

Análisis psicométrico con el modelo de Rasch de la escala Health of the Nation Outcome Scales (HoNOS) en población uruguaya

Psychometric analysis with the Rasch model of the Health of the Nation Outcome Scales (HoNOS) in Uruguayan population



Denisse Dogmanas
Gerardo Prieto

Psychology

ID: 10.33881/2027-1786.RIP.14106

Title: Psychometric analysis with the Rasch model of the Health of the Nation Outcome Scales (HoNOS) in Uruguayan population

Título: Análisis psicométrico con el modelo de Rasch de la escala Health of the Nation Outcome Scales (HoNOS) en población uruguaya

Alt Title / Título alternativo:

[en]: Psychometric analysis with the Rasch model of the Health of the Nation Outcome Scales (HoNOS) in Uruguayan population

[es]: Análisis psicométrico con el modelo de Rasch de la escala Health of the Nation Outcome Scales (HoNOS) en población uruguaya

Author (s) / Autor (es):

Dogmanas & Prieto

Keywords / Palabras Clave:

[en]: Health of the Nation Outcome Scales; Mental Health Services; Rasch Analysis

[es]: Health of the Nation Outcome Scales; Servicios de Salud Mental; Análisis de Rasch;

Proyecto / Project:

En el marco de una tesis doctoral. Nombre "Adaptación y Validación en Población Uruguaya de la Escala HoNOS de Resultados en Servicios de Salud Mental"

Submitted: 2020-08-27

Accepted: 2020-11-08

Resumen

El objetivo principal del trabajo fue analizar las propiedades psicométricas de The Health of the Nation Outcome Scales (HoNOS) utilizando la Teoría Clásica de los Tests y el Modelo de las Escalas de Calificación, un modelo tipo Rasch para ítems politómicos. La HoNOS es una de las escalas más utilizadas para evaluar resultados de forma rutinaria en los servicios de salud mental. La versión original para adultos ha sido traducida al castellano (Uriarte et al., 1999) pero hasta el momento no se han realizado estudios empíricos sobre las propiedades psicométricas de esta versión. Participaron del estudio 402 usuarios de cuatro servicios de salud mental ambulatorios y del hospital psiquiátrico de Uruguay. Se analizó el sistema de categorías de respuesta, el ajuste de las personas e ítems al modelo de Rasch, la fiabilidad, validez discriminante y la sensibilidad al cambio.

Los resultados indicaron que el sistema original de cinco categorías de la HoNOS no funcionó correctamente de acuerdo con las recomendaciones de Linacre (2002). Se buscó una solución mediante la agregación de categorías adyacentes reduciendo las categorías iniciales a cuatro que mostró un mejor funcionamiento. Se observó un aceptable ajuste del instrumento a los requisitos del modelo. La escala mostró adecuada validez para discriminar a los pacientes con diferentes niveles de severidad y sensibilidad al cambio. Se concluye que la HoNOS cumple con algunas características deseables como medida de evaluación de resultados de los servicios de salud mental.

Presenta algunos problemas métricos que pueden verse mejorados reduciendo a cuatro las categorías de respuesta.

Citar como:

Dogmanas, D., & Prieto, G. (2021). Análisis psicométrico con el modelo de Rasch de la escala Health of the Nation Outcome Scales (HoNOS) en población uruguaya. *Revista Iberoamericana de Psicología*, 14 (1), 59-69. Obtenido de: <https://reviberopsicologia.ibero.edu.co/article/view/1941>

Abstract

The aim of the study was to analyze the psychometric properties of The Health of the Nation Outcome Scales (HoNOS) using the Classical Test Theory and the Rating Scale Model, a Rasch-type model used for polytomous items. The HoNOS is one of the most widely used scales to assess outcomes routinely in mental health services. The original version for adults has been translated into Spanish (Uriarte et al., 1999) but no empirical studies have been carried out on the psychometric properties of this version. Participants were 402 users of four outpatient mental health services and the psychiatric hospital in Uruguay. The response category system, the fit of people and items to the Rasch model, reliability, discriminant validity and sensitivity to change were analyzed. The results showed that the original HoNOS five-category system did not function correctly according to the Linacre (2002) recommendations. A solution was sought by adding adjacent categories, reducing the initial categories to four that showed better performance. An acceptable fit of the instrument to the model requirements was observed. The scale showed adequate validity to discriminate patients with different levels of severity and sensitivity to change. It is concluded that the HoNOS meets some desirable characteristics as a measure for evaluating the outcomes of mental health services. The original HoNOS presents some metric problems that can be improved by reducing the response categories to four.

Dra. Denisse **Dogmanas**

ORCID: [0000-0002-5922-9219](https://orcid.org/0000-0002-5922-9219)

Source | Filiación:

Universidad de la República, Uruguay

BIO:

Doctora en Psicología Clínica y de la Salud de la Universidad de Salamanca, España. Profesora Adjunta del Instituto de Psicología Clínica de la Facultad de Psicología de la Universidad de la República, Uruguay.

City | Ciudad:

Montevideo [uy]

Dr. Gerardo **Prieto**

AutorID: [55543476000](https://orcid.org/55543476000)

Research ID: [F-3397-2012](https://orcid.org/F-3397-2012)

ORCID: [0000-0001-8187-7784](https://orcid.org/0000-0001-8187-7784)

Source | Filiación:

Universidad de Salamanca, España

BIO:

Profesor en la Facultad de Psicología de la Universidad de Salamanca desde 1974; Catedrático de Psicometría 1987-2018; Profesor Emerito desde 2018.

City | Ciudad:

Salamanca [es]

Análisis psicométrico con el modelo de Rasch de la escala Health of the Nation Outcome Scales (HoNOS) en población uruguaya

Psychometric analysis with the Rasch model of the Health of the Nation Outcome Scales (HoNOS) in Uruguayan population

Denisse **Dogmanas**
Gerardo **Prieto**

La evaluación de la calidad de los servicios de salud mental es un prerrequisito para que los servicios puedan mejorar (**Trauer, 2010; Kilbourne, 2018**). El monitoreo rutinario de resultados de los servicios de salud mental es necesario para identificar las necesidades de atención en salud mental de las personas, las intervenciones más efectivas y optimizar así los recursos existentes.

The Health of the Nation Outcome Scales (HoNOS) fue desarrollada en la década del 90 en Gran Bretaña por la Unidad de Investigación de The Royal College of Psychiatrists' Research Unit (**Wing, Curtis y Beevor, 1996**) con el propósito de ser un instrumento para la medición rutinaria de los resultados en personas con trastornos mentales severos incluyendo aspectos biológicos, psicológicos y sociales (**Wing et al., 1996; Pirkis et al., 2005**). Ha sido diseñada para su uso en personas con una amplia gama de diagnósticos (**Andreas et al., 2011**) y en diferentes servicios de salud mental (**Delaffon, Anwar, Noushad, Ahmed y Brugha, 2012**).

