. Así, este modelo:

- Permite manejar diferentes conjuntos de categorías de respuesta para cada ítem.
- Los umbrales entre categorías de respuesta se estiman por ítem, no asumiéndose que son iguales a través de todos los ítems.
- Ofrece una mayor flexibilidad para manejar ítems con una estructura de respuesta variada.

¿Cuándo elegir uno u otro modelo?

La elección entre RSM y PCM depende de la estructura del instrumento de medición y los objetivos del análisis. El RSM puede ser preferible por su simplicidad y la facilidad de interpretación cuando todas las preguntas utilizan la misma escala de respuesta. PCM ofrece mayor flexibilidad y es adecuado para instrumentos con variedad en las escalas de respuesta, permitiendo un análisis más detallado y específico para cada ítem. Para datos dicotómicos, donde las respuestas a los ítems son de naturaleza binaria (p.ej., correcto/incorrecto, sí/no), el modelo de Rasch original, como hemos visto, es el más adecuado y directamente aplicable.

Metodología de aplicación del análisis Rasch

La implementación efectiva del análisis Rasch en la investigación implica varios pasos críticos, desde el diseño inicial del estudio hasta la interpretación de los resultados. A continuación, se detallan 11 pasos esenciales en este proceso (fig. 1):

1. Paso/definición de los objetivos y preparación: El primer paso es establecer claramente los objetivos de la investigación y determinar cómo el análisis Rasch puede contribuir a alcanzarlos, ya sea mediante la evaluación de la validez de un instrumento existente, el desarrollo de uno nuevo, el refinamiento o reducción de instrumentos ya existentes o, incluso, la equiparación de diferentes tests en la medida.
2. Paso/diseño del instrumento de medición: Seleccionar o diseñar ítems que sean representativos del constructo que se desea medir. Esto incluye asegurar que los ítems



Figura 1 Esquema que representa los 11 pasos para aplicar el análisis RASCH.

sean claros, unidimensionales y adecuados para el grupo objetivo.

- Paso/recolección de datos: Realizar la recolección de datos asegurando una muestra representativa y suficientemente amplia para el análisis estadístico. No obstante, en la literatura no hay acuerdo sobre el tamaño muestral necesario para este tipo de análisis. Así, Stolt et al.³ presentan estudios que usan muestras que van desde 43 hasta 13.113 participantes. El análisis Rasch puede funcionar relativamente bien independientemente del tamaño de la muestra por las características mencionadas, que le otorga una ventaja frente a otros tipos de análisis como el análisis factorial. Habitualmente, muchos estudios hacen referencia a la cita de Polit y Beck¹⁹ sobre la utilización de muestras en base a 5-20 participantes por cada ítem del instrumento a analizar.
- Paso/análisis inicial de los datos: Usar un *software* de análisis Rasch para realizar un análisis inicial, identificando ítems que no funcionan bien, patrones inesperados en las respuestas y la alineación general de los datos con el modelo Rasch. En muchas ocasiones, este análisis inicial se combina con análisis sencillos de la TCT para estudiar la dificultad, desconocimiento y discriminación. Esto permite identificar ítems que no aportan beneficios al constructo y que se pueden eliminar previamente a la realización del análisis Rasch.

En esta fase, es habitual, en la literatura, encontrar que para estudiar la unidimensionalidad se lleve a cabo un análisis de componentes principales de los residuos del modelo²⁰. Los criterios habituales son que el primer componente explique al menos el 50% de la varianza y que el segundo componente explique menos

del 5% (o un autovalor —Eigenvalue— inferior a 2,0)²¹. Otra opción es, en el marco de un análisis factorial, calcular diferentes índices de unidimensionalidad, tal y como proponen Ferrando y Lorenzo-Seva²². El cálculo de los valores *Unidimensional Congruence* (UniCo), *Explained Common Variance* (ECV) y *Mean of Item Residual Absolute Loadings* (MIREAL) pueden ayudar a evaluar si los datos pueden ser tratados como esencialmente unidimensionales²².

- Paso/evaluación de la bondad de ajuste: nos proporciona datos de validez interna de la escala y de validez de la respuesta de las personas. Se utilizan estadísticas de ajuste, normalmente, los denominados *infit* y *outfit*^{4,23}. Los ítems o respuestas que no se ajustan pueden indicar problemas con el ítem o la unidimensionalidad del constructo. Es crucial identificar los ítems que no se ajustan al modelo de Rasch. La inclusión de ítems con un ajuste deficiente dificulta la calidad de la medición y disminuye la precisión del instrumento. Estos ítems deben eliminarse, revisarse o reescribirse, y volver a someterse a prueba. Las estadísticas de bondad de ajuste indican hasta qué punto cada ítem se ajusta al constructo subyacente de la prueba⁴.

El modelo de estimación conjunta de máxima verosimilitud, o modelo *Joint Maximum Likelihood Estimation* (JMLE, por sus siglas en inglés) se refiere a una técnica estadística utilizada para la estimación de parámetros dentro de modelos de la TRI, incluido el análisis Rasch. Este enfoque se caracteriza por estimar simultáneamente los parámetros de los ítems y los parámetros de habilidad de los sujetos basándose en la máxima verosimilitud conjunta de todos estos parámetros dados los datos observados⁶.

FINAL JMLE ITEM STATISTICS						
Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
aciertos1	-1,29	0,11	1,05	0,86	1,06	0,76
aciertos2	-0,87	0,10	1,22	4,76	1,36	4,40
aciertos3	-0,38	0,10	1,08	2,11	1,04	0,55
aciertos4	-1,44	0,11	1,03	0,60	1,01	0,14
aciertos5	-1,21	0,11	1,05	1,06	1,07	0,92
aciertos7	-1,41	0,11	1,00	-0,01	1,02	0,29
aciertos8	1,72	0,13	0,95	-0,65	0,98	-0,01
aciertos9	-0,61	0,10	1,06	1,42	1,06	0,83
aciertos11	-1,65	0,11	1,07	1,11	1,09	0,91
aciertos12	-1,06	0,10	0,96	-0,94	0,90	-1,26
aciertos13	-2,98	0,16	1,14	1,25	1,46	2,07
aciertos14	0,72	0,10	1,00	0,06	1,06	0,53
aciertos15	1,27	0,12	1,00	0,04	1,10	0,63
aciertos18	-0,40	0,10	0,99	-0,21	0,98	-0,27
aciertos19	0,21	0,10	0,99	-0,34	0,95	-0,51
aciertos20	-1,06	0,10	1,13	2,59	1,16	2,00
aciertos22	0,02	0,10	1,05	1,48	1,13	1,44

Figura 2 Representación parcial de una tabla de valores estimados por JMLE para un hipotético instrumento de medida (dificultad, *infit* [WMS] y *outfit* [UMS]).

