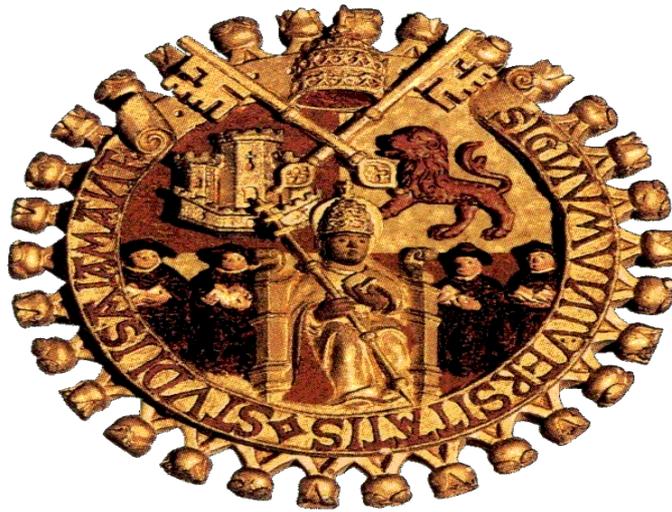


UNIVERSIDAD DE SALAMANCA

FACULTAD DE PSICOLOGÍA

Departamento de Personalidad, Evaluación y Tratamiento Psicológicos



TESIS DOCTORAL

Adaptación y Validación en Población Uruguaya de la Escala HoNOS de

Resultados en Servicios de Salud Mental

DENISSE DOGMANAS

Salamanca, 2014

Dr. D. GERARDO PRIETO ADÁNEZ, Catedrático de Metodología de las Ciencias del Comportamiento. Departamento de Psicología Básica, Psicobiología y Metodología. Facultad de Psicología. Universidad de Salamanca.

DR. D. FERNANDO JIMÉNEZ GÓMEZ, Profesor Titular del Departamento de Personalidad, Evaluación y Tratamiento Psicológicos. Facultad de Psicología. Universidad de Salamanca.

CERTIFICAN:

Que el trabajo titulado *Adaptación y Validación en Población Uruguaya de la Escala HoNOS de Resultados en Servicios de Salud Mental*, realizado bajo su dirección por Dña. DENISSE DOGMANAS PERELMAN, licenciada en Psicología y alumna del Programa de Doctorado en Psicología Clínica y de la Salud, reúne los requisitos necesarios para optar al GRADO DE DOCTOR por la Universidad de Salamanca.

Salamanca, Diciembre de 2014

Fdo.: Gerardo Prieto Adánez

Fdo.: Fernando Jiménez Gomez

Resumen

The Health of the Nation Outcome Scales (HoNOS) (Wing, Curtis, y Beevor, 1996; Wing, Beevor, Curtis, Park, Hadden y Burns, 1998) es una de las escalas más utilizadas para evaluar los resultados de forma rutinaria en los servicios de salud mental.

El presente trabajo puso a prueba la versión al castellano de la HoNOS (Uriarte, Beramendi, Medrano, Wing, Beevor, y Curtis, 1999) en población clínica de servicios públicos de salud mental de Uruguay. El objetivo principal fue analizar sus propiedades psicométricas utilizando la Teoría Clásica de los Tests y el Modelo de las Escalas de Calificación (una extensión para ítems politómicos del modelo de Rasch). Se analizaron el sistema de categorías de respuesta, el ajuste de las personas e ítems al modelo de Rasch, la fiabilidad, la validez discriminante y la sensibilidad al cambio. Además se examinó la viabilidad de la HoNOS como instrumento de evaluación rutinaria de resultados de los servicios de salud mental de Uruguay.

Previo al estudio de validación, se realizó un estudio piloto para analizar la necesidad de adaptar la HoNOS y un cuestionario socio-demográfico y clínico complementario para su uso en los servicios de salud de Uruguay. Participaron ocho clínicos en calidad de expertos. Los resultados mostraron que los instrumentos resultaron aceptables y adecuados, y se propuso adaptar lingüísticamente dos ítems de la HoNOS en dos palabras del glosario.

En los estudios empíricos de validación, participaron 402 usuarios de cuatro servicios de salud mental ambulatorios y del hospital psiquiátrico de referencia nacional: 285 (71%) en régimen de atención ambulatoria y 117 (29%) en régimen de internación psiquiátrica. En la recolección de datos participaron 22 clínicos; 21 respondió además un cuestionario de valoración de la HoNOS para analizar la viabilidad de su utilización en Uruguay.

Los resultados indican que el sistema original de cinco categorías de la HoNOS no funciona correctamente de acuerdo con las recomendaciones de Linacre (2002). Se buscó una solución mediante la agregación de categorías adyacentes reduciendo las cinco categorías iniciales a cuatro, y su re-análisis mostró un mejor ajuste y funcionamiento. Se observó un aceptable ajuste del instrumento al modelo, y adecuada validez para discriminar a los pacientes con diferentes niveles de severidad de trastorno mental.

El estudio del cambio en los 75 pacientes evaluados en dos tiempos, mostró cambios estadísticamente significativos y un tamaño de efecto entre moderado a alto entre las medias de las puntuaciones en el pre-tratamiento y el post-tratamiento, lo que demuestra que la HoNOS es una medida sensible al cambio. En el análisis a nivel individual del cambio medido con el Índice de Cambio Fiable (ICF) y el Índice de Mejoría Fiable (IMF), se observó que con la puntuación clásica el 10.7% de la muestra presentó un cambio fiable y el 14.7% una mejoría fiable; y con los indicadores de cambio y mejoría correspondientes al modelo de Rasch el 20% y 29.3% respectivamente. El análisis de la convergencia entre la clasificación de los pacientes con los puntos de corte de la HoNOS medidos con el ICF, el IMF y los indicadores correspondientes al modelo de Rasch, y el punto de vista del criterio de los clínicos considerando a los pacientes que estaban en condiciones de egresar del hospital, mostró unos índices de sensibilidad entre bajos y moderadamente bajos.

El análisis de la viabilidad muestra que la HoNOS fue valorada positivamente por los clínicos participantes, pero con diverso grado de aceptación.

Se concluye que la HoNOS presenta algunos problemas métricos y algunas características deseables como medida de evaluación rutinaria de resultados de los servicios

Adaptación y Validación en Población Uruguaya de la Escala HoNOS

de salud mental, y que presenta mejores propiedades cuando se utiliza en pacientes con trastornos mentales más severos.

Palabras clave: Servicios de salud mental, Health of the Nation Outcome Scales (HoNOS), Modelo de Rasch.

Agradecimientos

En primer lugar quiero agradecer a mis directores de tesis, Gerardo Prieto y Fernando Jiménez, por su generosa transmisión de conocimientos y constante disposición a orientarme y a ayudarme durante el proceso de desarrollo de este trabajo.

A la Facultad de Psicología y La Comisión Sectorial de Investigación Científica de la Universidad de la República, por apoyar el desarrollo de la formación doctoral en Psicología, facilitando las condiciones para que este trabajo haya podido realizarse.

A la Dirección de Salud Mental y Poblaciones Vulnerables de la Administración de Servicios de Salud del Estado (ASSE), en particular a Gabriela Novoa y Horacio Porciúncula; y a los profesionales de los centros de salud participantes porque sin su colaboración en la recogida de los datos, esta investigación no hubiera sido posible.

A los profesionales participantes del estudio piloto, y a Daniela Gleizer y Gabriela Fernández por colaborar durante el ingreso de los datos. A José Juan Uriarte por permitirme con amabilidad trabajar e intercambiar sobre la versión al castellano de la HoNOS.

A mis compañeros y amigos de trabajo de la Facultad de Psicología, especialmente a Rosa Zytner, Liliana Guerrero, Laura de Souza, Daniela Gleizer, Paola Premuda, Robert Pérez, Andrea Bielli y Alejandra Arias; y del Ministerio de Salud Pública, a Ariel Montalbán, Gimena Paroli, Fernanda Porteiro, Gabriela Suárez y previamente, Renée del Castillo y Hebert Tenenbaum, por su apoyo y cariño durante todas las etapas de este trabajo.

Otro agradecimiento quiero transmitir a todos aquellos compañeros y colegas que con su tiempo y sus consejos me han dado aliento en diferentes momentos de este proceso:

Fernanda Barboza, Ricardo Bernardi, Fernando Berriel, Juan Dapuetto, Susana Martínez, Ana Luz Protesoni, y especialmente a Damián Schroerder.

Adaptación y Validación en Población Uruguaya de la Escala HoNOS

A mi compañero de la vida Sergio, por todo su amor, apoyo y comprensión. A mi querida familia; padres, hermanos, tíos, primos, sobrinos y amigas, por su afecto incondicional en todo momento.

A todos ellos, mi eterno agradecimiento.

Tabla de Contenido

Lista de Tablas	iv
Lista de Figuras	vi
Lista de Apéndices	vii
INTRODUCCIÓN	1
I. REVISIÓN TEÓRICA	5
1. Medición de Resultados en los Servicios de Salud Mental	6
1.1. Surgimiento	6
1.2. La evaluación de resultados como práctica rutinaria de los servicios	11
1.3. El uso de medidas estandarizadas	13
1.4. Beneficios de utilizar instrumentos para evaluar resultados de manera rutinaria	14
2. Instrumentos de Evaluación de Resultados de Servicios de Salud Mental	17
2.1. Panorama actual	18
2.2. Requisitos de los instrumentos de ROA en salud mental	22
2.3. Principales instrumentos de evaluación rutinaria de resultados en salud mental	26
23. Prácticas Actuales de ROA en los Servicios de Salud Mental	29
3.1. Nivel internacional	29
3.2. Uruguay	32
3.3. Desafíos actuales	34
4. The Health of the Nation Outcome Scales (HoNOS)	37

4.1. Características	37
4.2. Uso a nivel internacional	41
4.3. Propiedades psicométricas	43
5. El Modelo de Rasch	54
5.1. Características del Modelo de Rasch	55
5.2. El Modelo de Escalas de Calificación	58
II. ESTUDIOS EMPÍRICOS	60
Introducción	61
Estudio piloto de la HoNOS	63
1. Objetivos	63
2. Metodología	63
2.1. Participantes	63
2.2. Instrumentos	63
2.3. Procedimiento	65
3. Resultados	66
3.1. HoNOS	66
3.2. Ficha de información complementaria a la Escala HoNOS (CCH)	69
4. Conclusiones	70
Estudio de Validación de la HoNOS	72
1. Objetivos	72
1.1. Objetivos generales	72
1.2. Objetivos específicos	72

2. Metodología	72
2.1. Participantes	72
2.2. Instrumentos	76
2.3. Procedimiento	78
2.4. Análisis de datos	80
3. Resultados	86
3.1. Análisis del funcionamiento de las categorías de respuesta de la HoNOS	86
3.2. Construcción de un sistema apropiado de categorías de la HoNOS	88
3.3. Análisis de la dimensionalidad	91
3.4. Análisis de los valores de las personas e ítems de la HoNOS	91
3.5. Validez discriminante de la HoNOS	100
3.6. Sensibilidad al cambio de la HoNOS	101
3.7. Viabilidad de la HoNOS	111
4. Discusión	112
III. CONCLUSIONES	126
Referencias	132
Apéndices	A-1