Cumple con varios requisitos que se consideran claves para un instrumento que ha de ser aplicado de manera rutinaria por los servicios de salud dado que es breve, comprensivo, fácil de aplicar, útil (**Pirkis et al., 2005**) y su uso no implica un costo económico. Distintas revisiones la han identificado como candidata adecuada para la evaluación rutinaria de los servicios de salud mental (**Harris et al., 2018**), sirviendo además para comparar el desempeño de los servicios y el costo-efectividad (**Slade et al., 2010**).

La HoNOS original se desarrolló para los servicios de salud mental de adultos, de 18 a 64 años inclusive. Posteriormente, se desarrollaron otras versiones: para niños y adolescentes (HoNOSCA) (Gowers et al., 1999); adultos mayores (HoNOS 65+) (Burns, Beevor, Lelliott y Wing, 1999) y otras poblaciones. La versión original de la HoNOS es utilizada como parte de la recolección rutinaria y obligatoria de datos de servicios públicos de salud mental de Reino Unido, Nueva Zelanda y Australia (Coombs, Stapley y Pirkis, 2011) y en algunas ciudades o contextos en otros países como Alemania, Holanda, Canadá (Harris et al., 2018). Además, ha sido traducida a diferentes idiomas incluyendo Alemán, Danés, Español, Griego, Italiano, Lituano, Noruego, Francés y Tailandés (Harris et al., 2018).

Mediante la Teoría Clásica de los Test (TCT), Pirkis et al. (2005) realizaron una revisión de los estudios de validación de la HoNOS, concluyendo que muestra adecuada validez concurrente, predictiva y de contenido, adecuada fiabilidad y sensibilidad al cambio.

Los datos sobre la estructura interna de la HoNOS no son concluyentes, ya que se han sugerido diferentes estructuras factoriales. Algunos estudios apoyan una estructura de cuatro factores que corresponde a las subescalas originales de HoNOS (Pirkis et al., 2005) mientras que otros sugieren otras estructuras (Speak y Muncer, 2016).

Estudios realizados más recientemente, han utilizado el modelo de Rasch para evaluar las propiedades psicométricas de la escala. Rasch es un modelo de medida alternativo con potenciales ventajas sobre la TCT, principalmente por la posibilidad de medir en una única dimensión en la que se sitúan conjuntamente las personas y los ítems (Wright y Stone, 1979; Prieto, Delgado, Perea y Ladera, 2010). Utilizando el análisis de Rasch en una muestra de daneses, Palic, Kappel y Makransky (2016) encontraron una versión de la HoNOS unidimensional de 10 ítems, contradiciendo los hallazgos previos de Lovalgio y Monzani (2012) en Italia, que no apoyaron una estructura unidimensional.

Uriarte et al. (1999) han llevado a cabo la traducción oficial al castellano de la HoNOS. También han sido traducidas la HoNOSCA para niños y adolescentes (Ballesteros-Urpí, 2018) y la HoNOS65+ para personas mayores de 65 años (Ausín, Muñoz, Quiroga, 2007). La HoNOSCA y HoNOS65+ cuentan con estudios de validación al castellano. Sin embargo, hasta el momento no se han realizado estudios empíricos sobre las propiedades psicométricas de la versión original de la HoNOS para adultos.

El presente estudio se propuso evaluar las propiedades psicométricas de la HoNOS para adultos en su versión al castellano a partir de la TCT y el modelo de Rasch (MR). De forma específica se pretendió analizar el sistema de categorías de respuesta, la dimensionalidad de la escala, el ajuste de las personas e ítems al modelo de Rasch, la fiabilidad, la validez discriminante y la sensibilidad al cambio.

El modelo de Rasch permite realizar un análisis más sofisticado que la TCT (Prieto et al., 2010; Delgado, Prieto y Burin, 2017). En el MR la probabilidad de que el sujeto n domine (acierte, asevere, etc) el ítem i se modela como $P_{ni} = \exp(B_n - D_i) / (1 + \exp(B_n - D_i))$, donde B_n es el nivel de persona y D_i es la ubicación del ítem. Una de las principales ventajas respecto de la TCT se deriva del hecho de que es un modelo de medición conjunta: si los datos empíricos se ajustan adecuadamente al modelo, entonces las medidas de las personas y los valores de los ítems pueden ubicarse conjuntamente en una escala de intervalos cuya unidad es el logit. Además, el modelo de Rasch posee la propiedad de objetividad específica permitiendo la separación algebraica de los parámetros de los ítems y de las personas (el parámetro de la per-

sona puede ser eliminado cuando se estiman los parámetros de los ítems; puede eliminarse asimismo el parámetro del ítem para estimar los parámetros de las personas). Por tanto, las estimaciones de los parámetros de los ítems son independientes de la muestra de personas y las estimaciones de los parámetros de las personas son independientes de los ítems concretos que se han utilizado; esto no ocurre en el modelo de medición clásico. La medición invariante es una de las principales propiedades del modelo. A diferencia de la TCT en la que la dificultad de los ítems (proporción de acertantes) puede variar sensiblemente entre muestras de personas con distinto nivel en el atributo, los parámetros de localización de los ítems son invariantes a la muestra de personas; asimismo, los parámetros de los sujetos son invariantes a la muestra de ítems empleada para estimarlos.

Método

Participantes

Los estudios fueron llevados a cabo en cuatro servicios ambulatorios, y el centro de internación psiquiátrica de referencia nacional de la Administración de Servicios de Salud del Estado (ASSE) -principal prestador de salud del sector público de Uruguay- en Montevideo, la capital del país.

La muestra estuvo integrada por 402 usuarios de los servicios de salud mental; 285 (71%) se encontraban recibiendo tratamiento en atención ambulatoria y 117 (29%) en régimen de internación psiquiátrica. La edad media del grupo ambulatorio fue de 44.42 años (SD = 15.27) y la del grupo internado de 39.28 (SD = 13.83) no resultando la diferencia estadísticamente significativa ($p > .05$). En el grupo ambulatorio el diagnóstico principal más frecuente fue el Trastorno del estado de ánimo (33%), mientras que en la mayoría de los usuarios del grupo de internados el más frecuente fue la Esquizofrenia y trastornos delirantes (54.7%).

En la recolección de datos participaron 22 clínicos: 14 psiquiatras y 8 psicólogos.

Instrumentos

Se utilizó la versión traducida al castellano de la escala HoNOS (Uriarte et al., 1999). La HoNOS es una escala de 12 ítems que evalúan síntomas psicológicos, problemas del comportamiento, orgánicos y sociales (Wing et al., 1996). Cada ítem es puntuado en una escala de cinco opciones de respuesta: 0 (sin problema), 1 (problema menor que no requiere intervención), 2 (problema leve pero claramente presente), 3 (problema de moderada gravedad), 4 (problema grave o muy grave) (Uriarte et al., 1999). Para puntuar la HoNOS el profesional debe considerar los problemas más severos ocurridos durante las últimas dos semanas previas a la evaluación, excepto en los ítems 10-12, para los que puede considerarse un período más extenso.

Procedimiento

Antes de iniciar los estudios se solicitó y obtuvo autorización para la utilización de la Escala HoNOS al Royal College of Psychiatrist (RCP) de UK, y al autor de la versión oficial al castellano.

El proyecto de investigación fue aprobado por el Comité de ética en investigación de la Facultad de Psicología de la Universidad de la República del Uruguay.