En el contexto de la TRI y el análisis Rasch, los parámetros de los ítems generalmente incluyen la dificultad de los ítems (y potencialmente otros parámetros, como la discriminación, dependiendo del modelo específico de TRI utilizado). Los parámetros de los sujetos, por otro lado, suelen referirse a las habilidades o competencias de los individuos que responden a los ítems. La estimación conjunta significa que el proceso de estimación busca aquel conjunto de valores de parámetros de ítems y sujetos que, en conjunto, tienen la mayor probabilidad de haber generado los datos observados.

Este modelo presenta ventajas como son la «eficiencia», pues permite la estimación simultánea de parámetros de ítems y sujetos, lo que puede ser más eficiente desde el punto de vista computacional en comparación con los enfoques que estiman estos parámetros por separado, y la «aplicabilidad», que es especialmente útil en situaciones donde no se dispone de información previa sobre los parámetros de los ítems o las habilidades de los sujetos, permitiendo una estimación directa a partir de los datos recogidos. Pero este modelo también presenta limitaciones como son la «inexactitud en extremos» (la estimación conjunta puede ser menos precisa para los sujetos con habilidades extremadamente altas o bajas, ya que la información para estimar sus habilidades de manera precisa puede ser limitada en los datos observados) y la «estabilidad» (pues puede ser susceptible a problemas de estabilidad y convergencia, especialmente en conjuntos de datos pequeños o con patrones de respuesta poco comunes).

Aunque el JMLE es una técnica poderosa dentro del análisis Rasch y otros modelos de la TRI, es importante que los investigadores sean conscientes de sus limitaciones y consideren la posibilidad de utilizar métodos alternativos de estimación, como el *Marginal Maximum Likelihood Estimation* (MMLE) o el *Conditional Maximum Likelihood Estimation* (CMLE), dependiendo de las características específicas de sus datos y los objetivos de su estudio. La elección del método de estimación debe basarse en una consideración cuidadosa de estos factores

para asegurar la precisión y la validez de los resultados del análisis.

Desde el modelo JMLE, el *infit* (información/ponderada), o *Weighted Mean Square Fit Statistic* (WMS), es una medida de ajuste que es especialmente sensible a los patrones de respuesta en los ítems que están cerca del nivel de habilidad del individuo. El *outfit* (*Outlier-Sensitive Fit*), o *Unweighted Mean Square Fit Statistic* (UMS), es una estadística de ajuste crudo, más sensible a respuestas inusuales o atípicas («outliers») entre los ítems que están lejos del nivel de habilidad del individuo. Estos valores se expresan como cuadrados medios. Un valor de 1 indica un ajuste perfecto al modelo. Así, los valores de los índices de ajuste comprendidos entre 0,8 y 1,2 representan un buen ajuste¹⁵ y los valores comprendidos entre 0,5 y 1,5 significan un ajuste aceptable¹¹, pero también se recomienda ajustar los valores según el tamaño de la muestra²⁴. En la figura 2 se representa un resultado parcial de un hipotético análisis, donde se pueden ver: la dificultad y los valores de *infit* (WMS) y *outfit* (UMS) con sus respectivos errores estándar.

A modo de ejemplo, el ítem 13 sería el más fácil, el 22 estaría cercano a 0 (dificultad media) y el 15 sería el más difícil de los presentados en esta tabla.

En el caso de la dificultad, valores cercanos a cero estarían en la mitad de dificultad, mientras valores mayores de cero, a medida que se alejan indican mayor dificultad; valores menores de cero indican una menor dificultad. En dos los valores se encuentran en el rango de valores que indican buen ajuste al modelo.

Como hemos visto, el análisis de estos parámetros es de vital importancia pues nos permite el «Diagnóstico de ítems». Estas estadísticas permiten identificar ítems que no se comportan según lo esperado, ya sea porque son demasiado predecibles o porque presentan una variabilidad inesperada en las respuestas y la «Mejora del instrumento», ya que, al analizar el *infit* y el *outfit*, los desarrolladores de pruebas pueden tomar decisiones informadas sobre la revisión o eliminación de ítems para

mejorar la validez y fiabilidad del instrumento. Finalmente, este análisis se relaciona directamente con la «*Adaptación al modelo*»; la evaluación del ajuste de los ítems al modelo Rasch (o cualquier modelo de TRI aplicado) es crucial para asegurar que las mediciones reflejen con precisión el constructo de interés.

Otras medidas que nos aporta el análisis de bondad de ajuste son los estadísticos de calidad de la escala⁶, *fiabilidad* y *separación*, que pueden calcularse tanto para los ítems como para las personas. La fiabilidad de la persona es similar al coeficiente de fiabilidad de la TCT, siendo deseables valores superiores a 0,8. La separación de personas es similar a la fiabilidad en el sentido de que representa hasta qué punto una medida puede reproducir y clasificar las puntuaciones de forma coherente. Los valores de separación superiores a 2 son deseables, lo que indica que el instrumento puede separar a las personas en al menos 2 estratos, por ejemplo, baja y alta capacidad²⁵. La fiabilidad de los ítems se refiere al grado en que las dificultades de los ítems pueden ordenarse por rango. La separación entre ítems proporciona información similar sobre la calidad de la localización de los ítems en el rasgo latente.

6. Paso/evaluación de la independencia local: Normalmente, se lleva a cabo un análisis de la correlación los residuos estandarizados, lo que se conoce como el coeficiente Q3 de Yen, una medida estadística desarrollada por Yen²⁶ en 1984, y utilizada para evaluar la independencia local de los ítems en el contexto de los modelos de la TRI, incluido el análisis Rasch. Es una medida de correlación residual entre pares de ítems. En otras palabras, evalúa la correlación entre las respuestas a 2 ítems después de haber controlado por la habilidad general del respondiente. Si 2 ítems son verdaderamente independientes uno del otro, dada la habilidad de un individuo, entonces se espera que el coeficiente Q3 sea cercano a cero. Un valor de Q3 significativamente distinto de cero sugeriría que existe una dependencia entre los ítems, lo cual podría indicar una violación de la suposición de independencia local, propiedad necesaria para asumir el modelo Rasch, tal y como se expuso inicialmente.