Lista de Tablas

Tabla 1. Principales instrumentos que cumplen con los requisitos de ROA	28
Tabla 2. Ítems que componen la HoNOS	39
Tabla 3. Evidencia de las propiedades psicométricas de la HoNOS según revisión de Pirkis	45
Tabla 4. Valores promedio de las respuestas de los expertos sobre la relevancia y claridad de los ítems de la HoNOS	68
Tabla 5. Características sociodemográficas de la muestra	74
Tabla 6. Frecuencia de pacientes por grupo y diagnóstico principal	76
Tabla 7. Calidad psicométrica de las categorías de respuesta de la HoNOS según criterios de Linacre	87
Tabla 8. Calidad psicométrica de las categorías de respuesta de la HoNOS, con agrupación de categorías según criterios de Linacre	89
Tabla 9. Estadísticos descriptivos de los valores de las personas	93
Tabla 10. Nivel de dificultad, error estándar asociado, estadísticos de ajuste y correlaciones ítem-escala (r_{ix}) de los ítems de HoNOS	95
Tabla 11. Estadísticos descriptivos y diferencias en logit entre las medias de los hombres y de las mujeres	98
Tabla 12. Estadísticos descriptivos en los diferentes grupos diagnósticos	99
Tabla 13. Estadísticos descriptivos y diferencias entre las medias del puntaje total de HoNOS del grupo ambulatorio y del grupo internado	101
Tabla 14. Diagnósticos principales del grupo ambulatorio e internado evaluado en el pre-	102

tratamiento y post-tratamiento	
Tabla 15. Estadísticos descriptivos y diferencia entre las medias de las puntuaciones en el pre y post-tratamiento	103
Tabla 16. Estadísticos descriptivos y diferencias entre el pre-tratamiento y el post-tratamiento en los ítems de la HoNOS	105
Tabla 17. Estadísticos descriptivos y diferencias entre las medias del pre- y post-tratamiento en el grupo ambulatorio y el grupo internado	106
Tabla 18. Estadísticos descriptivos de las puntuaciones de los sujetos en el pre-tratamiento y post-tratamiento y datos de anclaje	108
Tabla 19. Análisis del cambio individual en HoNOS entre el pre-tratamiento y el post-tratamiento medidos con los índices ICF_{TCT} , IMF_{TCT} , $Cambio_{Rasch}$ y $Mejoría_{Rasch}$	110
Tabla 20. Frecuencias, medias y desviaciones típicas de las respuestas de los clínicos en el cuestionario sobre viabilidad de la HoNOS	112

Lista de Figuras

Figura 1. Representación gráfica de las curvas características de las categorías de respuesta de la HoNOS	88
Figura 2. Representación gráfica de las curvas características de las categorías de respuesta de la HoNOS agrupadas	90
Figura 3. Representación conjunta de personas e ítems de la HoNOS	97

Lista de Apéndices

Apéndice A. Escalas de evolución HoNOS: versión en lengua castellana	A-2
Apéndice B. Ficha de información complementaria a la Escala HoNOS	A-10
Apéndice C. Cuestionario para la adaptación de la HoNOS en Uruguay	A-12
Apéndice D. Cuestionario sobre la ficha complementaria a la Escala HoNOS Uruguay	A-15
Apéndice E. Hoja de registro de HoNOS	A-17
Apéndice F. Cuestionario de Viabilidad de HoNOS	A-20

INTRODUCCIÓN

INTRODUCCIÓN

Desde hace tiempo se reconoce que la investigación de resultados es un pilar para mejorar la calidad de la asistencia y el impacto de la atención de los servicios de salud mental (Heinemann, Fisher, y Gershon, 2006; Kazdin, 2008).

Evaluar los resultados de los servicios de salud permite una práctica reflexiva que contribuye a mejorar los tratamientos que se brindan (National Institute for Mental Health [NIMH], 2008) y reporta múltiples beneficios a todos los involucrados: clínicos, usuarios, directores de servicios, investigadores y administración pública.

En la actualidad existe un consenso internacional respecto a que los servicios de salud mental deben evaluar sistemáticamente sus resultados como parte de la práctica clínica habitual (Jacobs y Moran, 2010; Slade et al., 2006). Esta práctica se ha denominado Routine Outcome Assessment (ROA) o Routine Outcome Measurement (ROM). La ROA se ha expandido internacionalmente y en la actualidad se implementa en los servicios de salud mental de diferente manera alrededor del mundo (Jacobs, 2009; Trauer, 2010a).

A medida que se ha ido avanzando en la teorización y experiencia en ROA se han establecido varios consensos; uno de los principales plantea que es necesario utilizar medidas estandarizadas que cumplan con ciertos requisitos. Este consenso ha llevado a un desarrollo tan importante de este tipo de medidas que existe en la actualidad un enorme número de instrumentos que se usan para evaluar resultados en salud mental, especialmente de adultos (NIMH, 2008).

Dentro de la vasta cantidad de medidas de resultados, The Health of the Nation Outcome Scales (Wing, Curtis, y Beevor, 1996; Wing et al., 1998) cumple con la mayoría de los requisitos que se consideran clave para ROA.

The Health of the Nation Outcome Scales (HoNOS) se ha convertido en uno de los instrumentos más utilizados a nivel internacional, y ha pasado a formar parte del conjunto de datos básicos de recolección obligatoria en los países que poseen mayor desarrollo en el campo de la evaluación de resultados en servicios de salud mental. Desde su creación hasta el momento, la HoNOS ha sido traducida a varios idiomas y cuenta con numerosos estudios de validación en distintos países. Uriarte et al. (1999) han llevado a cabo la traducción oficial al castellano en 1999, pero hasta el momento no se han realizado estudios empíricos sobre las propiedades psicométricas de esta versión (ver Apéndice A).

Por otra parte, la HoNOS ha sido desarrollada y tradicionalmente estudiada desde el modelo de la Teoría Clásica de los Tests (TCT). En la actualidad, existen nuevos modelos psicométricos más sofisticados, como el modelo de Rasch (1960), que constituye un acercamiento alternativo a la TCT y resuelve algunas de sus desventajas metodológicas (Wright y Stone, 1979). Así, se pretende con la presente investigación analizar con la TCT y el modelo de Rasch las propiedades de la HoNOS, poniendo a prueba si es efectivamente una medida adecuada para la evaluación de los resultados de los servicios de salud mental.

En Uruguay se han realizado en los últimos años importantes avances en el desarrollo y recolección de datos básicos de salud mental, especialmente a nivel del Sistema de Salud (Ministerio de Salud Pública [MSP], 2011). Sin embargo, el sistema de recolección de datos actual no incluye ningún instrumento estandarizado para evaluar resultados. Tampoco existe ningún instrumento de ROA con estudios de validación en el país. Al poner a prueba la HoNOS en servicios públicos de salud mental, se pretende contribuir a la utilización de instrumentos que permitan medir de manera rutinaria los resultados de los servicios de salud mental del país.

El trabajo se organiza de la siguiente manera: el capítulo I consiste en una revisión teórica, en el capítulo II se presenta la investigación empírica realizada, y en el capítulo III se presentan las conclusiones.

La revisión teórica se divide en cinco partes. La primera parte se dedica al campo de la evaluación de resultados en salud mental y el movimiento hacia su práctica de manera rutinaria (ROA). La segunda parte aborda los instrumentos de medida de resultados en servicios de salud mental y los principales requisitos para las medidas de ROA. La tercera parte se dedica a revisar las prácticas actuales de utilización de instrumentos de ROA en los servicios de salud mental, e incluye un resumen sobre la situación actual, y algunos de los principales desafíos que están planteados en el tema. Posteriormente se dedica una sección a describir las características y estudios realizados con la escala HoNOS. La revisión teórica culmina con la descripción del modelo de Rasch.

El capítulo II se inicia con una introducción a los estudios empíricos llevados a cabo en esta investigación. Posteriormente se describen los objetivos, metodología, resultados y discusión del estudio piloto y para el de validación.

En el capítulo III se presentan las conclusiones del trabajo. También se establecen las limitaciones de la investigación y algunas sugerencias para futuros estudios en el tema. Para finalizar, se analizan algunas de las implicaciones de los resultados de este trabajo con miras a la potencial utilización de HoNOS en los servicios de salud mental de Uruguay.

I. REVISIÓN TEÓRICA

I. REVISIÓN TEÓRICA

1. Medición de Resultados en los Servicios de Salud Mental

1.1. Surgimiento

La investigación de resultados representa una de las áreas de actividad de la salud pública más extensas y de más rápido desarrollo (Clark, 2002). El objetivo de este tipo de investigación es detectar cuáles son las intervenciones más efectivas para mejorar la calidad de la atención y, como consecuencia, la calidad de vida de los usuarios.

El movimiento de medición (*measurement*) y gestión (*management*) de resultados en salud surge como consecuencia de varios factores: los costos, la aceleración en el desarrollo de nuevas tecnologías en salud, y la evidencia sobre las variaciones en la forma en la que se brinda la atención (Gilbody, House, y Sheldon, 2002a). A la vez, la heterogeneidad de la atención trae como consecuencia diferencias en los resultados y en la calidad entre los servicios (Clark, 2002). La evidencia sobre la gran variabilidad, tanto en los recursos que se destinan a los servicios como en los resultados que se obtienen, ha generado la preocupación por una distribución productiva (Fonagy, 2011).

En el campo de la salud mental se vuelve fundamental lograr una distribución de recursos basada en los resultados, debido a la alta prevalencia de los trastornos mentales y el desajuste entre las necesidades y los recursos disponibles. Los problemas de salud mental constituyen 5 de las 10 causas principales de discapacidad y un 14% de toda la carga global de enfermedades (Prince et al., 2007).

A la alta prevalencia se agrega la importante dificultad en el acceso a los servicios de las personas con trastornos mentales. Este fenómeno se ha denominado *brecha de*

tratamiento, porque expresa la proporción de personas que necesitan atención y no la reciben. Según la Organización Mundial de la Salud [OMS] (2013), “en los países de ingresos bajos y medios, entre un 76% y un 85% de las personas con trastornos mentales graves no reciben tratamiento” (p. 6). Los sistemas de salud no han logrado aun brindar una respuesta adecuada a las personas con trastornos mentales (OMS, 2013; Organización Panamericana de la Salud [OPS], 2009). “Sin embargo, los recursos disponibles para afrontar la enorme carga que representan las enfermedades mentales son insuficientes, están distribuidos desigualmente y en ocasiones son utilizados de modo ineficaz” (OPS, 2009, p. 1).

Frente a recursos finitos, y a la brecha de tratamiento existente en la atención en salud mental, se hace necesario generar información respecto a las intervenciones más efectivas y a su impacto en las personas. Esta información permite evaluar el destino de los recursos (Fonagy, 2011) para lograr una atención efectiva y eficiente (Jacobs, 2007; Maloney y Chaiken, 1999). La práctica de investigación de resultados otorga mayores garantías de que los servicios están siendo efectivos y útiles (Jacobs, 2007) y cada vez es más utilizada en la dirección de los servicios y en el diseño de las políticas gubernamentales (Lutchman et al., 2007). En Estados Unidos y en Europa tiene un rol cada vez más importante en la valoración de la efectividad y costo-eficiencia de los servicios de salud, así como en su financiación (Newnham y Page, 2010).

Otra de las razones que explican el importante crecimiento de la medición de resultados en salud en general, y en la salud mental en particular, es el movimiento de la práctica basada en evidencias (PBE).