Previamente al estudio de validación, se realizó un estudio piloto con el objetivo de someter al juicio de expertos la adecuación de la HoNOS para Uruguay. El grupo de expertos (n=8) integrado por psiquiatras y psicólogos con experiencia de trabajo en servicios de salud con pacientes con trastorno mental consideraron que la escala era apropiada. Propusieron dos modificaciones menores en lo que refiere a una adaptación lingüística para Uruguay de dos palabras en dos ítems del glosario.

La HoNOS fue completada por los profesionales de los centros participantes como parte de la consulta rutinaria con sus pacientes, considerando que la HoNOS fue desarrollada para ser utilizada por los propios clínicos en su práctica habitual en los servicios de salud.

Para el estudio de sensibilidad al cambio, en el caso de los pacientes ambulatorios; los datos fueron completados al inicio del tratamiento y en la consulta de control y en los pacientes internados; al ingreso y al egreso de la internación.

Antes de iniciar el estudio, los profesionales recibieron una capacitación de una duración aproximada de cuatro horas, diseñada a partir de los materiales de apoyo enviados por el RCP y de acuerdo a las sugerencias del manual para el entrenador de la Escala HoNOS (Wing et al., 1996).

Los datos para el estudio fueron obtenidos previamente anonimizados por los profesionales.

Análisis Estadístico y Psicométrico

Los datos recogidos fueron introducidos en una base de datos Excel y posteriormente en el paquete estadístico SPSS 15, que fue utilizado para el procesamiento de datos desde la TCT. El análisis Rasch de las puntuaciones de la HoNOS se realizó mediante el software WINSTEPS 3.74.0 (Linacre y Wright, 2011).

El modelo propuesto por Rasch (1960) se usa cuando los datos son dicotómicos: la persona puede ser calificada usando dos categorías de respuesta (sí/no, acierto/error, acuerdo/desacuerdo, etc). Para analizar ítems politómicos (más de dos categorías), como los de HoNOS, se empleó el Modelo de Escalas de Calificación (MEC) (Andrich, 1978), un modelo de tipo Rasch que permite analizar empíricamente la calidad métrica de las categorías de respuesta. La formulación de este modelo ($\log(P_{nik}/P_{ni(k-1)}) = B_n - D_i - F_k$) indica que el logaritmo del cociente entre la probabilidad de que una persona n sea calificada con la categoría k en el ítem i (P_{nik}) y la probabilidad de que sea calificada con la categoría inferior ($P_{ni(k-1)}$) es igual a la diferencia en la variable entre el parámetro de la persona (B_n), la localización del ítem (D_i) y el valor en logit del punto en el que las categorías k y $k-1$ son equiprobables (F_k). F_k suele denominarse paso o umbral. El número de pasos equivale al número de categorías menos uno. El modelo asume que los pasos entre las categorías adyacentes son iguales en todos los ítems.

Análisis de las categorías

Se evaluó la funcionalidad métrica de las categorías de respuesta mediante los criterios de Linacre (2002): a) un mínimo de 10 respuestas por categoría; b) distribución regular de las respuestas entre las categorías (uniforme o unimodal); c) incremento monótono de las medias de las categorías (las categorías superiores corresponden a los mayores niveles en el rasgo latente); d) inexistencia de categorías desajustadas severamente ($\text{Outfit} < 2.0$) y e) pasos entre las categorías adyacentes ordenados monótonicamente. La presencia de umbrales desordenados sugiere que las categorías no son claramente diferenciables por los calificadores (Dzingina, Higginson, McCrone y Murtagh, 2017). Si las categorías no cumplen los criterios, se aconseja agrupar categorías adyacentes y verificar la funcionalidad de las nuevas categorías (Colvin y Gorgun, 2020).

Unidimensionalidad

Se analizó el ajuste de los datos al supuesto de unidimensionalidad. Para esto se utilizaron dos procedimientos: el Análisis Paralelo (Horn, 1965) en el contexto del Análisis Factorial Exploratorio y el Análisis de Componentes Principales de los Residuos en el marco del modelo de Rasch. Para realizar el Análisis Paralelo (AP) se utilizó el programa Factor 7.2 (Lorenzo-Seva y Ferrando, 2012). Para realizar el AP con FACTOR se siguió el procedimiento sugerido por Ferrando y Anguiano-Carrasco (2010). En el marco del MR, se inspecciona la unidimensionalidad mediante el Análisis de Componentes Principales de los residuos (diferencias entre los valores observados y los predichos por el modelo). Puede considerarse que los datos son esencialmente unidimensionales si la medida estimada con el modelo de Rasch da cuenta de un porcentaje de varianza moderadamente alto, en opinión de Reckase (1979) al menos un 20%, y además el valor propio del primer componente de los residuos es inferior a 3 (Miguel, Silva y Prieto, 2013).

Ajuste

Smith, Rush, Fallowfield, Velikova y Sharpe (2008) analizaron la relación entre el tamaño de la muestra y los estadísticos de ajuste al modelo. Estos autores concluyeron que los estadísticos outfit e infit (respectivamente la media no ponderada y ponderada con su varianza de los cuadrados de los residuos) son relativamente independientes del tamaño de la muestra en datos politómicos y que son los estadísticos de ajuste más usados en la investigación en ciencias de la salud. Por ello, se cuantificó el ajuste de las personas e ítems al modelo de Rasch usando los estadísticos Outfit (la media no ponderada de los residuos estandarizados al cuadrado) e Infit (la media de los residuos estandarizados al cuadrado ponderados con su valor de información). Se han propuesto distintos rangos de valores para detectar el desajuste (Bond y Fox, 2015). En nuestro caso, hemos optado por utilizar los criterios propuestos por Linacre (2011) debido a su frecuente uso. Cuando los valores de estos indicadores oscilan entre 0.5 y 1.5 se considera que los datos tienen un ajuste adecuado al modelo; entre 1.5 y 2.0, el desajuste es considerado moderado y sin graves consecuencias para las medidas, mientras que los valores mayores a 2.0 indican un desajuste severo.

Fiabilidad

A diferencia de la TCT, el modelo de Rasch aporta un estadístico de precisión, denominado error estándar (SE), para las estimaciones del parámetro de cada persona y de cada ítem, siendo mayor la precisión de las medidas cuanto menor sea SE. A partir del promedio de los cuadrados de los errores estándar (SE^2), se pueden obtener coeficientes de fiabilidad para las medidas de las personas ($R_p = (SD_p^2 - SE_p^2) / SD_p^2$) y de los ítems ($R_i = (SD_i^2 - SE_i^2) / SD_i^2$). R_p y R_i indican respectivamente la proporción de la variabilidad de las personas y de los ítems que es reproducible a partir del modelo, mostrando el grado en el que se puede diferenciar entre la habilidad de las personas y entre la dificultad de los ítems en la variable (Bond y Fox, 2015). Ambos estadísticos pueden variar entre 0 y 1. Los valores $R_p > 0.70$ y $R_i > 0.90$ son indicadores de una fiabilidad aceptable.