El coeficiente Q3 de Yen se aplica en el análisis de la validez de un instrumento de medición por varias razones:

- Identificación de Ítems dependientes: Permite identificar pares de ítems que pueden estar influyéndose mutuamente, lo cual es particularmente útil en la revisión y mejora de instrumentos de medición. Por ejemplo, 2 ítems pueden ser redundantes o pueden estar midiendo el mismo sub-rasgo o habilidad específica, en lugar de aspectos únicos del constructo general.
- Mejora del diseño del instrumento: Al identificar ítems con dependencias, los investigadores y desarrolladores de pruebas pueden tomar decisiones informadas sobre qué ítems modificar o eliminar para mejorar la independencia local, la validez y la fiabilidad del instrumento.
- Evaluación de la estructura del constructo: En algunos casos, las dependencias identificadas entre ítems pueden proporcionar información valiosa sobre la estructura subyacente del constructo que se está

midiendo. Esto puede llevar a una mejor comprensión de las dimensiones del constructo y a la refinación teórica del mismo.

Es importante notar que la implementación de este análisis requiere el uso de *software* estadístico especializado que pueda manejar modelos de la TRI y calcular correlaciones residuales. Valores altos de Q3 pueden indicar problemas con la independencia de los ítems, pero también pueden reflejar aspectos multidimensionales del constructo que no necesariamente invalidan el uso del instrumento. Por tanto, los hallazgos deben ser evaluados en el contexto de la teoría subyacente del constructo y el propósito del instrumento de medición. Así pues, la interpretación de los valores del coeficiente Q3 de Yen, específicamente el umbral en el cual se considera que un valor de Q3 indica una dependencia inaceptable entre ítems, puede variar dependiendo del contexto del estudio y las normas establecidas por el investigador o la disciplina. Sin embargo, hay algunas guías generales que se pueden considerar. Yen, en su trabajo original, no especificó un umbral estricto para determinar cuándo un valor de Q3 es inaceptablemente alto, lo que sugiere que la interpretación depende en gran medida del contexto específico del análisis y del instrumento. En la práctica, algunos investigadores y metodólogos han adoptado enfoques pragmáticos para interpretar los valores de Q3, considerando valores por encima de ciertos puntos de corte como indicativos de una posible dependencia entre ítems. Un enfoque común es considerar valores de Q3 superiores a 0,2 o 0,3 como indicativos de una correlación residual significativa entre pares de ítems, sugiriendo que estos ítems podrían no ser independientes entre sí después de controlar por la habilidad general de los respondientes. Esta recomendación general debe usarse con precaución; algunos contextos pueden requerir un umbral más conservador o más liberal, dependiendo de la naturaleza del constructo medido y el propósito del instrumento de evaluación. En realidad, no existe un criterio uniforme, ya que este valor depende del tamaño de la muestra, del número de ítems y del número de categorías de respuesta²⁷. A modo de ejemplo, en la [figura 3](#), se presenta una sección de una matriz de correlaciones de Q3 de Yen donde se puede observar que todos los valores son inferiores a 0,2 en valor absoluto, lo que estaría indicando independencia local de los ítems.

También es importante destacar que un valor alto de Q3 no necesariamente indica que los ítems deban ser eliminados o que el instrumento sea inválido. En su lugar sugiere la necesidad de una revisión más detallada de los ítems implicados para entender la razón de la dependencia. Esto podría llevar a considerar aspectos como la redacción de los ítems, la estructura del cuestionario, o incluso la posibilidad de dimensiones subyacentes no reconocidas dentro del constructo que se está midiendo.

7. Paso/calibración de ítems y estimación de habilidades¹: A través del análisis Rasch se estima la dificultad de cada ítem y se calculan las habilidades o niveles del constructo para cada participante. Una vez llevado a cabo todo lo anterior, los datos pueden representarse en gráficos que permiten una mejor interpretación. Son comunes los mapas ítems-personas o gráfico de Wright, que se compone de 2 grandes histogramas (graficados de forma

Q3 Correlation Matrix

	aciertos3	aciertos5	aciertos7	aciertos9	aciertos11	aciertos12
aciertos3	—					
aciertos5	0.055	—				
aciertos7	-0.003	0.098	—			
aciertos9	-0.074	-0.020	-0.068	—		
aciertos11	0.012	0.147	-0.024	0.003	—	
aciertos12	-0.025	-0.051	0.034	0.010	0.011	—
aciertos14	-0.013	-0.064	0.023	-0.002	0.021	-0.047
aciertos15	0.037	-0.010	-0.059	-0.053	-0.064	0.016
aciertos20	-0.029	0.029	-0.012	0.044	0.052	-0.052
aciertos23	-0.055	-0.009	0.079	-0.004	-0.026	-0.134
aciertos28	0.019	-0.040	-0.056	-0.003	0.063	-0.054
aciertos29	-0.017	0.047	0.039	-0.046	0.030	0.074

Figura 3 Representación parcial de una matriz de correlaciones Q3 de Yen.