La práctica basada en evidencias (PBE). La PBE ha sido considerada una de las tendencias más importantes en la historia de la atención a los servicios de salud mental

(Hardiman, Theriot, y Hodges, 2005). A través de la literatura se puede observar que una amplia gama de profesionales y organizaciones ha trabajado sistemáticamente para lograr una práctica basada en evidencias (Digennaro Reed y Reed, 2008). Algunas de estas disciplinas son la medicina, la psicología clínica, el análisis del comportamiento, el counselling, la educación, y la enfermería (Kazdin, 2008).

La PBE comenzó con los cambios en los reembolsos de los seguros de salud, en los que se exige generar guías de práctica clínica como política central para la prestación de servicios de salud (Institute of Medicine, 2001, citado en Lane-Herman, 2012). En salud mental, la PBE surgió a partir de la Medicina Basada en la Evidencia (Evidence Based Medicine; EBM), adhiriéndose a los mismos principios que esta (Hardiman et al., 2005). La EBM propone utilizar la investigación para informar sobre la práctica clínica y mejorar los resultados (Sox y Woolf, 1993; Woolf y Atkins, 2001, citado en APA Presidential Task Force on Evidence-Based Practice, 2006).

En psicología ha sido definida por la APA Presidential Task Force on Evidence-Based Practice (2006) como “la integración de la mejor evidencia disponible con el juicio clínico en el contexto de las características del paciente, la cultura y preferencias” (p. 273). La definición sigue los lineamientos de la utilizada por el Institute of Medicine (IOM) para definir la PBE adoptada por Sackett, Straus, Richardson, Rosenberg y Haynes (2000).

En su origen la PBE se apoyó en los estudios de eficacia de los tratamientos (llevados a cabo en situaciones de máximo control experimental), con su patrón de oro, el ensayo clínico aleatorizado (*randomized clinical trial*; RCT). Los RCT se caracterizan por el control mediante aleatorización, manipulación de variables, tratamientos manualizados y criterios

exigentes de inclusión de pacientes, y se enfocan sobre todo en la validez interna más que en la validez externa (Seligman, 1995).

Sin embargo, la evidencia de los RCT ha sido criticada por los clínicos por considerar que las condiciones de laboratorio de los estudios de eficacia de los tratamientos son distintas a las de la práctica clínica diaria y que, por lo tanto los resultados podrían no ser generalizables (Kazdin, 2008; Newnham y Page, 2010). En la práctica habitual de los servicios las intervenciones que se aplican no siguen los manuales y los pacientes no son homogéneos (como en los ensayos controlados), sino que en general presentan varios problemas y comorbilidad diagnóstica (Howard, Moras, Brill, Martinovich, y Lutz, 1996; National Institute of Mental Health [NIMH], 2004; Seligman, 1995). Esto generó interés tanto en los estudios de eficacia como de efectividad de los tratamientos (Seligman, 1995). Los estudios de efectividad están basados en la evaluación de los resultados de las intervenciones en las condiciones reales de la práctica clínica, con los terapeutas y pacientes de la práctica habitual de los servicios. Estos estudios poseen la ventaja de evitar la pérdida de generalizabilidad y de validez externa que se les critican a los estudios de eficacia (Jacobs, 2007).

En la actualidad se considera que ambos tipos de estudios son importantes y contribuyen a generar la evidencia sobre los tratamientos. Los servicios de salud mental basados en la evidencia deberían utilizar tratamientos que hayan demostrado su eficacia en condiciones controladas y que luego hayan sido probados en las condiciones del mundo real para determinar la medida en que la eficacia se generaliza y son efectivos (Newnham y Page, 2010).

Definición de resultados de los servicios de salud mental. En lo que respecta a las personas con trastornos mentales, resulta complejo acordar una definición de resultado (Trauer, 2010a). Existen muchos puntos de vista sobre lo que representa un resultado útil (NIMH, 2008). Fonagy, Matthews, Pilling y Glover (2005) afirman que la medición de resultados se ha visto afectada por una “falta de claridad respecto a las naturaleza de los resultados en sí mismos” (p. 7). Además, definir lo que se entiende por resultado no solo afecta a su medición, sino que tiene consecuencias teóricas y profesionales, y que afectan también la administración pública de los recursos. El debate sobre cómo es mejor medir la salud y el bienestar siempre estará presente, especialmente cuando esas mediciones pueden ser usadas para sancionar determinados tratamientos, racionalizar otros, y evaluar la actividad de los servicios y los beneficios del gasto público (NIMH, 2008).

En la década del 60, Donabedian (1966), considerado uno de los principales fundadores de los sistemas de gestión de la calidad, propuso un modelo para evaluar científicamente la calidad de la práctica en el sector de la salud; dicho modelo consistía en un esquema con tres componentes que se encuentran relacionados: estructura, proceso y resultados. La estructura alude a las características relativamente estables de la institución de salud (su infraestructura física y recursos). El proceso hace referencia a la forma de organizar y prestar la asistencia al usuario, la cantidad de tratamientos administrados, la accesibilidad, la relación terapéutica (Clark, 2002). El componente de resultado se refiere a los efectos o resultados (*outcomes*) de la atención sobre el estado de salud de los usuarios (Donabedian, 1980). Este modelo se constituyó en el núcleo del campo de estudio de investigación sobre servicios de salud (Clark, 2002), y fue luego ampliamente utilizado por muchas disciplinas del campo de la salud (Maloney y Chaiken, 1999).

Una definición de resultados muy utilizada en el campo de la salud mental es la planteada por Andrews, Peters y Teesson (1994): “el efecto en el estado de salud de la persona que es atribuible a una intervención hecha por un profesional de la salud o un servicio de salud” (p. 14). Si bien esta definición ha recibido objeciones por parte de algunos autores (por ejemplo, Slade, 2002) sigue siendo muy utilizada y, según Trauer (2010b), es la más citada ya que resulta operativa. Por estas razones es la que se utiliza en este estudio.

Sin embargo, hay aspectos que es necesario considerar a la hora de analizar los resultados en salud mental para evitar caer en reduccionismos. Uno de ellos es que no existe relación directa entre una intervención y los resultados; los cambios en el estado del usuario pueden estar influidos por otros aspectos que ocurran simultáneamente con el tratamiento (como la autoayuda), o ser producto de dichos aspectos (Slade, 2002a). Por otra parte, los efectos de las intervenciones no siempre son positivos, sino que pueden también resultar negativos o no producirse cambios. Uno de los problemas de las personas que poseen trastornos mentales crónicos es justamente la falta de cambio que los caracteriza (Trauer, 2010b). Además, en la atención en salud mental, que no se produzcan cambios podría llegar a ser un resultado positivo, indicando que el usuario no tuvo una recaída (Fonagy, 2011; Trauer, 2010b).

1.2. La evaluación de resultados como práctica rutinaria de los servicios

A los servicios de salud se les ha venido requiriendo cada vez más demostrar la mejora de la calidad, y se ha llegado, así, a un consenso internacional de que los resultados deben ser medidos de manera rutinaria. La práctica rutinaria de evaluación de resultados por parte de los clínicos (denominada Routine Outcome Assessment; ROA), significa que forma parte de la práctica clínica habitual de los servicios de salud y no se realiza en el contexto de

una investigación puntual, con recursos humanos y materiales adicionales destinados específicamente para eso (Slade, 2004).

La ROA es consecuencia del movimiento de evaluación de resultados para la gestión de los servicios (*outcomes management*), que encuentra su origen en el trabajo de Ellwood de 1988 (Gilbody, House, y Sheldon, 2003). Ellwood sugirió que la evaluación rutinaria y frecuente por parte de los clínicos permitiría la construcción de bases de datos que se convertirían en un recurso para los clínicos y la investigación de los servicios de salud, y se podrían usar con diferentes objetivos, como comparar tratamientos, y servicios o evaluar nuevas tecnologías (Beurs et al., 2011; Jacobs, 2007). Ellwood propone la gestión de resultados como una tecnología capaz de ser utilizada por los clínicos para evaluar, monitorear, asegurar una atención de calidad y mejorar la calidad de vida de sus pacientes (Andrews y Page, 2005; Gilbody et al., 2002a).

La ROA implica un monitoreo de la práctica clínica de manera regular con el objetivo de impactar en sus resultados. Jacobs (2009) revisa estudios que aportan evidencia a favor de la evaluación de manera sistemática en los servicios de salud mental, y concluye:

El mensaje clave de estos estudios es que una sola medición de resultados (o medición de resultados infrecuente) parece tener resultados equívocos en cuanto a los términos en los que realmente se mejoran los resultados subjetivos, pero la medición de resultados que es hecha longitudinalmente y más regularmente (con más de una o dos administraciones) puede significativamente mejorar la calidad de vida del paciente o reducir las admisiones psiquiátricas. (p. 12)

La evaluación periódica permite además, tomar decisiones basadas en el progreso del paciente, impactando en la toma de decisiones sobre continuar, alterar o terminar el tratamiento (Kazdin, 2008).

1.3. El uso de medidas estandarizadas

Tradicionalmente los resultados de la atención en salud se evaluaban usando el juicio clínico. Sin embargo, actualmente en la política de salud mental y la investigación se reconoce que además del juicio clínico, es importante usar métodos estandarizados formales (Slade, 2006). De hecho, frente a la creciente demanda de rendición de cuentas de los servicios de salud, una de las formas de respuesta ha sido la introducción de la medición rutinaria de los resultados utilizando para ello instrumentos estandarizados (Fonagy et al., 2005).

En la literatura especializada se puede encontrar frecuentemente el término “medidas de resultados” (*outcome measures*) para referirse a un formato estandarizado para recoger información sobre los resultados de los usuarios. En este trabajo se utiliza los términos *instrumento y/o medidas de resultados* para referirse a los instrumentos psicométricos. La evaluación rutinaria de los resultados en servicios de salud mental (ROA) se refiere en esta investigación al uso continuado de instrumentos psicométricos para medir los resultados de los tratamientos que brindan los servicios de salud mental.

Son varias las ventajas que pueden mencionarse al medir los resultados de los tratamientos utilizando instrumentos (Trauer, 2010b). En la siguiente sección se describen las principales aplicaciones y beneficios de la práctica de evaluación de resultados de servicios de salud mental cuando se realiza de manera sistemática y utilizando instrumentos de resultados.

1.4. Beneficios de utilizar instrumentos para evaluar resultados de manera rutinaria

La ROA provee información de los resultados de una forma accesible y común a todos los involucrados (Fonagy et al., 2005) incluyendo a usuarios, familiares, clínicos, directores de servicios, administración pública, agencias financieras, investigadores y sociedad.

En lo que respecta a los clínicos uno de los beneficios de utilizar instrumentos es que esto aporta información para la planificación y la toma de decisiones clínicas que permiten mejorar la práctica y la calidad de la atención. La recolección de datos sobre los resultados tiene sentido “cuando se aplica y traslada en cambios positivos en la práctica, cuando resulta en una mejora de la calidad de los servicios” (Fonagy et al., 2005, p. 9). Los instrumentos permiten realizar mejores definiciones y apreciaciones más precisas que el juicio clínico global (Kazdin, 2008).

Las medidas pueden ser útiles al inicio, durante y al final del tratamiento. Antes de iniciar el tratamiento brindan información sobre el tipo y la severidad del trastorno, lo que puede ser utilizado para optimizar la distribución de los pacientes a los diversos tipos de tratamientos (Beurs et al., 2011). Durante el tratamiento permite definir con mayor precisión los cambios. Los pacientes cambian de manera distinta y con tiempos disímiles: en algunos usuarios se pueden observar cambios rápidos, mientras que en otros se requiere de más tiempo, o puede no haber cambios (Beurs et al., 2011; Trauer, 2010a). También existen diferencias entre las áreas y las intensidades en las que se va produciendo el cambio en los diferentes usuarios. En algunos usuarios los cambios se producen de manera gradual e imperceptible, y los instrumentos permiten observar estas pequeñas diferencias (Kazdin, 2008).