Validación de criterio

Se obtuvieron dos evidencias criterio. Por un lado, la eficiencia de HoNOS para discriminar entre grupos con distinto nivel de gravedad y por otro, su sensibilidad para detectar presuntas mejoras en la severidad del trastorno mental.

En el primer caso, se analizaron las diferencias entre el grupo atendido en los servicios de salud mental ambulatorios y el grupo internado en el hospital psiquiátrico. El tipo de servicio en el que están siendo atendidas las personas es una información externa asociada a la severidad de los problemas de salud mental de los pacientes: los servicios ambulatorios atienden a personas que presentan menores niveles de severidad, mientras que los de internación a personas con niveles más altos de severidad. De acuerdo al criterio que plantea que la mayor severidad de los problemas se corresponderá con una mayor puntuación en HoNOS, se considera que una diferencia significativa y moderadamente alta entre las medias de ambos grupos será una evidencia favorable a la validez criterial.

El análisis de la sensibilidad al cambio de HoNOS se realizó, tanto con las puntuaciones clásicas como con las puntuaciones Rasch analizando la diferencia entre las medias de las puntuaciones obtenidas en la primera evaluación (pre-tratamiento) y en la segunda evaluación (post-tratamiento). En el caso de los pacientes ambulatorios; al inicio y en la consulta de control. En el caso de los pacientes internados; al

ingreso y al egreso de la internación. Para contrastar la significación del cambio entre las medias de las puntuaciones clásicas del pre-tratamiento (pre) y el post-tratamiento (post), se utilizó la prueba t de Student para muestras apareadas. Asimismo, se calculó el tamaño del efecto (d de Cohen) para valorar la magnitud de la diferencia. Para realizar el análisis del cambio entre las medias de las puntuaciones logit en el pre y el post se requiere un marco de medición estable en ambas ocasiones; es necesario realizar un escalamiento de ambas escalas en la misma métrica, debido a que en las calibraciones separadas del pre y el post el punto cero de cada escala se sitúa por defecto en la dificultad media de los ítems (Wright, 1996). El procedimiento consistió en calibrar en primer lugar los datos por separado, calculando la dificultad de los ítems (D_i) en cada una de las aplicaciones. Después se contrastó la invarianza entre ambas aplicaciones de los parámetros de dificultad de los ítems y se estimaron las puntuaciones en el post fijando los parámetros de dificultad de los ítems invariantes (Prieto y Velasco, 2010). La significación de la diferencia entre las medias de las puntuaciones logit pre y post se contrastó con el estadístico t de Welch. Asimismo, se calculó el tamaño del efecto (d de Cohen) para valorar la magnitud de la diferencia.

Resultados

El análisis de las categorías de respuesta de la HoNOS mediante el MEC mostró que se cumplen la mayoría de las recomendaciones de Linacre (2002). Sin embargo, los pasos entre las categorías sucesivas no están ordenados monotónicamente (Tabla 1). Esto significa que no todas las categorías de respuesta son funcionales. De acuerdo con Linacre (2002) las categorías han de ser modales (la elección de cada categoría ha de ser la más probable en algún rango de la variable). En la Figura 1 puede observarse que la curva correspondiente a la categoría 1 está muy aplanada, lo que implica que la probabilidad de que sea utilizada es mucho más pequeña que la del resto de las categorías. En consecuencia, las categorías debieron ser redefinidas mediante una agrupación de categorías adyacentes, lo que llevó a reducir las cinco categorías originales a cuatro. La agrupación se realizó de la siguiente forma: 0 = sin problema (0), 1 = problema leve (1+2), 2 = problema de moderada gravedad (3) y 3 = problema grave o muy grave (4) (Figura 1). Puede observarse que el sistema de cuatro categorías presenta una funcionalidad óptima. En consecuencia, el resto de los análisis se llevó a cabo con el sistema reducido de categorías de respuesta.

Tabla 1 Análisis de las categorías de respuesta originales (C=5) y agrupadas (C=4) de HoNOS

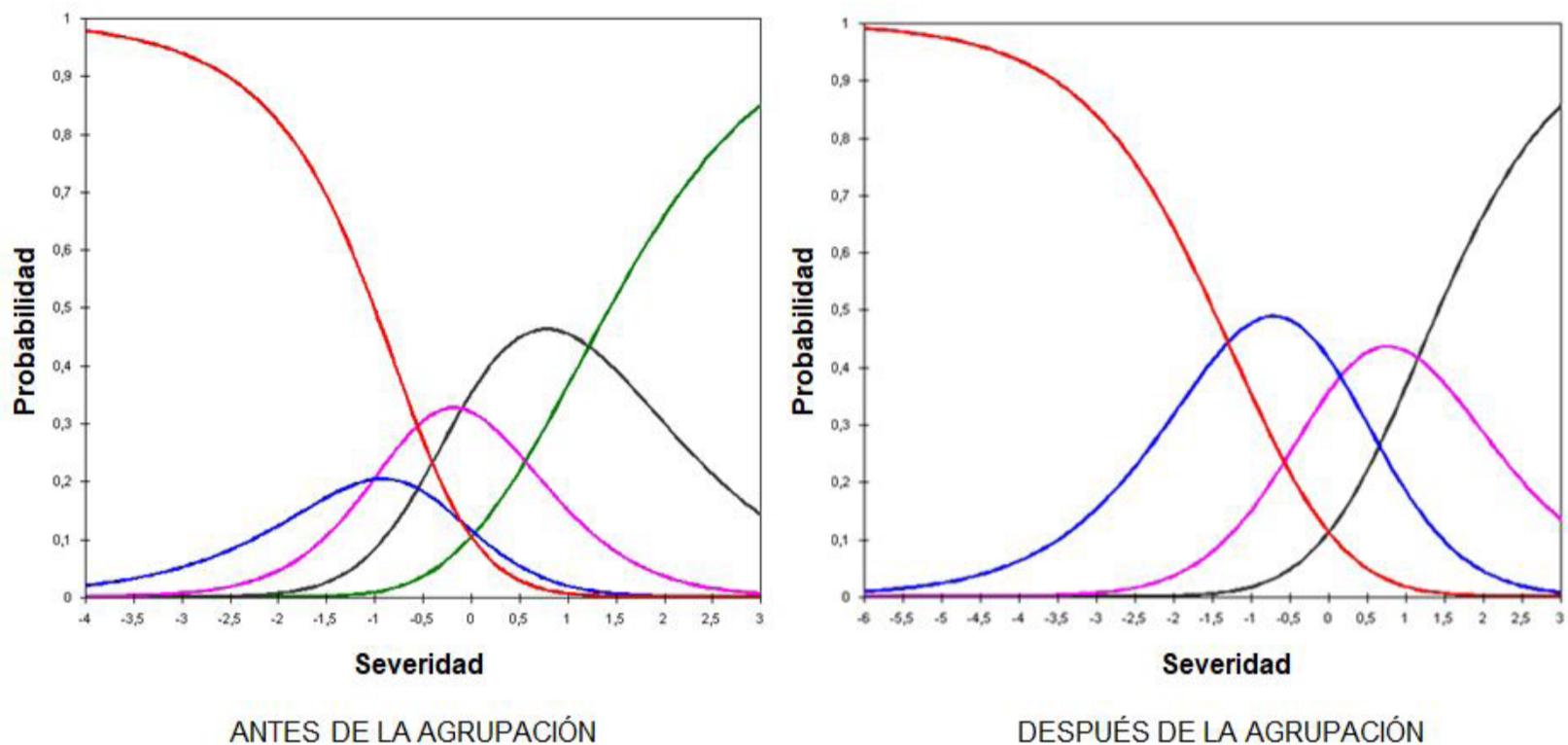
| Categoría de respuesta | Observaciones (número y %) | Promedio de categorías | Infit | Outfit | Pasos o umbrales |
|------------------------|----------------------------|------------------------|-------|--------|------------------|
| Primer análisis | | | | | |
| 0 | 2528 (52%) | -1.51 | .97 | .99 | -- |
| 1 | 777 (16%) | -1.00 | .93 | .69 | -.11 |
| 2 | 862 (18%) | -.70 | .87 | .82 | -1.01 |
| 3 | 535 (11%) | -.47 | 1.10 | 1.41 | -.10 |
| 4 | 122 (3%) | -.23 | 1.18 | 1.29 | 1.22 |
| Segundo análisis | | | | | |
| 0 | 2528 (52%) | -2.20 | .96 | .97 | -- |
| 1 | 1639 (34%) | -1.21 | .85 | .73 | -1.30 |
| 2 | 535 (11%) | -.72 | 1.10 | 1.36 | .15 |
| 3 | 122 (3%) | -.24 | 1.18 | 1.35 | 1.15 |