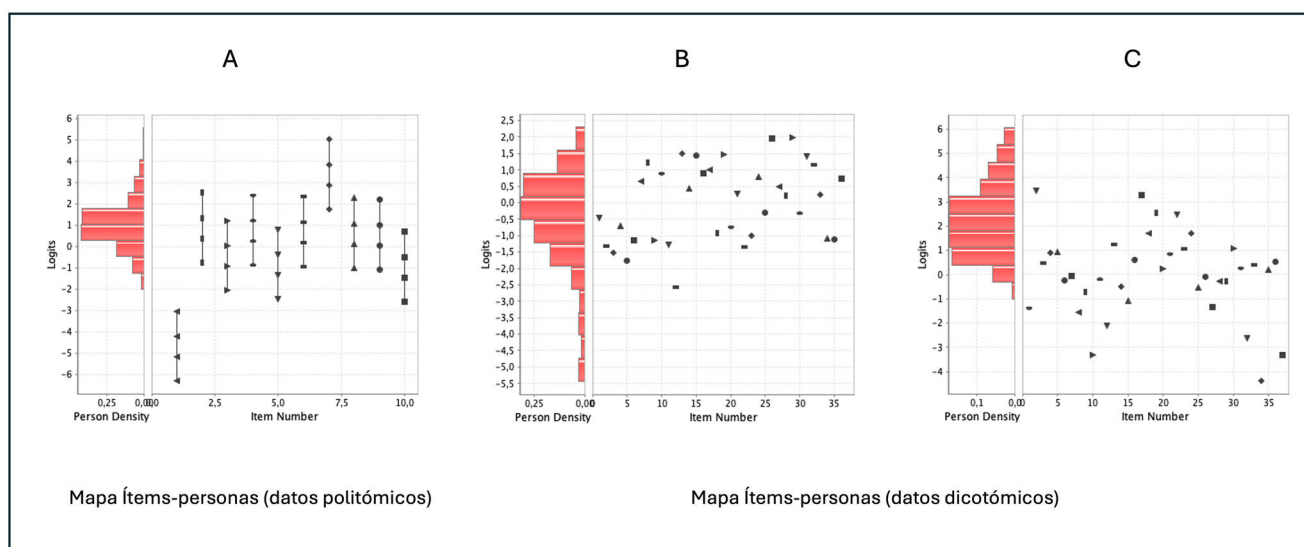


Figura 4 Ejemplos de gráficos de Wright o mapas de ítems/personas.

vertical). El lado izquierdo presenta el histograma de la distribución de habilidades de las personas. El lado derecho presenta el histograma de la distribución de dificultades de los ítems. Los 2 histogramas comparten el eje de valores (línea vertical).

En la figura 4 se presentan 3 gráficos de Wright diferentes. En el primero es un gráfico de un modelo con datos politómicos. Por ello, en la parte derecha del gráfico aparecen líneas verticales con 4 marcas que representan cada ítem y la dificultad de cada categoría (10 ítems con 4 categorías de respuesta). En la izquierda estaría el histograma de personas. Ambos están en la misma escala logit, lo que permite comparar ambos elementos en la misma escala. Vemos, que, aunque se ve una distribución bastante normal, esta está desplazada hacia arriba en habilidad. En el caso de los ítems, hay una distribución homogénea a lo largo de la escala logit, lo que indica que hay ítems más fáciles (el que está más abajo) y otros más difíciles (el que está más arriba), aunque excepto para el primero, la distribución de personas e ítems se encuentra prácticamente igual.

En el caso de los otros 2 gráficos, son modelos con datos dicotómicos (los ítem se representan con diferentes símbolos). En el gráfico central vemos que hay un grupo de personas que tienen una tendencia a puntuaciones bajas en el instrumento y que los ítems están más arriba, lo que puede suponer que este instrumento (si fuera de conocimientos) presenta preguntas relativamente difíciles para una parte de la muestra. En el gráfico de la derecha es todo lo contrario, las personas se desempeñan muy bien y los ítems parecen más fáciles.

Los gráficos de la curva característica de los ítems (CCI), que ofrece una representación gráfica de las características del ítem, y por otra parte muestra la relación entre el nivel de habilidad y la repuesta al ítem, en términos probabilísticos; y la curva característica del test (CCT) (fig. 5), que muestra la relación existente entre la puntuación total en un test (no solamente en un ítem como sucede con la CCI) y el nivel de habilidad de una persona. En el caso de la figura 5 el gráfico presenta la curva característica de un test que se ajusta al modelo Rasch con una curva sigmoide que enfrenta la puntuación en la variable

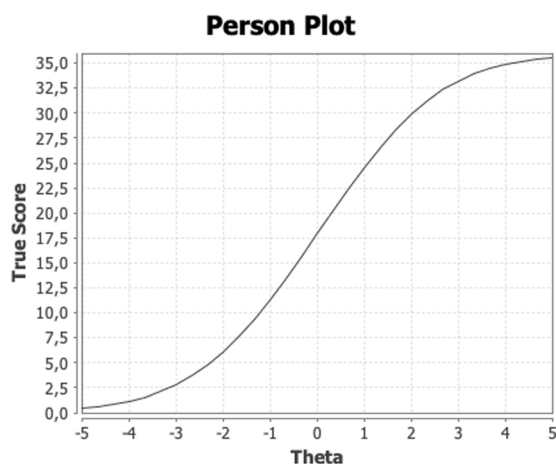


Figura 5 Curva característica del test.

latente (theta) frente a las puntuaciones verdaderas de las personas.

8. Paso/análisis de la invarianza de las medidas: asegurando que las mediciones son consistentes a través de diferentes subgrupos de la muestra, como por género, edad u otras características. Esto es crucial para confirmar que el instrumento mide de manera equitativa para todos los participantes. Normalmente, se lleva a cabo mediante un análisis del funcionamiento diferencial de los ítems (DIF)^{4,6}.

El análisis de funcionamiento diferencial del ítem (DIF, por sus siglas en inglés) es una técnica estadística que se utiliza para identificar si diferentes grupos de individuos responden de manera distinta a un ítem específico, a pesar de tener el mismo nivel de habilidad o rasgo latente.

Aunque el análisis DIF y el análisis Rasch son conceptos distintos, pueden estar relacionados en la práctica de la evaluación psicométrica. El análisis Rasch proporciona un marco para medir un rasgo latente basado en las respuestas a los ítems de un test, y uno de sus principales atractivos es la capacidad de generar mediciones independientes de la muestra de ítems y la población de respondientes. Incorporar el análisis DIF en este contexto ayuda a garantizar que los ítems del instrumento sean justos para todos los grupos de individuos, lo que es esencial para la validez de las mediciones obtenidas a través del modelo Rasch. Por tanto, el propósito principal del análisis DIF es asegurar la equidad y la imparcialidad en la evaluación, identificando ítems que puedan ser sesgados o funcionar de manera diferente para subgrupos específicos dentro de la población, como pueden ser subgrupos definidos por género, etnicidad, edad, o cualquier otra variable demográfica o de agrupación relevante. Un ítem que muestra DIF sugiere que algo, además de la habilidad general que se está midiendo afecta cómo diferentes grupos responden a ese ítem.

El análisis DIF se puede realizar mediante varios métodos, incluidos enfoques gráficos, pruebas estadísticas y técnicas basadas en modelos como la regresión logística o el propio análisis Rasch. Los métodos basados en modelos Rasch, en particular, permiten evaluar el DIF mientras se mantiene la

estructura del modelo, facilitando la interpretación de cómo el sesgo de ítem afecta la medición del rasgo latente.