Los instrumentos brindan también la posibilidad de realizar durante el tratamiento modificaciones que disminuyan las probabilidades de abandono. Muchas veces los clínicos no pueden anticipar los tratamientos que tendrán escasos resultados o “fracasarán” y las medidas de resultados pueden ayudar a re-evaluar el plan de tratamiento (Barry, 2011). Algunos estudios han demostrado que a través de la información recolectada y el feedback que los clínicos reciben con la evaluación, algunos casos en los que era previsible que el tratamiento fallara no solo mejoraron en comparación con los usuarios del grupo control, sino que además se redujeron los abandonos y las tasas de deterioro (Lambert et al., 2001; Lambert et al., 2002, citados por Barry, 2011).

Slade et al. (2006) realizaron una investigación utilizando medidas estandarizadas en una muestra representativa del Reino Unido para evaluar la efectividad de ROA. El estudio, que utilizó como metodología el RCT, incluyó 101 pares de pacientes y profesionales de salud mental de adultos. Los participantes debieron completar cuestionarios todos los meses y recibieron cada 3 meses feedback del tratamiento. El grupo control (n = 59) recibió el tratamiento usual. En este estudio observaron en los pacientes una reducción en las admisiones a internación psiquiátrica, y por tanto también, según los autores, un ahorro en los costos provenientes de la utilización de los servicios de salud.

Por otra parte, contar con datos externos sobre la efectividad del tratamiento complementa el juicio clínico a nivel de las limitaciones o sesgos que pueden ocurrir en la percepción e implicación del clínico para recoger la información (Kazdin, 2008).

Hatfield y Ogles (2006) demostraron, en una investigación empírica en la que encuestaron a 810 psicólogos, que a pesar de que los profesionales afirmaban que el aspecto que más incidía para tomar decisiones clínicas eran las manifestaciones verbales de los

pacientes, también se basaban en los datos que aportaban los instrumentos de resultados. En esta misma investigación los autores encontraron que una de las razones más importantes por las cuales los psicólogos usaban medidas de resultados era por considerar que forman parte de una práctica ética. Brindar el tratamiento que más garantías de mejoría ofrezca (Londoño y Valencia, 2005) forma parte de la ética y responsabilidad social por parte de los profesionales.

Los usuarios y familiares de los servicios de salud se benefician, ya que la medición de los resultados contribuye al monitoreo sistemático y mejorar la atención que reciben.

Por otra parte, tradicionalmente el rol del paciente en la evaluación se consideró secundario (Jacobs, 2009). Sin embargo, en los últimos años se ha venido valorando cada vez más la importancia de la participación de los pacientes para la gestión de los servicios, tanto en las decisiones como en la evaluación de los resultados (Erbes et al., 2004; Slade, 2002a; Smith, Fischer, Nordquist, Mosley, y Ledbetter, 1997). La perspectiva de los usuarios puede o no coincidir con la de los clínicos. Existe abundante evidencia de la falta de concordancia entre usuarios y proveedores (Trauer, 2010b). La falta de concordancia se ha visto tanto en la valoración que realizan de los resultados obtenidos como en lo que respecta a cuáles son más importantes para su medición (Crawford et al., 2011). Los instrumentos de resultados facilitan que la perspectiva de los usuarios y cuidadores sea tomada en cuenta sistemáticamente y formalmente (Trauer, 2010b).

A los directores de las instituciones y los servicios les permite mejorar la gestión, planificar el desarrollo del servicio, determinar las prioridades en la asignación de los recursos y tomar decisiones informadas.

Por otra parte, a nivel de la administración pública, la exigencia de rendición de cuentas exige contar con información sobre los resultados de los servicios de salud, para los que se han destinado fondos públicos.

Los costos de la enfermedad mental son muy altos para los individuos, sus familiares, los empleadores y el gobierno. Los trastornos mentales producen costos directos de la atención en salud, e indirectos debido a la ausencia laboral o la discapacidad para trabajar (Cyhlarov, McCulloch, McGuffin, y Wykes, 2010).

En este sentido, generar información a través de los instrumentos sobre las estrategias más efectivas en el tratamiento y reducción de trastornos mentales impacta en la economía y en la sociedad, en tanto brinda información para el diseño, planificación, y evaluación de las políticas gubernamentales en salud, así como para su financiación. Contar con datos globales permite ajustar las políticas y dirigir los recursos de acuerdo a las necesidades (Trauer, 2010a).

Por último, evaluar los tratamientos que se utilizan en la práctica clínica real de los servicios de salud aporta al desarrollo del conocimiento científico en salud mental. Los datos de la práctica clínica pueden contribuir directamente a generar hipótesis que se estudien en investigaciones, y a ajustar la investigación de los tratamientos a temas relevantes de la práctica clínica (Kazdin, 2008).

2. Instrumentos de Evaluación de Resultados de Servicios de Salud Mental

En este apartado se describen algunas características del panorama actual de los instrumentos de resultados en salud mental, los requisitos para los instrumentos de ROA, y por último algunos de los principales instrumentos.

2.1. Panorama actual

Uno de los consensos fundamentales en el campo de ROA es la utilización de medidas estandarizadas. Sin embargo, en lo que respecta a los métodos y la selección de instrumentos, la falta de consenso y organización ha sido una característica (Gilbody et al., 2003). Una causa del problema radica en que existen miles de instrumentos de resultados disponibles y que se están utilizando especialmente en relación a la salud mental del adulto (NIMH, 2008).

El trabajo de Froyd, Lambert y Froyd (1996) revisó estudios de resultados publicados entre 1983 y 1988 en 20 revistas clave, y observaron que en 348 estudios se aplicaron 1430 medidas diferentes. Por su parte, la revisión de Herman et al. (2002, citado en Burlingame, Dunn, y Chen, 2005) observó más de 300 medidas propuestas para evaluar la calidad de los servicios de salud mental. El estudio de Gilbody et al. (2003) sobre los instrumentos utilizados por los psiquiatras en su práctica clínica en el Reino Unido observó, de manera similar, gran diversidad y heterogeneidad.

En la actualidad se presenta un amplio menú de opciones en cuanto a instrumentos de resultados para evaluar distintas dimensiones, enfoques, abordajes, y perspectivas. Dentro de esta gran oferta de opciones existen instrumentos genéricos o específicos para determinados trastornos o problemas de salud mental. Los genéricos permiten evaluar a cualquier paciente independientemente de su trastorno, edad o tratamiento (Maloney, 1999). Este tipo de instrumentos se pueden utilizar con grupos amplios, y pretenden aplicarse universalmente a todos los usuarios dentro del sistema de salud mental (Trauer, 2010b). Presentan la ventaja de permitir las comparaciones e insumir menor tiempo a los clínicos, ya que la misma escala se usa con todos los pacientes (Slade, 2004). Los instrumentos específicos, en cambio, se administran solo a aquellos pacientes con un determinado trastorno (por ejemplo, ansiedad o

depresión), pero poseen la ventaja de brindar información más precisa para el tratamiento con grupos de poblaciones concretas (Fitzpatrick, Davey, Buxton, y Jones, 1998).

Existen, además, múltiples instrumentos desarrollados por cada una de las diversas dimensiones del cambio que se decida evaluar. El término *dimensión*, en este sentido, se refiere al área que se selecciona para evaluar el cambio del sujeto.

Tradicionalmente, en los estudios de resultados en los servicios de salud mental la dimensión prevalente de evaluación ha sido la sintomatología (Roth y Fonagy, 2005; Slade, 2002a). Por ejemplo, Gilbody et al. (2003) realizaron una revisión de la literatura de las medidas de resultados usadas en ensayos clínicos en psiquiatría entre los años 1955 y 2000, y encontraron que la mayoría usaba escalas basadas en la sintomatología. Las medidas de sintomatología se han utilizado convencionalmente y aun muchos estudios de resultados y servicios las utilizan como única medida (NIMH, 2008).

A lo largo del desarrollo de este campo de estudio el abordaje tradicional de evaluar la sintomatología como única dimensión para medir el cambio ha sido considerado insuficiente, mereciendo las críticas de los clínicos (Fonagy et al., 2005; Trauer, 2010b); esto condujo, a la vez, al interés por medir otras áreas del sujeto.

En 1994 la American Psychological Association y la Society for Psychotherapy convocaron a una conferencia (*Core Battery*) con el objetivo de desarrollar criterios para una batería universal que pudiera ser usada como herramienta en todos los estudios de resultados de psicoterapia (Horowitz, Strupp, Lambert, y Elkin, 1997). En ella se acordaron tres dominios básicos de evaluación: el malestar subjetivo, la discapacidad resultante en las distintas áreas de funcionamiento del sujeto, y la sintomatología.

Slade (2002a), en la revisión sistemática de estudios ROA en los servicios de salud mental de adultos entre los años 1993 a 2000, sintetizó los 16 dominios más utilizados en siete categorías: el bienestar (y la calidad de vida), la cognición/emoción, el comportamiento, la salud física (y discapacidad), el funcionamiento interpersonal y social, los aspectos relacionados con la sociedad (el cuidado y los costos de la seguridad social), y los servicios.

Según la revisión de Trauer (2010b), en la actualidad las principales áreas de evaluación de los instrumentos de resultados en salud mental son: la severidad de la enfermedad, el funcionamiento, la calidad de vida, las necesidades percibidas, y la recuperación.

Otra dimensión que ha planteado interés en los últimos años en el campo de los servicios de salud mental es la productividad asociada a la calidad, criterio que se puede observar en los trabajos de Jacobs (2009), y en el de Hampson, Killaspy, Mynors-Wallis y Meier (2011).

La calidad de vida y la satisfacción son dimensiones que también han tenido relevancia y han sido objeto de varios estudios de evaluación en salud mental, si bien por diferentes motivos en ambos casos algunos autores han presentado reparos para su uso con fines de ROA. En el caso de la calidad de vida, se trata de un aspecto sobre el que tradicionalmente ha existido consenso en considerarlo un objetivo importante a lograr en pacientes con problemas de salud mental. Sin embargo, se ha considerado inapropiado su uso con el objetivo de ROA porque en general, los instrumentos se componen de una cantidad importante de ítems e insumen demasiado tiempo (Andrews et al., 1994; Flynn, Adams, y Alligood, 2002).

La satisfacción del usuario como medida de efectividad de los tratamientos ha dado lugar a controversias (Knaup, Koesters, Schoefer, Becker, y Puschner, 2009); por ejemplo, existen estudios que evidencian que puede no haber relación entre la satisfacción del usuario y el resultado obtenido en el estado de salud (Sperlinger, 2002).

Otra distinción que puede establecerse entre los instrumentos disponibles es según la perspectiva que recogen: los que están basados en la evaluación del clínico (*clinician rated* o *clinician based*) versus los basados en la evaluación de los pacientes (*patient-rated* o *patient-based*), generalmente con auto-reportes. Cada uno de los abordajes presenta sus ventajas y desventajas. Los instrumentos que requieren de la evaluación del clínico presentan problemas como el costo, el entrenamiento y el tiempo. Los auto-informes de los usuarios presentan potenciales problemas, como la deseabilidad social, la dificultad de su uso con pacientes que están cursando un cuadro agudo, y la idiosincrasia personal (Ford, 2005, citado en Jacobs, 2009).