Fuente: elaboración propia

En el contexto del AFE, el Análisis Paralelo (AP) revela que existe una única dimensión básica, dado que el porcentaje de la varianza explicado por el autovalor del primer factor empírico (32.1%) fue superior al porcentaje de la varianza explicado por el promedio de los autovalores

aleatorios correspondientes a la primera dimensión (16.9%). De hecho, el porcentaje correspondiente al segundo factor empírico (14.5%) fue menor que el porcentaje correspondiente al autovalor del segundo factor aleatorio (15.1%).

Figura 1. Curvas características de las categorías de respuesta de la HoNOS original y con categorías agrupadas



En lo que se refiere a la inspección de la unidimensionalidad de las medidas mediante el ACP de los residuos, se observó que el porcentaje de la varianza de las observaciones explicada por la dimensión extraída por el modelo de Rasch es moderadamente alto (32.2%) y el autovalor correspondiente al primer componente de los residuos es inferior a 3 (2.1). En consecuencia, los datos derivados del AP y del ACP de los residuos permiten concluir que las medidas obtenidas con la HoNOS son fundamentalmente unidimensionales.

En la tabla 2 se observa que el promedio de los estadísticos de ajuste de las personas es razonable: los valores de la media y de la des-

viación estándar son respectivamente de 1.0 y .54 (Infit) y 1.0 y .66 (Outfit). La mayoría de las personas se ajustaron al modelo: 44 personas (10.9%) presentan un desajuste moderado (infit y/o outfit > 1.5 y < 2) y 37 personas (9.2%) alto desajuste (infit y/o outfit > 2).

Los errores típicos de medida de las personas van de .36 a 1.84. La fiabilidad global de las personas ($R_p = .66$, α de Cronbach = .69), resulta levemente inferior al valor recomendado para una fiabilidad aceptable (.70).

Tabla 2 Estadísticos descriptivos de los valores de las personas

| Estadístico | X | B | Infit | Outfit |
|-------------|-----|-------|-------|--------|
| M | 7.6 | -1.67 | 1.0 | 1.0 |
| SD | 4.2 | .88 | .54 | .66 |
| Min | 0 | -5.18 | .16 | .20 |
| Máx | 24 | .82 | 3.0 | 4.5 |
| α | .69 | -- | -- | -- |
| PSR | -- | .66 | -- | -- |

Nota. X = Puntuación clásica. B = valores Rasch (logits). M = Media. SD = desviación estándar. Min = valor mínimo. Max = valor máximo. α = alpha de Cronbach. PSR = Person Separation Reliability (fiabilidad Rasch).

Fuente: elaboración propia

En relación al ajuste de los ítems al modelo de Rasch (Tabla 3), pese a que tres ítems presentan un desajuste moderado con valores de Infit y/o Outfit superiores a 1.5 (ítem 3: consumo problemático de alcohol o drogas; ítem 5: enfermedad física; e ítem 6: ideas delirantes y alucinaciones); ningún ítem presenta un desajuste severo con valores superiores a 2.00.

La mayoría de los ítems mostró una adecuada correlación ítem-test ($r_{iX} > .30$), excepto los ítems 5 (enfermedad física) y 7 (problemas en relación con el humor depresivo) que presentaron correlaciones de .12 y .24 respectivamente.

Tabla 3 Dificultad, error estándar, estadísticos de ajuste y correlaciones ítems-escala (r_{ix}) de los ítems de HoNOS

| Ítem | Dificultad (en logit) | Error estándar (en logit) | Infit | Outfit | (r_{ix}) |
|---|-----------------------|---------------------------|-------|--------|--------------|
| 1. Conducta hiperactiva, agresiva, disruptiva o agitada | .25 | .09 | 1.00 | .78 | .46 |
| 2. Autolesiones no accidentales | 1.29 | .14 | 1.35 | .79 | .33 |
| 3. Consumo problemático de alcohol o drogas | .66 | .10 | 1.37 | 1.27 | .32 |
| 4. Problemas cognitivos | .40 | .09 | 1.01 | 1.07 | .38 |
| 5. Problemas por enfermedad física o discapacidad | .10 | .08 | 1.64 | 1.76 | .16 |
| 6. Problemas asociados a la presencia de ideas delirantes y alucinaciones | .20 | .08 | 1.46 | 1.31 | .36 |
| 7. Problemas en relación con el humor depresivo | -.44 | .07 | 1.28 | 1.34 | .27 |
| 8. Otros problemas mentales o conductuales | -.81 | .06 | 1.17 | 1.21 | .41 |
| 9. Problemas con las relaciones | -.91 | .06 | .53 | .55 | .66 |
| 10. Problemas con las actividades de la vida cotidiana | -.40 | .07 | .53 | .52 | .68 |
| 11. Problemas con las condiciones de vida | -.10 | .07 | .88 | .81 | .59 |
| 12. Problemas en relación con la ocupación y las actividades | -.23 | .07 | .75 | .70 | .62 |

Fuente: elaboración propia

Con el análisis Rasch de la precisión de las medidas de los ítems, se observan valores apropiados: los errores típicos de medida de los ítems oscilan entre .07 a .12, y la fiabilidad global evaluada mediante el estadístico Ri es muy alta (.99).

En la figura 2 se muestra la distribución de las personas e ítems en la que puede observarse que la media de la localización de las personas (-1.67) es muy inferior a la de los ítems (.00). Además, la mayor parte de las personas se encuentran por debajo del rango de la variable en el que los ítems se sitúan; el nivel de severidad de las personas de la muestra es muy bajo. Se observa que el ítem 2 resulta el indicador de mayor nivel de gravedad de la HoNOS y el ítem 9 el de menor nivel.