En el contexto del análisis Rasch, el DIF se evalúa comparando la dificultad del ítem (parámetro de localización del ítem) entre diferentes grupos, controlado por la habilidad general. Este proceso se realiza manteniendo la estructura del modelo Rasch, lo que significa que la estimación de la dificultad del ítem y la habilidad del respondiente se basa en el principio de máxima verosimilitud, asumiendo que los ítems son independientes y que la probabilidad de una respuesta correcta solo depende de la diferencia entre la habilidad del respondiente y la dificultad del ítem.

Algunos métodos que pueden servir a este cometido son:

- Análisis de contraste de grupos: una técnica común es dividir la muestra según el grupo de interés (p. ej., género, etnicidad) y realizar un análisis Rasch por separado para cada grupo. Luego, se compara la dificultad estimada de cada ítem entre los grupos. Un cambio significativo en la dificultad del ítem indica la presencia de DIF.
- Modelos de Rasch multifacéticos: Algunas versiones del modelo Rasch permiten incorporar directamente el grupo como una «faceta» en el análisis. Esto facilita la estimación simultánea de los efectos del grupo sobre la dificultad del ítem, proporcionando una forma directa de evaluar el DIF.
- Pruebas estadísticas específicas: Se utilizan pruebas estadísticas diseñadas para detectar diferencias en los parámetros de los ítems entre grupos, como el test de Mantel-Haenszel o pruebas basadas en la regresión logística, adaptadas al contexto del modelo Rasch. Estas pruebas comparan la dificultad del ítem entre grupos después de ajustar por la habilidad del respondiente.

La evaluación DIF en el contexto del análisis Rasch puede utilizar las *odds ratio* (OR) como una herramienta para medir la magnitud del DIF entre grupos. Las OR ofrecen una forma cuantitativa de comparar las probabilidades de que los individuos de diferentes grupos (p. ej., varones vs. mujeres, o grupos etarios distintos) respondan correctamente a un ítem, dado un nivel similar de habilidad o rasgo latente. Así, las OR se calculan para cada ítem y comparan la probabilidad de una respuesta correcta entre 2 grupos de interés. Un valor de OR de 1 indica que los grupos tienen la misma probabilidad de responder correctamente al ítem, después de ajustar por la habilidad. Valores significativamente mayores o menores que 1 sugieren la presencia de DIF, es decir, que un ítem favorece a un grupo sobre el otro.

Proceso para evaluar DIF con *odds ratio* en análisis Rasch

La secuencia sería la siguiente:

- Estimación de la habilidad: Primero, se utiliza el análisis Rasch para estimar las habilidades de los individuos basadas en sus patrones de respuesta a los ítems del test.
- Cálculo de las OR: Luego, para cada ítem, se calculan las OR de responder correctamente para los individuos en diferentes grupos, controlando por la habilidad. Esto se hace comparando las OR de éxito en un grupo con las del otro.

Item	Chi-square	p-value	Valid N	Q.R. (95% C.I.)			Class
aciertos1	6,97	0,01	417	1,88 (	1,18,	3,01)	B-
aciertos2	0,15	0,69	450	0,92 (	0,59,	1,41)	A
aciertos3	1,00	0,32	441	1,26 (	0,81,	1,95)	A
aciertos4	6,64	0,01	433	1,83 (	1,15,	2,92)	B-
aciertos5	2,76	0,10	433	1,46 (	0,93,	2,29)	A
aciertos7	0,01	0,90	423	0,97 (	0,60,	1,56)	A
aciertos8	11,40	0,00	356	0,36 (	0,19,	0,69)	B+
aciertos9	2,52	0,11	433	1,43 (	0,92,	2,21)	A
aciertos11	0,21	0,65	407	1,12 (	0,69,	1,82)	A
aciertos12	2,87	0,09	390	1,51 (	0,94,	2,42)	A
aciertos13	11,40	0,00	358	2,82 (	1,47,	5,44)	B-
aciertos14	0,02	0,89	407	1,03 (	0,64,	1,66)	A
aciertos15	3,38	0,07	402	1,75 (	0,96,	3,18)	A
aciertos18	10,40	0,00	432	2,07 (	1,33,	3,23)	B-
aciertos19	8,12	0,00	437	1,90 (	1,21,	2,98)	B-
aciertos20	3,39	0,07	450	1,51 (	0,97,	2,34)	A
aciertos22	7,06	0,01	415	0,54 (	0,34,	0,85)	B+

Figura 6 Representación parcial de un análisis DIF por medio de la OR.

- Análisis estadístico: Se realiza un análisis estadístico para determinar si las diferencias en las OR son significativas, lo que puede implicar el uso de pruebas de hipótesis o intervalos de confianza. La interpretación según la OR:
- $OR > 1$: Indica que el grupo de referencia tiene mayores odds de responder correctamente al ítem en comparación con el grupo de comparación, sugiriendo un posible sesgo a favor del grupo de referencia.
- $OR < 1$: Sugiere que el grupo de comparación tiene mayores odds de responder correctamente, indicando un sesgo a favor de este grupo.
- $OR \approx 1$: Implica que no hay evidencia significativa de DIF para ese ítem entre los grupos comparados.

Es importante tener en cuenta que al utilizar las OR para evaluar DIF en el análisis Rasch, es crucial considerar el tamaño del efecto y la significación estadística. No todos los ítems con OR significativamente distintos de 1 necesariamente tienen un impacto práctico significativo en la medición. Por lo tanto, es esencial interpretar las OR en el contexto de la relevancia práctica y en combinación con otras medidas de DIF y análisis cualitativos de los ítems. En la figura 6 se presentan resultados parciales de un análisis DIF mediante prueba de Mantel-Haenszel para estimar la OR de un grupo de referencia frente a un grupo focal (estudiantes de 1.º-2.º curso frente a estudiantes de 3.º-4.º curso en un test de conocimientos). Como se puede observar hay una serie de ítems que presentan DIF moderado (B+ o B- en función de qué grupo tiene más habilidad para ese ítem). Esto nos ayuda a evaluar los ítems que tienen un funcionamiento diferente y que, por tanto, pueden estar midiendo otro constructo o que presentan diferencias por determinados elementos del grupo en cuestión. Esto violaría los supuestos del modelo Rasch. Por tanto, los ítems que presentan DIF (en este caso moderado, representado por la letra B) deberían ser eliminados.