Otro aspecto por el que se puede clasificar a los instrumentos es según el enfoque que pretenda recoger: ideográfico o nomotético según si está enfocado en medir los resultados grupales o los individuales (Aitchison et al., 2007). El abordaje tradicional en la investigación de resultados ha sido el nomotético (Boswell, Kraus, Miller, y Lambert, 2013).

Falta de un método estandarizado para la selección de instrumentos. Como se dijo anteriormente, la selección de los instrumentos es otro de los problemas del campo de la investigación de resultados de salud mental (Slade, 2002a). Se puede observar que se utilizan diversos criterios y procedimientos (Erbes et al., 2004) pero no hay un método estandarizado específico para su selección ni para su revisión (Burlingame et al., 2005; Erbes et al., 2004). Para contribuir a la solución de este problema se han creado en los últimos años guías,

manuales y compendios de medidas de resultados. Un ejemplo de un exhaustivo compendio de medidas de resultados es el del Instituto Nacional de Salud Mental de Inglaterra (NIMH, 2008), que ha sido desarrollado para guiar a los clínicos que se inician en la medición de resultados en la elección más apropiada y relevante según el área.

En contraste con el problema de la gran oferta de instrumentos de resultados disponibles reiterado en la literatura, los países no angloparlantes que pretenden introducir la práctica de ROA se enfrentan a un escaso número de instrumentos que poseen traducción y/o estudios de validación para diferentes lenguajes y culturas. Esta carencia no favorece el desarrollo de la práctica de evaluación de resultados, a la vez que ha llevado a que en muchas ocasiones se utilicen instrumentos validados para otras culturas sin tomar en cuenta las peculiaridades (Prasad-Ildes y Wright, 2004).

2.2. Requisitos de los instrumentos de ROA en salud mental

Relevancia, utilidad. De manera ideal, las medidas deben resultar útiles y brindar información relevante para todos aquellos que van a utilizarlas: clínicos, usuarios, directores de servicios, administración pública. Considerando su uso a nivel nacional, es necesario que una medida pueda ser aplicada en un amplio rango de poblaciones y en diferentes ámbitos, permitiendo las comparaciones (Fitzpatrick et al., 1998; Horowitz et al., 1997).

En cuanto al contenido, en varios de los trabajos donde se discuten los criterios que deberían tener las medidas de ROA se constata la necesidad de incluir medidas que reflejen distintos aspectos del individuo (Andrews et al., 1994; Erbes et al., 2004; Horowitz, et al., 1997; Trauer 2010b). Las medidas que toman en cuenta solo un aspecto del sujeto como la sintomatología, han sido criticadas por los clínicos por considerarlas poco útiles (Gilbody et

al., 2003), dando lugar a un mayor énfasis en las evaluaciones más globales que reflejen diferentes aspectos del ser humano (Slade, 2002a).

Adecuadas propiedades psicométricas. Otro de los requisitos indiscutible para la selección de instrumentos de resultados en salud mental, es que cuenten con adecuadas propiedades psicométricas. Además de las clásicas propiedades de validez y fiabilidad, una propiedad fundamental para los instrumentos de resultados es la sensibilidad al cambio (Andrews et al., 1994; Erbes et al., 2004; Flynn et al., 2002; Horowitz et al., 1997; NIMH, 2008; Schmidt, Garratt, y Fitzpatrick, 2000; Smith et al., 1997; Sperlinger, 2002).

La sensibilidad al cambio en un instrumento es un indicador de la capacidad para registrar el cambio producido en la variable que se pretende medir a lo largo del tiempo.

Para medir los resultados en salud mental, es importante que el instrumento permita detectar los cambios que se producen como consecuencia del tratamiento que reciben los pacientes.

Los métodos tradicionales para valorar los cambios como resultado de los tratamientos han sido las pruebas de significación estadísticas y las medidas del tamaño del efecto (Cohen, 1988). El tamaño del efecto (Cohen, 1988), se utiliza para estimar la magnitud de la diferencia entre medidas de resultados obtenidas en distintos momentos temporales.

Para evaluar el cambio a nivel individual, uno de los métodos más usados es el Índice de Cambio Fiable (ICF) que informa si la magnitud del cambio observado excede el rango que podría ser atribuido al monto probable por error o azar. Este método posee la ventaja de permitir el análisis del cambio a nivel de las personas concretas y saber cuál es el porcentaje de las personas que cambian dentro de un grupo.

Los contrastes estadísticos de significación han sido criticados por los clínicos, quienes argumentan que identificar diferencias estadísticamente significativas no supone necesariamente un cambio clínico relevante (Cohen, 1988; Jacobson y Truax, 1991). Esto ha llevado al interés en evaluar la significación clínica de los cambios observados.

Jacobson y Truax (1991) proponen dos pasos para valorar el Cambio Clínico Significativo (CCS): en primer lugar, estimar el cambio mediante el Índice de Cambio Fiable y, en segundo lugar, estimar el cambio significativo que refiere a si el sujeto ha pasado a estar más cerca de la media del grupo de población funcional que de la del grupo no funcional. Dicho de otro modo, si se desplaza del puntaje de una población clínica hacia una “no-clínica”. Ambos criterios deben cumplirse para establecer que se ha producido un CCS de acuerdo a estos autores. El problema que reporta el CCS es que para poder calcularlo se debe poseer los datos de la población “no clínica”, los cuales no siempre están disponibles.

Viabilidad (*feasibility*). Existe un amplio consenso en la literatura acerca de que una cualidad fundamental para un instrumento de evaluación de resultados es la *feasibility*, que puede traducirse al castellano como viabilidad. Su estudio es importante porque no es lo mismo una medida de resultados cuando es utilizada por académicos o investigadores en un contexto de investigación, que cuando es utilizada por los clínicos con el objetivo de monitorear rutinariamente el progreso de sus pacientes (Slade, 2004). Los investigadores, en general, están entrenados para usar medidas de resultados y cuentan con tiempo, dedicación e infraestructura para realizar la tarea; pero esto no es lo habitual de ROA (Slade, 2004). Por el contrario, la ROA involucra “pedirle a clínicos que están ocupados, y que tienen poco interés o entrenamiento en usar medidas de resultados, que agreguen a su trabajo ya sobrecargado el llenado de formularios, en una cultura donde el llenado de formularios es visto como tiempo

muerto” (Slade, 2004, pp. 62–63). La viabilidad indica que la medida es adecuada, relevante y sustentable para su uso de manera rutinaria, cuando se usa con un propósito específico en ámbitos de atención clínicos típicos (Slade, Thornicroft, y Glover, 1999). Para poder evaluarla se debe considerar el contexto: dónde, cuándo, para qué, quién, y a quién se realiza la evaluación. (Slade et al., 1999).

Diversos autores señalan la falta de una definición consistente y consensuada del término *feasibility* y de cómo medirlo (Fitzpatrick et al., 1998; Slade et al., 1999). Andrews et al. (1994) identificaron tres dimensiones de viabilidad: *la aplicabilidad, la aceptabilidad y la practicidad*. La *aplicabilidad* se define como el grado en que toma en cuenta las dimensiones importantes para el evaluador, el significado para los propósitos. La *aceptabilidad* alude a la facilidad con que se la puede utilizar. La *practicidad* se refiere al costo de implementación, en términos del entrenamiento que requiere el instrumento y la complejidad en puntuar e interpretar los datos.

En la revisión de Pirkis, Burgess, Kirk, et al. (2005) el término *viabilidad* es usado de manera equivalente a *utility*, indicando el grado en que un instrumento resulta aceptable y útil para los involucrados.

Según Fitzpatrick et al. (1998), la propiedad viabilidad implica que el instrumento sea breve, sencillo de aplicar y que no resulte una carga para el clínico.

Thornicroft y Slade (2000) sugieren operacionalizar el concepto de la siguiente manera: “que sea corto, simple, relevante, aceptable, disponible y valorable” (p. 84).

Otro criterio de viabilidad, según algunos autores, es la disponibilidad del instrumento, en el sentido de que sea accesible, que su uso no implique un costo económico o que, en todo caso, este sea mínimo (Andrews et al., 1994; Horowitz, et al., 1997).

Ante tantos requisitos para las medidas de ROA se ha afirmado que no existe, ni podrá existir, la medida “ideal” que pueda servir para medir los resultados de los distintos ámbitos de los servicios de salud mental, ni tampoco un sistema de medición único que resuelva la complejidad del asunto (Mellsop y Wilson, 2006; NIMH, 2008). Sin embargo, algunos grupos de investigación, muchas veces con el apoyo de iniciativas gubernamentales, han aceptado el desafío de desarrollar instrumentos que intenten cumplir los requisitos de ROA y que puedan ser usados a escala nacional. La HoNOS, es uno de ellos.

2.3. Principales instrumentos de evaluación rutinaria de resultados en salud mental

Se identifican en este apartado algunos de los principales instrumentos para ROA en los servicios de salud mental de adultos.

La revisión de Andrews et al. (1994), que constituyó un trabajo de referencia para Australia, realizó una revisión exhaustiva de los instrumentos disponibles. Los instrumentos recomendados en la revisión fueron the Behaviour and Symptom Identification Scale (BASIS-32) (Eisen, Dill, y Grob, 1994), the Mental Health Inventory (MHI) (Veit y Ware, 1983), the Medical Outcomes Study 36-Item Short-Form Survey (SF-36) (Ware y Sherbourne, 1992), the Health of the Nation Outcome Scales (HoNOS) (Wing et al., 1996; Wing et al., 1998), the Life Skills Profile (LSP) (Rosen, Hadzi-Pavlovic, y Parker, 1989), y the Role Functioning Scale (RFS) (McPheeters, 1984).

Stedman, Stedman, Yellowlees, Mellsop, Clarke, y Drake (1997) pusieron a prueba en Australia los seis instrumentos recomendados por Andrews et al. (1994) y estudiaron sus propiedades psicométricas. Luego del estudio recomendaron para el entorno australiano la HoNOS como medida para el clínico y la MHI como medida del usuario.

The Global Assessment Functioning (GAF), que equivale al eje V del DSM IV (American Psychiatric Association, 1994), es la reportada por algunos autores como la medida más usada para evaluar el funcionamiento en los pacientes psiquiátricos. (Burlingame et al., 2005; Flynn et al., 2002).

Otros instrumentos explícitamente desarrollados con el propósito de ROA son the Threshold Assessment Grid (TAG) (Slade, Powell, y Rosen, 2000), The Clinical Outcomes in Routine Evaluation - Outcome Measure (CORE- OM) (Evans et al., 2002) y The Outcome Questionnaire- 45 (OQ-45) (Lambert et al., 2004). En la Tabla 1 se describen las características principales de cada instrumento: la perspectiva que recoge (del clínico o del usuario), los objetivos, la cantidad de ítems que contienen y el contexto de aplicación. Estos instrumentos cumplen en distinto grado con los requisitos para ROA en su condición de ser breves, evaluar diferentes aspectos del ser humano, sus propiedades psicométricas han sido ampliamente estudiadas, y poder aplicarse en una amplia gama de poblaciones con diversos diagnósticos y problemas de salud mental.

Entre las medidas disponibles en la actualidad, la HoNOS posee una serie de ventajas que se describen en profundidad en el apartado 4.