El contraste de diferencia de medias del grupo ambulatorio y del grupo internado (Tabla 4) muestra que los pacientes internados presentan puntuaciones más altas que los del grupo ambulatorio y las diferencias son estadísticamente significativas en las puntuaciones Rasch y en las puntuaciones clásicas. Con los valores Rasch se observó un tamaño del efecto grande (d de Cohen = .82) y con las puntuaciones clásicas mediano (d de Cohen de .73). La diferencia entre las puntuaciones de severidad de ambos grupos es un indicio de la validez predictiva y de la utilidad diagnóstica de HoNOS.

La precisión de la medida de las personas es mayor en el grupo internado (Person Separation Reliability = .75, α de Cronbach = .71) que en el grupo ambulatorio (Person Separation Reliability = .54, α de Cronbach = .61).

Sensibilidad al cambio

Se analizaron 75 personas de la muestra total que presentaron dos evaluaciones (pre-tratamiento y post-tratamiento): 24 hombres y 51 mujeres con una media de edad de 38.21 (SD = 12.00); 39 pertenecían al grupo ambulatorio y 36 al grupo internado.

La diferencia entre las medias de las puntuaciones directas en el pre-tratamiento (M=9.36, DT=4.53) y el post-tratamiento (M=6.2, SD= 3.39) muestran que, con la puntuación clásica se observa un cambio en los pacientes en la puntuación media total de HoNOS que resultó estadísticamente significativa (t de Student= 7.66, p < .001) y el tamaño del efecto de esta diferencia es mediano (d de Cohen = .79). De igual modo, con el análisis Rasch la diferencia en la puntuación entre el pre-tratamiento (-1.18, SD=.79) y post-tratamiento (-2.22, SD= 1.10) resultó estadísticamente significativa (t=10.56, p< .001), siendo grande el tamaño del efecto correspondiente (d de Cohen = 1.08). El cambio revela que las personas habían mejorado en el post de manera notoria.

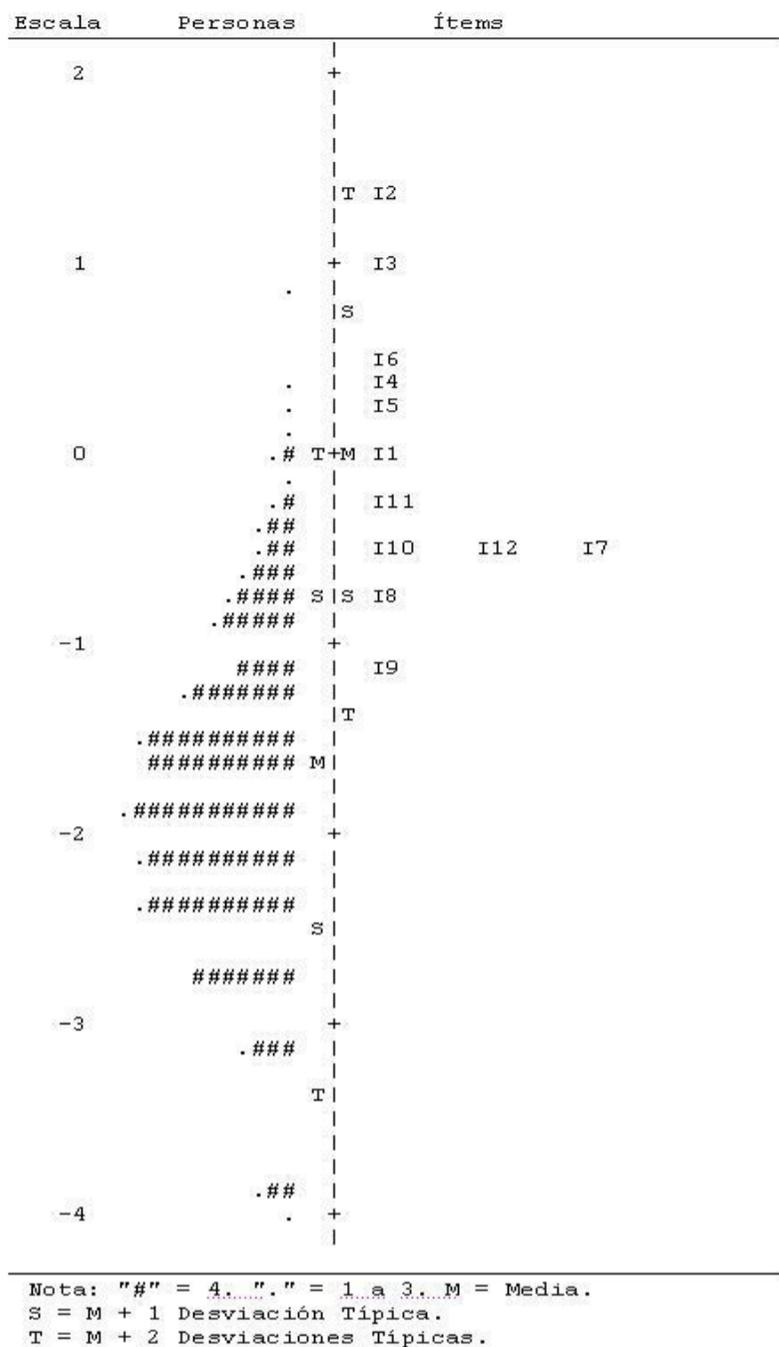
Tabla 4 Estadísticos descriptivos y diferencias entre las medias del puntaje total de HoNOS del grupo ambulatorio y del grupo internado

| | Grupo | | | | Diferencia | | |
|-------|-------------------|------|-----------------|------|------------|-----|------|
| | Ambulatorio (285) | | Internado (117) | | t | gl | Sig. |
| | Media | SD | Media | SD | | | |
| Rasch | -1.85 | .79 | -1.22 | .92 | -6.50 | 189 | .000 |
| TCT | 6.64 | 3.29 | 10.11 | 5.04 | -6.80 | 189 | .000 |

Nota. SD = desviación estándar. t = t de Welch. g. l. = grados de libertad. Sig. = p.

Fuente: elaboración propia

Figura 2. Distribución de personas (izquierda) e ítems (derecha) a lo largo del continuo de severidad. Personas e ítems por encima del punto 0 muestran mayor severidad.



Fuente: elaboración propia

Discusión y Conclusiones

Se pretendió con este trabajo generar evidencias sobre la validez, fiabilidad y sensibilidad al cambio de la versión al castellano de la HoNOS en servicios de salud mental públicos de Uruguay.

Hasta donde sabemos es el primer estudio empírico sobre las propiedades psicométricas de la HoNOS para adultos en su versión al castellano.

Se utilizaron tanto procedimientos derivados de la TCT como modelos psicométricos avanzados como el Modelo de Escalas de Calificación (MEC), una extensión para ítems politómicos del modelo dicotómico de Rasch. MEC permite una mayor calidad de la medición y una profundización en el análisis de las propiedades psicométricas del instrumento: calidad psicométrica de las categorías de respuesta, medición conjunta de los ítems y de las personas en una dimensión intervalar, medición invariante, cuantificación de la precisión de los paráme-

tros, entre otros. Este estudio permite ver algunos de los beneficios de su aplicación tanto para la investigación como para la práctica clínica al ser usado en el análisis de instrumentos y la evaluación del cambio en los pacientes.