9. Paso/visión y modificación del instrumento: Basándose en los resultados del análisis inicial (pasos 4, 5, 6 y 7), revisar y posiblemente modificar el instrumento. Esto puede implicar eliminar ítems que no se ajustan

bien, ajustar el formato de respuesta o clarificar el enunciado de los ítems. Este paso puede requerir más de un ajuste.

10. Paso/análisis final y validación del instrumento: Realizar un análisis Rasch final con el instrumento revisado para validar su estructura y determinar la fiabilidad y validez del mismo (repitiendo los pasos anteriores).
11. Paso/interpretación de resultados: Interpretar los resultados en el contexto de los objetivos del estudio, centrándose en la calidad de las mediciones, la validez del instrumento y las implicaciones para la práctica y la investigación futura. Esto implica examinar varios aspectos clave de los datos y el instrumento de medición. También incluye interpretar las escalas, los mapas de ítems/personas y las curvas de información.

Algunos elementos a tener en cuenta:

- a. Ajuste al modelo: Evaluar cómo los datos se ajustan al modelo Rasch, incluyendo el ajuste de los ítems individuales y el ajuste general del instrumento. Esto se hace mediante estadísticas de ajuste, como los residuos estandarizados y las medidas de *infit* y *outfit*.
- b. Unidimensionalidad: Confirmar que el conjunto de ítems mide un único constructo o dimensión, lo cual es fundamental para la validez del instrumento.
- c. Independencia local: Verificar que las respuestas a los ítems son independientes entre sí, después de controlar por la habilidad de la persona, lo que es esencial para la fiabilidad del instrumento.
- d. Estimación de parámetros: Interpretar los parámetros estimados por el modelo, incluyendo la dificultad de los ítems y la habilidad de las personas. Esto proporciona información valiosa sobre las propiedades del instrumento y las características de la población estudiada.
- e. Implicaciones prácticas: Finalmente, es crucial considerar las implicaciones prácticas de los resultados, incluyendo cómo pueden informar la práctica clínica, la educación en enfermería o la investigación futura.

Herramientas y *software* recomendados

Para llevar a cabo el análisis Rasch, existen varias herramientas y *software* especializados que facilitan este proceso. Algunas de las opciones más populares incluyen:

- RUMM2030® (<https://www.rummlab.com.au>): Una herramienta poderosa para el análisis Rasch, especialmente útil para el análisis de datos politómicos y la evaluación de la calidad del ajuste de los ítems.
- Winsteps® (<http://www.winsteps.com>)¹¹: Ampliamente utilizado en la comunidad de análisis Rasch, Winsteps ofrece una amplia gama de funcionalidades para el análisis de datos dicotómicos y politómicos.
- FACETS²⁸: Específicamente diseñado para el análisis de datos multifacéticos, FACETS es ideal para estudios que involucran evaluadores múltiples o ítems con complejidades adicionales.
- JMetrik® [<https://itemanalysis.com>]⁶: Una herramienta gratuita que ofrece capacidades para el análisis Rasch, junto con otras técnicas estadísticas comunes.

La elección del *software* adecuado dependerá de las necesidades específicas del estudio, incluyendo el tipo de datos a analizar, la complejidad del modelo y las preferencias del investigador.

Aplicación del análisis Rasch en enfermería

En esta sección, se explora cómo el análisis Rasch se aplica en la disciplina, destacando estudios de caso relevantes que ilustran su impacto práctico.

- Evaluación de competencias y habilidades clínicas

La evaluación precisa de las competencias y habilidades clínicas es fundamental en la formación de enfermería y en la práctica profesional continua. El análisis Rasch ofrece un marco robusto para esta evaluación, permitiendo a los educadores y administradores identificar con precisión las áreas de fortaleza y las necesidades de desarrollo entre los estudiantes y las enfermeras.

- Modelado de competencias: Mediante el uso del análisis Rasch, se puede modelar la complejidad de las competencias clínicas, diferenciando entre niveles de habilidad y asegurando que las evaluaciones sean adecuadas para el nivel de competencia esperado.
- Desarrollo curricular: Los resultados del análisis Rasch pueden guiar el desarrollo curricular y las intervenciones educativas, asegurando que los programas de formación en enfermería estén alineados con las necesidades de competencia identificadas.
 - Desarrollo y validación de instrumentos de medición en enfermería.

El desarrollo de instrumentos de medición fiables y válidos es otro ámbito crucial donde el análisis Rasch tiene un impacto significativo. Estos instrumentos son esenciales para la investigación en enfermería, la evaluación de programas educativos y la medición de resultados en la práctica clínica.

- Construcción de instrumentos: El análisis Rasch facilita la construcción de instrumentos de medición al proporcionar un método para evaluar la unidimensionalidad de las escalas y la adecuación de los ítems, lo que permite la obtención de herramientas de evaluación más precisas y fiables.
- Validación de escalas: A través del análisis Rasch, es posible validar escalas de medición existentes, refinando los ítems para mejorar la precisión y la relevancia clínica de las mediciones. Esto incluye la capacidad de identificar ítems sesgados culturalmente o que no funcionan de manera uniforme entre diferentes subgrupos de población.
 - Estudios de caso relevantes

Los siguientes estudios de caso ilustran la potencial aplicación práctica y los beneficios del análisis Rasch en el campo de la enfermería:

- Uso y calidad de reporte sobre análisis Rasch en enfermería: Stolt et al³ llevan a cabo una revisión exploratoria donde destacan que el uso de esta metodología en enfermería es poco sistemática, pero que es recomendable el uso de este enfoque metodológico.
- Desarrollo de instrumentos para medir el conocimiento en profesionales y estudiantes en diferentes áreas: diferentes autores han utilizado el análisis Rasch para desarrollar y validar instrumentos para medir conocimiento en diferentes áreas, por ejemplo, el cuidado del Alzheimer²⁹, instrumentos de calidad de vida en adolescentes³⁰ o conocimientos sobre la prevención de las lesiones por presión^{31,32}.
- Instrumentos para medir condiciones clínicas: como por ejemplo el desarrollado para medir el síndrome visual por ordenador en el lugar de trabajo³³. En este caso, también es un ejemplo de combinación de métodos desde la perspectiva de TCT y la TRI.
- Refinamiento o revalidación de instrumentos: en este caso, mediante el análisis Rasch se redujo el número de ítems de la escala de úlceras por presión en lesionados medulares, mejorando su validez y fiabilidad³⁴.
- Vinculación o equiparación de diferentes instrumentos: La vinculación de las escalas de 2 puntos de medición es un requisito previo para examinar un cambio en la competencia a lo largo del tiempo. En las evaluaciones educativas a gran escala, los formularios de pruebas no idénticas que comparten una serie de ítems suelen escalarse y vincularse utilizando modelos de respuesta al ítem, como en el caso del artículo publicado por Fischer et al³⁵.