Tabla 1

Principales instrumentos que cumplen con los requisitos de ROA

Instrumento	Perspectiva	Objetivo	N° Ítems	Aplicación
BASIS-32	Usuario	Resultados en salud mental	32	SSM: Segundo y tercer nivel
CORE-OM	Usuario	Malestar psicológico y severidad de problemas	34	Psicoterapia, Counselling: Primer y segundo nivel
GAF	Clínico	Funcionamiento global	1	SSM: Segundo y tercer nivel
HoNOS	Clínico	Resultados en salud mental	12	SSM Segundo y tercer nivel
LSP-39	Usuario	Funcionamiento general y discapacidad	39	SSM: Segundo y tercer nivel
MHI	Usuario	Bienestar/malestar psicológico	38	Población general
OQ-45	Clínico	Resultados en psicoterapia	45	Psicoterapia: Primer y segundo nivel
TAG	Clínico	Severidad de problemas de salud mental	7	Todos los niveles

Nota. BASIS-32 = the Behaviour and Symptom Identification Scale (Eisen et al., 1994); CORE-OM = Clinical Outcomes in Routine Evaluation - Outcome Measure, (Evans et al., 2002); GAF = Global Assessment Functioning (American Psychiatric Association, 1994); HoNOS = the Health of the Nation Outcome Scales (Wing et al., 1996, 1998); LSP-39 = the Life Skills Profile (Rosen et al., 1989); MHI = the Mental Health Inventory (Veit y Ware, 1983), OQ-45 = Outcome Questionnaire (Lambert et al., 2004); TAG = the Threshold Assessment Grid (Slade et al., 2000). SSM = Servicios de Salud Mental.

3. Prácticas Actuales de ROA en los Servicios de Salud Mental

En el siguiente apartado se sintetiza las prácticas actuales de ROA destacadas a nivel internacional y en Uruguay. Se presentan, por último, algunos de los desafíos actuales que se plantean en esta área.

3.1. Nivel internacional

La ROA ha sido implementada en diferentes lugares del mundo (Jacobs 2009; Trauer, 2010a). Según la revisión de Trauer (2010a) los países con mayor desarrollo en medición de resultados en salud mental son Alemania, Australia, Canadá, Estados Unidos, Italia, Noruega, Nueva Zelanda y el Reino Unido. Este autor advierte que podría haber algún otro país, pero estos serían los que poseen un desarrollo más claro en este campo y en los que el tema ha sido reportado a través de diversos medios.

En Inglaterra, Nueva Zelanda y Australia pueden ubicarse importantes iniciativas, y esfuerzos para adoptar ROA en los servicios de salud mental (Gilbody et al. 2002a), y la utilización rutinaria de medidas de resultados en los servicios de salud mental es obligatoria; entre ellas, la HoNOS ocupa un lugar central.

Australia es considerada por varios autores como el líder en ROA en los servicios de salud mental, por sus importantes avances y logros (Jacobs, 2007; Pirkis, Burgess, Coombs, et al., 2005; Slade, 2002b; Trauer, 2010a).

El ejemplo más temprano de investigación de resultados en salud mental aparece en Estados Unidos: el Medical Outcomes Study (MOS) conducido por RAND Corporation a fines de los años 80 (Tarlov et al., 1989; Wells et al., 1989, 1996, citados en Gilbody et al.,

2002a). Una de sus contribuciones fundamentales fue desarrollar una medida de salud centrada en el paciente, que luego derivó en la Short- Form 36 (SF36). (Stewart, Hays y Ware, 1988)

Según Jacobs (2007), en los Estados Unidos “la medición de resultados está siendo implementada de manera creciente en programas públicos y privados” (p. 16). La incidencia del factor económico y las presiones por los gastos en salud fueron los impulsores de la práctica de ROA en este país, llevando al énfasis creciente en las medidas de resultados (Jacobs, 2007, 2009). Sin embargo, a pesar del impulso inicial con que la investigación de resultados fue adoptada, la ROA no ha tenido el impacto esperado (Gilbody, House y Sheldon, 2002b).

Respecto a los países de América Latina, la revisión de Trauer (2010a) sobre los países que implementan ROA a nivel internacional observó que solo en algunos países del hemisferio norte hay un desarrollo claro, o al menos, publicaciones sobre el tema a través de diversos medios.

Una de las principales conclusiones planteadas por De la Parra (2013) en su revisión sobre la investigación en psicoterapia en América Latina fue la existencia de obstáculos y debilidades para investigar y publicar en la región. Un indicador de ello son los estudios bibliométricos que muestran que América Latina representa aproximadamente el 5% de la producción mundial en psicología, expresado en el número de artículos aparecidos en The Web of Science entre 1999 y 2004 (Nacarrete-Cortes, Quevedo-Blasco, Chaichio-Moreno, Ríos, y Buela-Casal, citado en De la Parra, 2013).

La revisión de Saldivia, Grandón y Vicente (2012) sobre la situación de Chile en particular arriba a conclusiones similares, evidenciando la persistencia de un déficit en la

investigación, especialmente sobre los tratamientos y los resultados de los servicios de salud mental.

Para poder actualizar la información sobre el estado del tema en América Latina para el presente trabajo se revisaron bases de datos bibliográficas y literatura gris utilizando palabras clave en español e inglés, se contactó a cuatro asesores de la Organización Panamericana de la Salud en el área de salud mental y a siete coordinadores de los países del Capítulo Latinoamericano de la Sociedad de Investigación en Psicoterapia (SPR). También se envió un correo de solicitud de información a toda la lista de miembros internacional de la Sociedad de Investigación en Psicoterapia (SPR). No fue posible encontrar estudios con el objetivo de ROA a escala nacional, ni información sobre ninguna iniciativa gubernamental al respecto por parte de ningún país de América Latina.

El problema de la insuficiente investigación en salud mental en general, y los limitados recursos destinados a los investigadores de América Latina ha sido reconocido por diversos documentos de la Organización Panamericana de la Salud. Se afirma: “La salud mental enfrenta no solo una brecha en las respuestas de los servicios a las necesidades de la población sino también en la capacidad para generar, analizar y utilizar información necesaria para la toma de decisiones” (OPS, 2011, p. 130).

A pesar de no encontrarse evidencias de utilización de instrumentos de resultados de manera rutinaria en América Latina, se puede constatar un creciente aumento de estudios de validación de algunos de los instrumentos principales de ROA comentados en la sección anterior (por ejemplo, el OQ-45 en Chile por De la Parra y Bergen, 2000, y en Argentina por Fernández-Álvarez, Hirsh, Maristany, y Torrente, 2005).

3.2. Uruguay

Uruguay comparte las mismas características y problemas en este tema que sus contrapartes de la región.

La actualización del Programa Nacional de Salud Mental (PNSM) del Ministerio de Salud Pública (MSP) ubicó la evaluación de los servicios de salud mental entre las mayores dificultades para el país “El retraso de los sistemas de registro, elaboración de datos, evaluación de resultados, construcción de indicadores y controles de calidad”. (Ministerio de Salud Pública [MSP], 2005, p. 157). En este mismo documento programático nacional se planteó como prioritario fortalecer la investigación y evaluación de todos los planes y programas de asistencia en desarrollo, de manera de que pueda ser implementada en todos los servicios.

En el año 2011 el Ministerio de Salud Pública (MSP, 2011) implementó un plan que incorporó de manera obligatoria para todos los servicios, públicos y privados, del país nuevas prestaciones de salud mental en el sistema de salud. Las nuevas prestaciones incorporadas refieren a abordajes psicoterapéuticos y psicosociales que se agregaron a las prestaciones de psiquiatría tradicionales existentes. El plan se plantea como objetivos asegurar el acceso universal y equitativo a los servicios de salud mental (MSP, 2011). Para evaluar las nuevas prestaciones el Programa Nacional de Salud Mental (PNSM) definió un conjunto mínimos de indicadores y de datos a ser recolectados por todos los servicios a nivel nacional. En la primera etapa se incluyeron datos sobre estructura y en una segunda etapa se sumaron datos sobre los procesos de las intervenciones. La evaluación sobre resultados y el establecimiento de los indicadores correspondientes constituyen un objetivo pendiente, que se encuentra en

proceso de definición (Ministerio de Salud Pública [MSP], 2013). Además, en el sistema de recolección de datos actual no se utilizan instrumentos estandarizados.

Por otra parte, el único instrumento que posee un estudio de validación para el país es el Symptom Check List (SCL-90 R) (Najson, 2008). Sin embargo, el SCL-90 R no cumpliría con los requisitos de ROA descritos anteriormente, dado que mide solo sintomatología y no es breve.

Un campo con un desarrollo en aumento en la investigación de resultados es el de los servicios de psicoterapia. Desde la década del 90 se observa una creciente actividad de investigación empírica en psicoterapia (Montado, Palermo, Fernández y Gerpe, 2011) y el interés por parte de diversos grupos e instituciones de utilizar instrumentos para evaluar las psicoterapias que se brindan. Sin embargo, dentro de estos estudios algunos constituyen proyectos de investigación puntual, mientras que otros se enmarcan en servicios que poseen el objetivo de medición rutinaria de resultados pero utilizan instrumentos que no cumplen con los requisitos definidos para ROA.

Respecto a los servicios que utilizan instrumentos con el objetivo de ROA, el de más larga data es el programa de investigación de la Clínica Psiquiátrica de la Facultad de Medicina de la UdelaR (Fernández Castrillo, Gerpe, Montado, y Palermo, 2010; Montado et al., 2011). Para evaluar los tratamientos el programa incluye instrumentos ad hoc desarrollados por el equipo de investigación y protocolos de Howard (1988-1993) del Programa de Psicoterapia del Instituto de Psiquiatría de Northwestern Memorial Hospital (Montado et al., 2011).

Otra experiencia de utilización de instrumentos con el objetivo de ROA es la implementada por la Facultad de Psicología de la UdelaR, en el Servicio de Atención

Psicológica Preventivo - Asistencial. En esta se utilizó desde su creación una Encuesta de Satisfacción de Usuarios y un Formulario de Evaluación del Proceso Psicoterapéutico de los Profesionales Externos, desarrollados por el propio servicio (Zytner, 2011). Además, desde el año 2006 el servicio incorporó de manera rutinaria la evaluación de los resultados con instrumentos validados (Zytner, 2011). Se utilizan el OQ-45 en su versión validada para Chile (De la Parra y Von Bergen, 2000) y el SCL90-R para los usuarios mayores de 18 años. Como se dijo anteriormente, si bien el SCL-90 R posee el valor de ser el único instrumento con estudios de validación para el país no cumple con los requisitos para ROA.

3.3. Desafíos actuales

A pesar del importante impulso de ROA y el desarrollo importante en algunos países, la implementación de ROA no ha proliferado en la medida que se pretendía (Jacobs y Moran, 2010).

La experiencia de los países con mayor desarrollo en el tema (Australia, Inglaterra, Nueva Zelanda) ha demostrado que para una efectiva ROA no es suficiente contar con el instrumento adecuado. Tal como plantean algunos autores, incluso el mejor instrumento puede ser inútil si hay fallas en la implementación (Milne, Reichelt, y Wood, 2001). Ese parece ser uno de los desafíos más importantes en la actualidad para los países que ya han logrado avanzar en la validación de los instrumentos.