El análisis de las categorías de respuesta determinó que las cinco categorías del sistema original de HoNOS no funcionan de manera adecuada. Se buscó una solución de mejora reduciendo las categorías. La agrupación propuesta de cuatro categorías cumple los criterios de calidad propuestos por Linacre (2002).

Los datos de la HoNOS se adecuaron suficientemente a la unidimensionalidad como para permitir el análisis mediante el modelo de Rasch.

La fiabilidad de las medidas de los ítems cuantificada por el estadístico derivado del MEC (Ri) es muy alta, indicando que el tamaño de muestra permitió estimar con mucha precisión los valores escalares de los ítems.

La fiabilidad de las puntuaciones directas de las personas se estimó mediante el coeficiente alfa de Cronbach ($\alpha=.69$), un valor que se encuentra dentro del rango de lo reportado por la mayoría de los estudios sobre la HoNOS (.59 a .76, según la revisión de Pirkis et al., [2005]). La precisión de las puntuaciones en la escala logit se estimó mediante el estadístico Rp (.66). Ambos estadísticos son ligeramente inferiores a .70, considerado el nivel aceptable.

Algunos estudios sugieren que la capacitación y la experiencia en el uso de la escala afectan a la fiabilidad (Turner, 2004). Futuros estudios deberían evaluar si con una capacitación más extensa y con el reiterado uso de la escala aumenta la fiabilidad (Adams, Palmer, O'Brien y Crook, 2009).

Al analizar por separado los resultados de los dos grupos de pacientes evaluados, se observó que en el grupo internado la precisión de la medida de las personas aumentó, y la del grupo ambulatorio disminuyó. Se observó que el promedio de severidad de los ítems fue superior al nivel promedio de las personas; y no se observó un subconjunto de ítems para la medición de los niveles más bajos de severidad de los trastornos mentales. Este resultado podría explicarse por las características de los pacientes ambulatorios de los servicios de salud mental de Uruguay, que son la mayor parte de la muestra del estudio (70.9%).

La HoNOS se desarrolló para ser usada por los pacientes de los servicios de segundo nivel de atención en salud mental, la cual en el Reino Unido corresponde a aquellos usuarios de los servicios con trastornos mentales más severos (Wing et al., 1996). En Uruguay, los especialistas de salud mental (psiquiatras y psicólogos) de los equipos de salud mental, en general atienden a todos los pacientes que consultan por problemas de salud mental, inclusive a los que tienen trastornos más leves. Por tanto, un porcentaje de los pacientes que se atienden en los servicios de salud mental de Uruguay, presentan trastornos mentales más leves que los pacientes a los que pretende medir la HoNOS, lo que podría explicar las bajas puntuaciones.

En consecuencia, la HoNOS resultó más precisa para medir a los pacientes con trastornos mentales más graves. Resultados similares a los nuestros fueron encontrados por Audin et al. (2001). Según la revisión de Pirkis, Burgess, Kirk, et al., (2005) son pocos los estudios que usaron la HoNOS para evaluar a pacientes con trastornos mentales menos severos, por lo que nuestros resultados amplían la evidencia de que la HoNOS no permite medir con alta precisión a las personas con trastornos mentales leves.

Este resultado tiene implicaciones prácticas para la utilización de la HoNOS en los servicios de salud mental, dado que la escala se plantea como una medida para ser utilizada con todos los pacientes adultos de los servicios de salud mental del segundo y tercer nivel (Wing et al., 1998). Por tanto, sería conveniente que de utilizarse como medida genérica para evaluar los resultados de los tratamientos de los servicios de salud mental de Uruguay, se reservara para pacientes con trastornos mentales más severos, y se seleccionara otro instrumento de resultados para medir a los pacientes con trastornos mentales más leves.

En lo referido a la validez de criterio, se ha observado que los promedios de la HoNOS del grupo ambulatorio y del grupo internado difirieron netamente en el sentido esperado, confirmando la capacidad de la escala para discriminar entre pacientes con diferente grado de severidad.

Los resultados mostraron que la HoNOS resultó ser una medida sensible al cambio, capaz de detectar la mejoría de las personas con trastornos mentales que reciben tratamiento en los servicios de salud mental.

Es urgente que los servicios de salud mental implementen de manera rutinaria el uso de instrumentos estandarizados y validados que permitan generar información sobre los resultados que obtienen, y de ese modo poder mejorar la calidad de la atención.

En este estudio, se encontraron algunos problemas métricos y algunas características deseables de la HoNOS como medida de resultados rutinaria para los servicios de salud mental.

Será necesario ampliar la evidencia sobre esta versión con futuros estudios a mayor escala y con diferentes poblaciones.