Estos ejemplos muestran cómo el análisis Rasch se aplica para abordar desafíos específicos en la educación y la práctica de enfermería, desde la evaluación de competencias hasta el desarrollo y la validación de herramientas de medición. La metodología Rasch ofrece una base sólida para la mejora continua en estos ámbitos, contribuyendo a elevar los estándares de atención y educación en enfermería.

Ventajas, limitaciones y desafíos de la aplicación de este método

A modo de resumen, a continuación se presentan los beneficios de utilizar esta metodología:

- Medición precisa y objetiva: El análisis Rasch permite la conversión de respuestas cualitativas a mediciones cuantitativas en una escala de intervalo, lo que proporciona una base sólida para la toma de decisiones basada en evidencia.
- Validación de instrumentos: Facilita la validación y refinamiento de instrumentos de medición, asegurando que sean tanto fiables como válidos para evaluar las competencias y habilidades clínicas en enfermería.
- Comparabilidad de mediciones: Una de las ventajas más significativas es la capacidad de comparar mediciones a través de diferentes grupos de población o momentos en el tiempo, lo que es fundamental para los estudios longitudinales y la evaluación de intervenciones.
- Identificación de ítems problemáticos: El análisis detallado que ofrece el análisis Rasch permite identificar ítems con funcionamiento inadecuado, ya sea por su dificultad, ambigüedad o sesgo cultural, contribuyendo a mejorar la calidad de los instrumentos de medición.
- Desarrollo profesional y educativo: Los resultados del análisis Rasch pueden informar el desarrollo curricular y las estrategias de enseñanza en programas de enfermería, alineando la educación con las competencias clínicas requeridas en la práctica.

Por todo lo expuesto, se puede considerar el análisis Rasch una valiosa herramienta metodológica, pero que no está exenta de desafíos y limitaciones, como:

- Complejidad metodológica: A pesar de sus beneficios, el análisis Rasch puede ser metodológicamente complejo, requiriendo para el investigador de una comprensión sólida de sus principios y suposiciones para su aplicación adecuada.
- Requisitos de datos: La aplicación efectiva del modelo Rasch requiere conjuntos de datos completos y bien estructurados, lo que puede ser un desafío en estudios con tasas altas de no respuesta o datos faltantes.
- Interpretación de resultados: La interpretación de los resultados del análisis Rasch, especialmente en lo que respecta a las estadísticas de ajuste y la unidimensionalidad, puede ser compleja y requiere una consideración cuidadosa.
- Limitaciones en la flexibilidad del modelo: Aunque el análisis Rasch es poderoso, puede ser menos flexible que otros modelos de la TRI, en términos de modelado de datos que no se ajustan perfectamente a sus suposiciones. El análisis Rasch asume que todos los ítems de una prueba tienen la misma importancia y que las respuestas de los participantes se pueden modelar utilizando una sola dimensión latente. Esta suposición puede no ser válida para todos los conjuntos de datos, especialmente en áreas donde los constructos son multidimensionales⁴.

Algunas propuestas de solución a las limitaciones o recomendaciones para mejorar el conocimiento, difusión y aplicación del análisis Rasch en la investigación enfermera podrían ser:

- a. Formación y capacitación: Fomentar la formación en análisis Rasch entre los investigadores en enfermería, ofreciendo talleres, cursos y recursos en línea para mejorar la comprensión y aplicación de esta metodología.
- b. Uso de *software* especializado: Aprovechar las herramientas y software diseñados específicamente para el análisis RASCH, muchos de los cuales incluyen guías y soporte para facilitar su uso.
- c. Estrategias de recolección de datos: Implementar estrategias robustas para la recolección de datos, como el seguimiento intensivo de los participantes y el diseño cuidadoso de los cuestionarios, para minimizar los datos faltantes o incompletos y asegurar la calidad de los datos.
- d. Colaboración multidisciplinaria: Trabajar en colaboración con profesionales estadísticos o metodólogos especializados en análisis Rasch puede ayudar a superar algunos de los desafíos metodológicos y de interpretación, asegurando la aplicación correcta del modelo.
- e. Consideración de modelos alternativos: En casos donde el análisis Rasch pueda no ser el más adecuado debido a las características de los datos, considerar el uso de otros modelos de la TRI que puedan ofrecer la flexibilidad necesaria o, incluso, el uso de modelos basados en la TCT. También sería adecuado la combinación de metodologías de la TRI y de la TCT para obtener resultados más robustos.

Conclusiones y perspectivas de futuro

El análisis Rasch proporciona una metodología valiosa para la evaluación de competencias y habilidades clínicas, así como para el desarrollo y validación de instrumentos de medición en enfermería. A través de su aplicación práctica, se pueden obtener mediciones precisas y objetivas, lo que facilita la toma de decisiones basada en evidencia en la educación y la práctica de enfermería. Los estudios de caso destacados ilustran el potencial del análisis Rasch para mejorar la calidad de la atención al paciente y la efectividad de los programas de formación en enfermería, subrayando su importancia como herramienta en la investigación y práctica de enfermería.

La aplicación del análisis Rasch en la enfermería es un campo rico en oportunidades para la investigación futura. Algunas direcciones prometedoras incluyen:

- Innovación en instrumentos de medición: Desarrollo de nuevos instrumentos de medición para áreas emergentes de práctica y cuidado en enfermería, utilizando el análisis Rasch para garantizar su validez y fiabilidad.
- Evaluación longitudinal de competencias: Aplicación del análisis Rasch en estudios longitudinales para evaluar el desarrollo de competencias a lo largo del tiempo, identificando factores que contribuyen al crecimiento profesional continuo.

- Integración de tecnologías digitales: Exploración de cómo las tecnologías digitales y los datos masivos pueden integrarse con el análisis Rasch para mejorar la recopilación y análisis de datos en la investigación y práctica de enfermería.
- Estudios comparativos internacionales: Utilización del análisis Rasch para facilitar estudios comparativos internacionales sobre la educación y práctica de enfermería, contribuyendo a la armonización de estándares y prácticas a nivel global.