Por este motivo, una de las reiteradas propuestas actuales en los sistemas de medición de ROA es prestar mayor atención a la implementación (Coombs y McKay, 2012; Milne et al., 2001; Stedman et al., 1997). Lograr que el uso rutinario de medidas de resultados en la práctica clínica se implemente en el sistema de salud es una tarea compleja que tiene

múltiples determinantes, y que se ha enfrentado en general a varias dificultades y barreras (Duncan y Murray, 2012; Fonagy et al., 2005; Jacobs, 2009).

Diversos autores (Coombs y McKay, 2012; Milne et al., 2001; Stedman et al., 1997) coinciden en plantear que la estrategia de implementación debe tomar en cuenta las dificultades y barreras que pueden encontrarse a nivel de los clínicos, de los usuarios y de toda la organización. Una de las principales barreras que se han identificado en varios estudios ha sido la falta de motivación y compromiso de los involucrados (Fonagy et al., 2005). Varios estudios en diversos países (Fonagy et al., 2005; Jacobs, 2009) observaron actitudes mixtas y contradictorias por parte de los clínicos hacia la utilización de ROA. Son diversos los motivos que explicarían esto. Tanto Gilbody et al. (2002b) como Slade et al. (2006) plantean que esto puede relacionarse a una falta de convencimiento por parte de los clínicos sobre la efectividad de ROA, y a la dificultad para ver la relevancia para su práctica profesional y los beneficios en la atención a sus pacientes. Similares conclusiones han sido planteadas por otros estudios. (Jacobs, 2009; Meehan, McCombes, Hatzipetrou, y Catchpole, 2006). Otra de las explicaciones ha sido que existe una concepción tradicional en salud mental de que los instrumentos estandarizados no son necesarios porque alcanza con el juicio clínico (Duncan y Murray, 2012; Slade, 2006). Coombs y McKay (2012), Jacobs (2009) y Slade (2012), coinciden en señalar que la resistencia se relaciona con el cambio cultural que supone usar instrumentos en la práctica clínica, generando un “choque cultural”.

Otros estudios observaron como motivos de la resistencia por parte de los clínicos la percepción de que la iniciativa de ROA se relaciona con motivos de gestión o de control del gobierno (Trauer y Buckingham, 2006).

En varias revisiones se señala como barrera las carencias a nivel organizacional, como, por ejemplo, la falta de infraestructura que apoye el trabajo, sobre todo en lo que refiere a brindar el soporte informático adecuado (Slade, 2006; Trauer y Buckingham, 2006; Jacobs, 2009), no prever el tiempo que implica ROA (lo que lleva a sobrecargar al clínico con una tarea que se considera que va en desmedro de la actividad clínica) (Slade, 2006), y la falta de incentivos económicos o profesionales para realizarlo (Slade, 2004;).

Algunas de las recomendaciones identificadas en la revisión de la literatura para superar estos obstáculos en la implementación de ROA, han sido la necesidad de lograr motivación, compromiso y participación positiva por parte de los clínicos y de los demás involucrados en la medición: usuarios, familiares, administradores y toda la organización (Duncan y Murray, 2012; Fonagy et al., 2005; Jacobs, 2009; Pirkis, Burgess, Coombs, et al., 2005).

Una herramienta para lograr esto, en la que vienen haciendo énfasis varios estudios y revisiones, es en la devolución (*feedback*) de información a los clínicos sobre los datos recolectados (Jacobs, 2009; Jacobs y Moran, 2010; Knaup et al., 2009; Newnham y Page, 2010; Slade, 2004; Slade et al., 2006). Se recomienda también proveer la infraestructura adecuada y el apoyo a nivel del sistema de salud en las distintas etapas (recolección, análisis y devolución de los resultados). Se requiere realizar una inversión en el sistema de información, brindar tiempo y entrenamiento a los clínicos (Coombs y McKay, 2012; Jacobs, 2009; Slade, 2006; Slade, 2002b; Thornicroft y Slade, 2000). Y, en concordancia con lo dicho anteriormente sobre el choque cultural que implica el uso de instrumentos, se insiste en trabajar en el cambio de cultura, generando una cultura que valore la medición (Coombs y McKay, 2012).

4. The Health of the Nation Outcome Scales (HoNOS)

La escala HoNOS fue desarrollada en la década del 90 en Gran Bretaña por la Unidad de Investigación de The Royal College of Psychiatrists' Research Unit en el marco del Proyecto denominado "Health of the Nation Strategy", del Departamento de Salud (DS) del Reino Unido. En 1992 el DS identificó cinco áreas prioritarias para reducir la morbilidad y mortalidad de la población, entre ellas la salud mental. La escala se desarrolló con el propósito de constituirse en una medida de resultados sensible, con una amplia cobertura de problemas que pudiera ser utilizada de manera rutinaria en los servicios de salud mental de adultos para cuantificar la mejoría de la salud y el funcionamiento social de las personas con enfermedades mentales (Pirkis, Burgess, Kirk, et al., 2005; Wing et al., 1998).

4.1. Características

Se trata de una medida basada en el juicio del clínico (*clinician rated*). Contiene 12 ítems que cubren una gama de problemas que incluye aspectos biológicos, psicológicos y sociales (ver Tabla 2). Ofrece una gradación de la gravedad entre 0 a 4 para cada ítem, en la cual a mayor puntuación corresponde mayor gravedad. La puntuación máxima global en la escala es de 48. La severidad se mide 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).

Según la revisión realizada por Delaffon, Anwar, Noushad, Ahmed, y Brugha (2012) el tiempo reportado que se requiere para completarla es entre 2-3 minutos. Algunos autores reportan tiempos algo mayores: entre 5- 15 minutos (McClelland, 2000), y de 15 a 30

minutos (Pirkis, Burgess, Kirk, et al., 2005). Las variaciones dependen de las diferencias entre los clínicos, la familiaridad que tengan con la escala y la disponibilidad de la información necesaria (Delaffon et al., 2012).

Tabla 2

Ítems que componen HoNOS

Ítems
1. Conducta hiperactiva, agresiva, disruptiva o agitada
2. Autolesiones no accidentales
3. Consumo problemático de alcohol o drogas
4. Problemas cognitivos
5. Problemas por enfermedad física o discapacidad
6. Problemas asociados a la presencia de ideas delirantes y alucinaciones
7. Problemas en relación con el humor depresivo
8. Otros problemas mentales o conductuales
9. Problemas con las relaciones
10. Problemas con las actividades de la vida cotidiana
11. Problemas con las condiciones de vida
12. Problemas en relación con la ocupación y las actividades

Nota. Versión al castellano de Uriarte et al. (1999).

Para puntuar la HoNOS se deben considerar los problemas mas 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. El instrumento incluye un extenso glosario con instrucciones detalladas para realizar la evaluación. En el caso de que no sea utilizada por el clínico que brinda el tratamiento, previamente es necesario tener una entrevista semi-

estructurada. La evaluación se puede complementar utilizando información de otras fuentes (por ejemplo, historia clínica, otros profesionales, familiares).

La HoNOS ha sido diseñada para utilizarse con pacientes con trastornos mentales de los servicios de salud mental de segundo y tercer nivel, con una amplia gama de diagnósticos (Andreas et al., 2011) y servicios de atención: ambulatorio, de internación psiquiátrica, residencial. También se la puede utilizar en la evaluación en el primer nivel de atención, aunque su uso con pacientes menos severos está menos estudiado (Pirkis, Burgess, Kirk, et al., 2005). El estudio de Audin, Margison, Clark y Barkham (2001) encontró poca utilidad para evaluar a pacientes atendidos en los servicios de psicoterapia que presentan patologías menos severas.

La HoNOS original se desarrolló para los servicios de salud mental de adultos, en edad laboral, de 18 a 64 años inclusive. Posteriormente, se desarrollaron versiones para otras edades: HoNOSCA (versión para niños y adolescentes); HoNOS 65+ (versión para adultos mayores). También existen versiones para otros servicios específicos: HoNOS-secure (para servicios forenses) (Dickens, Sugarman, y Walker, 2007), HoNOS-LD para personas con discapacidad intelectual (Roy, Matthews, Clifford, Fowler, y Martin, 2002), HoNOS-ABI para personas con daño cerebral adquirido (Fleminger et al., 2005). Otras modificaciones de la HoNOS se produjeron en una versión de autoinforme en el marco del paradigma de recuperación (Stewart, 2009, citado en Delaffon et al., 2012).

Algunos autores y revisores (Delaffon et al., 2012; Wing et al., 1998) consideran el entrenamiento en la escala un factor importante para la fiabilidad de las puntuaciones, mientras que otros autores sugieren que esto no es una condición previa indispensable para utilizarla (Andreas et al., 2010). Existe evidencia internacional que sugiere que la puntuación

de HoNOS es menos fiable cuando se completa por los clínicos que cuando la evaluación la realizan investigadores (Lovaglio y Monzani; 2011b). Otros autores han encontrado que la experiencia en el uso de la Escala y las circunstancias en las cuales los datos se obtienen afectan también la fiabilidad (Macdonald, 2002 citado en Turner, 2004).

No existe un formato de entrenamiento estandarizado para la HoNOS. El Royal College of Psychiatrist del Reino Unido (RCP) ha desarrollado una serie de materiales que se utilizan en una jornada de un día, mientras que Australia desarrolló otros materiales para el entrenamiento de los clínicos de los servicios públicos de salud mental de Victoria que se dictan en una jornada de medio día (Trauer, y Buckingham, 2006).

4.2. Uso a nivel internacional

Internacionalmente la HoNOS ha pasado a ser un instrumento relevante entre los evaluados por los clínicos (*clinician-rated*) para mediciones diferenciadas de severidad, y de resultados en pacientes con una amplia gama de diagnósticos (Andreas et al., 2011). Es una de las escalas actualmente más utilizadas en ROA en los servicios de salud mental británicos, en todo el territorio australiano, en Nueva Zelanda y, en mayor o menor medida, en otros países como Canadá, Dinamarca, Francia, Italia, Alemania, Holanda y Noruega (Andreas et al., 2011; Pirkis, Burgess, Kirk, et al., 2005). En el Reino Unido, Australia y Nueva Zelanda es parte del conjunto mínimo de datos nacionales que deben recolectar los servicios de salud mental, por lo cual su uso es obligatorio.

Tanto el gobierno de Australia como el de Nueva Zelanda establecieron el uso de la familia de escalas HoNOS como el componente central de ROA en el sistema de salud mental (Coombs, Stapley, y Pirkis, 2011).

En los servicios de atención a adultos de Australia, se exige la HoNOS junto a otras escalas en todas las ocasiones de atención y en distintos tipos de servicios (Trauer, 2010a).

En Inglaterra se estableció la recolección de datos obligatoria con la Escala HoNOS para los servicios públicos de salud mental (NHS) desde el año 2003. En 2005 el NIMHE estableció un grupo denominado Outcomes Reference Group con el objetivo de identificar potenciales barreras para una sólida implementación nacional de ROA, y recomendar los pasos básicos que debían darse para su inicio (Fonagy et al., 2005). El informe recomendó empezar con las medidas clínicas de morbilidad y la familia de HoNOS: “El conjunto de medidas de HoNOS (incluyendo HoNOSCA, HoNOS65+ and HoNOS-LD) puede proveer un anclaje de desarrollo, y establecer los dominios mínimos para un número de poblaciones clave” (p. 21). En el informe también se recomendó utilizar HoNOS en todos los usuarios de los servicios, realizando “al menos una medición por año” (p. 12). Las recomendaciones de este grupo fueron luego adoptadas por el gobierno (Jacobs, 2009).