Referencias

- Adams, M., Palmer, A., O'Brien, J., Crook, W. (2009). Health of the nation outcome scales for psychiatry: Are they valid? *Journal of Mental Health*; 9: 2,193-198. DOI: <https://www.tandfonline.com/doi/abs/10.1080/09638230050009186>
- Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., Schulz, H. (2011). Does a more specified version of the HoNOS (Health of the Nation Outcome Scales) increase psychometric properties of the inventory? *Psychopathology*;44, 261-271. DOI: <https://www.karger.com/Article/Abstract/322690>
- Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika*; 43:561-73. DOI: <https://link.springer.com/article/10.1007/BF02293814>
- Audin, K., Margison, F. R., Clark, J. M., & Barkham, M. (2001). Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *The British Journal of Psychiatry*, 178, 561-566, DOI: <https://pubmed.ncbi.nlm.nih.gov/11388975/>
- Ausín, B., Muñoz, M., Quiroga, M.A. (2007). Adaptación española de las escalas de resultados para personas mayores HoNOS65+ (Health of the Nation Outcome Scales for Older Adults). *Revista Española de Geriatría y Gerontología*; 42(2), 88-95. DOI: <https://www.sciencedirect.com/science/article/abs/pii/S0211139X0773531X>
- Ballesteros-Urpí, A., Pardo-Hernández, H., Ferrero-Gregori, A., Torralbas-Ortega, J., Puntí-Vidal, J., Olivet, P. T., ... & Peramiqel, M. (2018). Validation of the Spanish and Catalan versions of the Health of the Nation Outcome Scale for Children and Adolescents (HoNOSCA). *Psychiatry Research*, 261, 554-559. DOI: <https://www.sciencedirect.com/science/article/abs/pii/S0165178117311757>
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. New York and London: Routledge.
- Burns, A., Beevor, A., Lelliott, P., Wing, J. (1999). Health of the Nation Outcome Scales for Elderly People (HoNOS 65+). *The Royal College of Psychiatrists*. DOI: <https://doi.org/10.1192/bjp.174.5.424>
- Colvin, K. F. & Gorgun, G. (2020). Collapsing Scale Categories: Comparing the Psychometric Properties of Resulting Scales. *Practical Assessment, Research, and Evaluation*, 25(6), 1-19. <https://scholarworks.umass.edu/pare/vol25/iss1/6>
- Coombs, T., Stapley, K., & Pirkis, J. (2011). The multiple uses of routine mental health outcome measures in Australia and New Zealand: experiences from the field. *Australasian Psychiatry*, 19(3), 247-253. DOI: <https://doi.org/10.3109%2F10398562.2011.562507>
- Delaffon, V., Anwar, Z., Noushad, F., Ahmed, A. S., & Brugha, T. S. (2012). Use of Health of the Nation Outcome Scales in psychiatry. *Advances in Psychiatric Treatment*, 18(3), 173-179. DOI: <https://doi.org/10.1192/apt.bp.110.008029>
- Delgado, A. R., Prieto, G. & Burin (2017). Constructing three emotion knowledge tests from the invariant measurement approach. *PeerJ* 5:e3755; DOI <http://dx.doi.org/10.7717/peerj.3755>
- Dzingina, M., Higginson, I. J., McCrone, P. & Murtagh, F. E. M. (2017). Development of a Patient-Reported Palliative Care-Specific Health Classification System: *The POS-E. Patient*, 10, 353-365. DOI <https://link.springer.com/article/10.1007/s40271-017-0224-1>
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33. Recuperado de <https://psycnet.apa.org/record/2010-04075-002>
- Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J., & Jezzard, R. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. *The British Journal of Psychiatry*, 174(5), 413-416. DOI: <https://search.proquest.com/openview/58cbc7b1722766c8001a03fe483a2afd/1?pq-origsite=gscholar&cbl=40635>
- Harris, M. G., Sparti, C., Scheurer, R., Coombs, T., Pirkis, J., Ruud, T., ... & Burgess, P. M. (2018). Measurement properties of the Health of the Nation Outcome Scales (HoNOS) family of measures: protocol for a systematic review. *BMJ Open*, 8(4). DOI: <http://dx.doi.org/10.1136/bmjopen-2017-021177>
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*; 30:179-85. DOI:<https://link.springer.com/article/10.1007/BF02289447>
- Kilbourne, A. M., Beck, K., Spaeth-Ruble, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: a global perspective. *World Psychiatry*, 17(1), 30-38. DOI: <https://doi.org/10.1002/wps.20482>
- Linacre, J. M. (2002). *Optimizing Rating Scale Category Effectiveness*. *Journal of Applied Measurement*; 3(1), 85-106. Recuperado de <https://pubmed.ncbi.nlm.nih.gov/11997586/>
- Linacre, J. M. (2011). *A user's guide to WINSTEPS/MINISTEPS*: Rasch-model computer programs. Recuperado de <https://www.winsteps.com/winman/copyright.htm>
- Lorenzo-Seva, U., & Ferrando, P. J. (2012). *Manual of the Program FACTOR*. Recuperado de <https://psico.fcep.urv.cat/utilitats/factor/>
- Lovaglio, P. G., & Monzani, E. (2012). Health of the nation outcome scales evaluation in a community setting population. *Quality of life research*, 21(9), 1643-1653. DOI: <https://link.springer.com/article/10.1007/s11136-011-0071-9>
- Miguel, J. P., Silva, J. T., & Prieto, G. (2013). Career decision self-efficacy scale—short form: a Rasch analysis of the Portuguese version. *Journal of Vocational Behavior*, 82(2), 116-123. DOI: <https://doi.org/10.1016/j.jvb.2012.12.001>

- Palic, S., Kappel, M. L., & Makransky, G. (2016). Rasch validation and cross-validation of the health of nation outcome scales for monitoring of psychiatric disability in traumatized refugees in Western psychiatric care. *Assessment*, 23(6), 734-743. DOI: <https://doi.org/10.1177%2F1073191115594690>
- Pirkis, J.E., Burgess, P.M., Kirk, P.K., Dodson, S., Coombs, T.J., Williamson, M.K. (2005). A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health and quality of life outcomes*; 3(76) DOI: <https://link.springer.com/article/10.1186/1477-7525-3-76>
- Prieto, G. y Velasco, A.D. (2010). Does spatial visualization ability improve after studying technical drawing? *Quality & Quantity*, 44, 1015-1024. DOI: <https://link.springer.com/article/10.1007/s11135-009-9252-9>
- Prieto, G., Delgado, A. R., Perea, M. V., & Ladera, V. (2010). Scoring neuropsychological tests using the Rasch model: An illustrative example with the Rey-Osterrieth Complex Figure. *The Clinical Neuropsychologist*, 24(1), 45-56. DOI: <https://doi.org/10.1080/13854040903074645>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of educational statistics*, 4(3), 207-230. DOI: <https://doi.org/10.3102%2F10769986004003207>
- Slade, M., Byford, S., Barrett, B., Lloyd-Evans, B., Gilbert, H., Osborn, D. P., Skinner, R., Leese, M., Thornicroft, G., & Johnson, S. (2010). Alternatives to standard acute in-patient care in England: short-term clinical outcomes and cost-effectiveness. *The British journal of psychiatry. Supplement*; 53,14-19. DOI: <https://doi.org/10.1192/bjp.bp.110.081059>
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G. & Sharpe, N. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8:33. DOI: <https://doi.org/10.1186/1471-2288-8-33>
- Speak, B. L., & Muncer, S. J. (2016). Factorial structure of the Health of the Nation Outcome Scales: An ordinal confirmatory factor analysis using a national sample of clinician ratings in England. *International Journal of Mental Health Nursing*, 25(1), 87-98. DOI: <https://doi.org/10.1111/inm.12199>
- Trauer, T. (Ed.). (2010). *Outcome measurement in mental health: theory and practice*. Cambridge University Press; 1-11.
- Turner, D. (2004). Are the Health of the nation outcome scales (HoNOS) useful for measuring outcomes in older people's mental health services? *Aging & Mental Health*;8, 387-396, DOI: <https://doi.org/10.1080/13607860410001725063>
- Uriarte, J. J., Beramendi, V., Medrano, J., Wing, J. K., Beevor, A. S., & Curtis, R. (1999). Presentación de la traducción al castellano de la escala HoNOS (Health of the Nation Outcome Scales). *Psiquiatría Pública*, 11(4), 93-101. Recuperado de <https://biadmin.cibersam.es/Intranet/Ficheros/GetFichero.aspx?FileName=HoNOS.pdf>
- Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B. G., Hadden, S., y Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry*, 172, 11-18. <https://doi.org/10.1192/bjp.172.1.11>
- Wing, J. K., Curtis, R., & Beevor, A. (1996). Health of the nation outcome scales. Report on research and development. London: *Royal College of Psychiatrists*. DOI: <https://search.proquest.com/openview/2c9695fd5b4a5338f4060190ea3c97f0/1?pq-origsite=gscholar&cbl=40635>
- Wright, B.D. (1996). Time 1 to Time 2 (Pre-test to Post-test) Comparison and Equating: Racking and Stacking. Rasch Measurement. *Transactions*. 10, 478-479.
- Wright, B.D., Stone, M.H. (1979) *Best test design*. Chicago: MESA Press.