Declaración de IA generativa y tecnologías asistidas por IA en el proceso de redacción

Durante la preparación de este trabajo los autores utilizaron «ChatGPT Plus/Academic Assistant Pro®» con el fin de mejorar/aumentar la calidad, la relevancia y el recuento preciso de palabras. Después de utilizar esta herramienta/servicio, se revisó y editó el contenido según fuera necesario, asumiendo los autores plena responsabilidad por el contenido de la publicación.

Conflicto de intereses

Los autores declaran que no tienen ningún conflicto de intereses en relación con este manuscrito.

Bibliografía

1. Matas A. Introducción al análisis de la Teoría de Respuesta al Ítem. Ediciones Aidesoc; 2010 [consultado 5 Jun 2024]. Disponible en: <http://hdl.handle.net/10630/4711>
2. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. Lippincott Williams & Wilkins; 2008.
3. Stolt M, Kottorp A, Suhonen R. The use and quality of reporting of Rasch analysis in nursing research: A methodological scoping review. *Int J Nurs Stud*. 2022;132:104244, <http://dx.doi.org/10.1016/j.ijnurstu.2022.104244>.
4. Bond TG, Fox CM. *Applying the Rasch model: Fundamental measurement in the human sciences*. 3rd ed. Nueva York: Routledge; 2015.
5. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research; 1960.
6. Meyer JP. *Applied Measurement with jMetrik*. New York, NY: Routledge; 2014.
7. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske Institut; 1980.
8. Lord FM. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum; 1980.
9. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes*. 2003;1:27, <http://dx.doi.org/10.1186/1477-7525-1-27>.
10. Lerdal A, Kottorp A, Gay C, Aouizerat BE, Lee KA, Miaskowski C. J Pain Symptom Manag. 2016;51:1002–12, <http://dx.doi.org/10.1016/j.jpainsymman.2015.12.331>.
11. Linacre JM. *Winsteps® Rasch measurement computer program User's Guide*. Version 5. 6. 0. Portland, Oregon: Winsteps.com; 2023.
12. Kline TJ. Modern test theory: Assumptions, equations, limitations, and item analyses. En: Kline TJ, editor. *Psychological Testing: A Practical Approach to Design and Evaluation*. CA: SAGE Publications Ltd, Thousand Oaks; 2005.
13. Prieto G, Delgado AR. Análisis de un test mediante el modelo de Rasch. *Psicothema*. 2003;15:94–100.
14. Rasch G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. En: Glegvad M, editor. *The Danish Yearbook of Philosophy*. Copenhagen: Munksgaard; 1977. p. 59–94.
15. Engelhard G Jr. *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York: Routledge; 2013.
16. Andrich D. *Rasch models for measurement*. Sage Publications; 1988.
17. Masters MN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–74.
18. Andrich D. Rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–73.
19. Polit DF, Beck CT. Developing and Testing Self-Report Scales. In *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 8th ed. Lippincott Williams & Wilkins: Tokyo; 2008. p. 474–505.
20. Linacre JM. Detecting multidimensionality: Which residual datatype works best? *J Outcome Meas*. 1998;2, 266e83.
21. Linacre JM. Rasch measures and Unidimensionality. *Rasch Meas Trans*. 2011;24:1310.
22. Ferrando PJ, Lorenzo-Seva U. Assessing the Quality and Appropriateness of Factor Solutions and Factor Score Estimates in Exploratory Item Factor Analysis. *Educ Psychol Meas*. 2018;78:762–80, <http://dx.doi.org/10.1177/0013164417719308>.
23. Wright BD, Masters GN. *Rating Scale Analysis: Rasch Measurement*. Chicago: Mesa Press; 1982.
24. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodol*. 2008;8:33, <http://dx.doi.org/10.1186/1471-2288-8-33>.
25. Fisher W. Reliability, separation, strata statistics. *Rasch Meas Trans*. 1992;6:238.
26. Yen WM. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Appl Psychol Meas*. 1984;8:125–45, <http://dx.doi.org/10.1177/014662168400800201>.
27. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas*. 2017;41:178–94, <http://dx.doi.org/10.1177/0146621616677520>.
28. Linacre JM. *Facets computer program for many-facet Rasch measurement, version 3.87.0*. Beaverton, Oregon: Winsteps.com; 2024.
29. Parra-Anguita L, Sánchez-García I, del Pino-Casado R, Pancorbo-Hidalgo PL. Measuring knowledge of Alzheimer's: Development and psychometric testing of the UJA Alzheimer's Care Scale. *BMC Geriatrics*. 2019;19:63, <http://dx.doi.org/10.1186/s12877-019-1086-2>.
30. Dabaghi S, Esmailzadeh F, Rohani C. Application of Rasch Analysis for Development and Psychometric Properties of Adolescents' Quality of Life Instruments: A Systematic Review. *Adolesc Health Med Ther*. 2020;11:173–97.
31. López-Franco MD, Parra-Anguita L, Comino-Sanz IM, Pancorbo-Hidalgo PL. Development and Psychometric Properties of the Pressure Injury Prevention Knowledge Questionnaire in Spanish Nurses. *Int J Environ Res Public Health*. 2020;17:3063, <http://dx.doi.org/10.3390/ijerph17093063>.
32. Pérez-López C, López-Franco MD, Comino-Sanz IM, Pancorbo-Hidalgo PL. Validación del cuestionario de Conocimientos sobre prevención de lesiones por presión en estudiantes de Enfermería: análisis Rasch. *Enferm Clin*. 2021;31:12–20, <http://dx.doi.org/10.1016/j.enfcli.2020.07.001>.

33. Seguí MM, Cabrero-García J, Crespo A, Verdú J, Ronda E. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *J Clin Epidemiol.* 2015;68:662–73, <http://dx.doi.org/10.1016/j.jclinepi.2015.01.015>.
34. Higgins J, Laramée MT, Harrison KR, Delparte JJ, Scoville CY, Flett HM, et al. The Spinal Cord Injury Pressure Ulcer Scale (SCIPUS): An assessment of validity using Rasch analysis. *Spinal Cord.* 2019;57:874–80, <http://dx.doi.org/10.1038/s41393-019-0287-z>.
35. Fischer L, Rohm T, Carstensen CH, Gnambs T. Linking of Rasch-Scaled Tests: Consequences of Limited Item Pools and Model Misfit. *Front. Psychol.* 12:633896. DOI: 10.3389/fpsyg.2021.633896.