Ha sido traducida a varios idiomas, incluyendo alemán, checo, danés, francés, español, italiano y tailandés (Delaffon et al., 2012; Pirkis, Burgess, Kirk, et al., 2005).

Existen distintos estudios sobre sus propiedades psicométricas realizados en diversos países. Los países en los que para este trabajo se encontró al menos una publicación en la que se analizan las propiedades de la HoNOS son Alemania (Andreas et al., 2007, 2010), Australia (Stedman et al., 1997; Trauer et al., 1999), Dinamarca (Bech et al., 2006), Francia (Lauzon et al., 2001), Holanda (Mulder et al., 2004), Irlanda (Browne, Doran, y McGauran, 2000), Italia (Lora, Bezzi, y Erlicher, 2007; Lovaglio, y Monzani, 2011b; Parabiaghi, Barbato, D’Avanzo, Erlicher, y Lora, 2005), Canadá (Kisely, Campbell, Crossman, Gleich, y Campbell, 2007; Kisely, Campbell, Cartwright, Cox, y Campbell, 2010), Reino Unido

(Bebbington, Brugha, Hill, Mardsen, y Window, 1999; McClelland, 2000; Wing et al., 1998), Suiza (Koch, Wasserfallen, Camus, Gebel, y Bonsack, 2013) y Tailandia (Phuaphanprasert, Srisurapanont, y Silpakit, 2007).

Uriarte et al. (1999) realizaron la traducción al castellano en 1999 mediante el método de traducción inversa, con la colaboración del RCP, pero no existen hasta el momento estudios empíricos sobre las propiedades psicométricas de esta versión. No se tiene conocimiento de ningún estudio de validación de la HoNOS original en ningún país de América Latina ni en España (Uriarte, comunicación personal, 2 de octubre de 2012). A pesar de eso existe experiencia de estudios en España que la utilizan para evaluar tratamientos (por ejemplo, Iglesias, Illescas, Lucio, Santamarina, y Alonso, 2005). En los servicios de salud mental públicos de Bilbao se utiliza de manera habitual la versión de Uriarte et al. (1999) para la evaluación de todos los planes de tratamiento de personas con trastorno mental severo. La HoNOS está integrada al conjunto de datos mínimos que se recogen en un protocolo estructurado en los distintos servicios de salud mental (ambulatorios y hospitalarios) (Matthies, Uriarte, Martínez, Rodríguez, y López, 2012).

Según la revisión realizada por Delaffon et al. (2012), si bien se han hecho estudios de validez y confiabilidad de la HoNOS utilizada en otros idiomas, se requieren más estudios.

4.3. Propiedades psicométricas

Al finalizar las cuatro fases de testeo de la escala original en el Reino Unido, Wing et al. (1998) concluyeron:

HoNOS es simple de usar y en general aceptable clínicamente; cubre una amplia gama de problemas clínicos y disfunciones sociales; es sensible al cambio o la falta de él a lo

largo del tiempo; posee aceptable confiabilidad y es compatible con instrumentos que están bien establecidos desde hace más tiempo (Wing, et al., 1998, p. 16).

Luego de esta afirmación se han realizado muchos estudios: la revisión de Delaffon et al. (2012) encontró 585 artículos que utilizan alguna versión de HoNOS. Entre tantos estudios se puede observar una amplia gama de diseños en cuanto a tamaño de la muestra utilizada (desde algunas decenas a miles de usuarios), tipo de población, alcance (uno o varios servicios, regional, nacional), así como las propiedades que analizan. En general la mayoría de los estudios sobre sus propiedades se centraron en la fiabilidad, validez de criterio y concurrente más que en sus propiedades longitudinales como sensibilidad al cambio y validez predictiva (Kisely et al., 2010; Pirkis, Burgess, Kirk, et al., 2005).

Pirkis, Burgess, Kirk, et al., (2005) realizan una extensa revisión de las propiedades psicométricas (evaluadas en la mayoría de los estudios desde el modelo de la Teoría Clásica de los Tests), incluyendo todos los trabajos que se realizaron con la escala desde su creación hasta el año 2005. Las principales conclusiones se muestran en la Tabla 3. Este trabajo puede considerarse clave porque es la única revisión sobre las propiedades y aparece citado en las múltiples publicaciones posteriores, entre ellas la revisión sobre el uso de la escala en psiquiatría llevada adelante por Delaffon et al. en 2012.

Tabla 3

*Evidencia de las propiedades psicométricas de HoNOS según revisión de Pirkis **

Propiedad	Análisis	Valoración
Validez	Contenido	Buena
	Constructo	Buena
	Concurrente	Buena
	Predictiva	Buena
Confiabilidad	Inter-jueces	Adecuada
	Test-retest	Adecuada
Sensibilidad al cambio		Adecuada
Viabilidad		Adecuada

Nota. * Adaptada de Pirkis, Burgess, Kirk, et al., (2005, p. 85).

En relación con la validez de contenido, según Pirkis, Burgess, Kirk, et al., (2005) en la mayoría de los estudios la valoración fue positiva, lo que sugiere que el instrumento es apropiado y está bien diseñado. Sin embargo, destacan los autores, algunos ítems han reportado algunos inconvenientes para su aplicación, entre ellos el ítem 8 (Otros problemas mentales o conductuales), que restringe al clínico a indicar solo un problema, y en la práctica clínica es frecuente que se presenten más de uno simultáneamente. Las instrucciones del glosario son: “*Puntuar únicamente el problema clínico más severo que no haya sido*

considerado en los ítems 6 y 7 como sigue. Especificar el tipo de problema consignando la letra apropiada: A fóbico; B ansiedad; C obsesivo-compulsivo; D sobrecarga mental y tensión; E disociativo; F somatoforme; G alimentación; H sueño; I sexual; J otros, especificar”.

Otros ítems que pueden presentar dificultad para otorgar una puntuación son los sociales (10-12) cuando no hay disponible suficiente información, lo que sucede en varias ocasiones –especialmente en los pacientes que se encuentran internados (Trauer et al., 1999). Además algunos autores han criticado a la HoNOS por no incluir en la valoración otros factores como la cultura, la pobreza, el riesgo y la adherencia a la medicación (Pirkis, Burgess, Kirk, et al., 2005).

Respecto a la estructura factorial de la HoNOS, los resultados de los distintos estudios han sido contradictorios (Lovaglio y Monzani, 2011a; Speak y Hay, 2012). Distintas publicaciones y revisiones mencionan un debate respecto a la dimensionalidad subyacente de la HoNOS (Lovaglio y Monzani, 2011a; Pirkis, Burgess, Kirk, et al., 2005; Speak y Hay, 2012). Algunos autores han sugerido que la HoNOS no debería ser considerada como medida de un solo constructo subyacente (Trauer et al., 1999).

El reporte del proceso de desarrollo y validación de la escala original (Wing et al., 1998) concluyó que “el total ha probado ser robusto durante las 4 etapas de las pruebas y los 4 sub puntajes podrían también ser útiles” (p. 17). Las cuatro subescalas propuestas en el reporte de Wing et al. (1998) fueron identificadas como: “Comportamiento” (ítems 1, 2 y 3), “Discapacidad” (ítems 4 y 5), “Síntomas” (ítems 6, 7 y 8) y “Social” (ítems 9, 10, 11 y 12).

Sin embargo, el modelo de cuatro factores original de Wing ha sido muy cuestionado por los diferentes estudios, que muestran distintas estructuras factoriales (Turner, 2004,

citado por Speak y Hay, 2012, p. 16). Los estudios de Preston (2000), Trauer et al. (1999) y McClelland (2000), que examinaron la estructura de las subescalas de la escala, encontraron diferentes estructuras factoriales (citados en Pirkis, Burgess, Kirk, et al., 2005, p. 80). Mientras que Preston observó un buen ajuste de las subescalas al modelo de cuatro factores que los autores originales sugirieron, Trauer et al. (1999) y McClelland (2000) obtuvieron otras estructuras factoriales totalmente diferentes (Pirkis, Burgess, Kirk, et al., 2005). Trauer et al. (1999), que encontró un ajuste pobre a la estructura factorial original, propuso una estructura alternativa de cinco factores. Por otro lado, ambos estudios poseen limitaciones (Speak y Hay, 2012): el de Trauer porque la muestra solo incluyó personas con diagnóstico de esquizofrenia, y por tanto se cuestionó su generalizabilidad, y el de McClelland (2000) porque la estructura factorial de cuatro factores extraída en el análisis de componentes principales no fue confirmada luego usando análisis factorial confirmatorio (Speak y Hay, 2012).

Lauzon et al. (2001) realizaron en la versión en francés (HoNOS-F) un análisis factorial confirmatorio del modelo de cuatro factores de Wing et al. (1998) con una muestra de 137 pacientes. En el estudio no pudieron comprobar el modelo de Wing, por lo que recomendaron usar solo el puntaje total de la HoNOS (citado por Speak y Hay, 2012, p. 17).

Con la versión alemana (HoNOS-D) Andreas et al. (2010) realizaron un análisis factorial exploratorio, tanto del modelo del Wing et al. (1998) como del propuesto por Trauer et al. (1999), y en ninguno de los dos casos pudieron confirmar la estructura factorial, por lo que también recomendaron el uso de los ítems individuales y del puntaje total.

Tradicionalmente el tema ha sido estudiado desde la Teoría Clásica de los Tests (TCT). Un estudio que utilizó la Teoría de Respuesta al Ítem (TRI) fue el de

Lovaglio y Monzani (2011a), quienes realizaron un análisis de validez de constructo de la versión italiana de la Escala utilizando el modelo Rasch. En este estudio concluyeron que los datos de la versión italiana indicaban una estructura factorial más multidimensional que unidimensional.

Respecto a la evidencia de validez concurrente, numerosos estudios evidenciaron una alta correlación de la HoNOS en comparación con otros instrumentos de evaluación del clínico, como the Role Functioning Scale (RFS); Brief Psychiatric Rating Scale (BPRS); Life Skills Profile (LSP) (Pirkis, Burgess, Kirk, et al., 2005). Según la revisión de (Pirkis, Burgess, Kirk, et al., 2005), la HoNOS no ha mostrado un buen desempeño en relación con las escalas de evaluación del usuario (por ejemplo, respecto al Symptom Checklist, SCL-90-R; Derogatis, 1983) ni con la versión de auto-reporte de la HoNOS. Sobre esta discrepancia advierten que los datos coinciden con los hallazgos de otros estudios en los que se utilizan instrumentos que recogen diferentes fuentes de información, observando una mayor correlación entre aquellos que se basan en la misma fuente (Pirkis, Burgess, Kirk, et al., 2005).

Respecto a la validez discriminante, Pirkis, Burgess, Kirk, et al. (2005) incluyeron en su revisión estudios en los que se observó una adecuada capacidad de la escala para discriminar entre grupos de usuarios con diferentes diagnósticos, así como con diferentes niveles de discapacidad en relación con el tratamiento recibido, y con los tipos de servicios de atención (residencial, centro diurno, ambulatorio, internación y de larga estadía).

Kisely et al. (2007) realizaron un estudio de validez discriminante de la HoNOS comparando los puntajes de 485 usuarios internados y ambulatorios, observando diferencias

estadísticamente significativas entre la media global del grupo de pacientes internado (17.1, SD = 5.38) comparada con la del grupo de pacientes ambulatorios (10.8, SD = 5.85).

Distintos estudios han demostrado buena validez predictiva de la HoNOS respecto a los contactos del servicio, el tiempo de estadía de internación y la tasa de readmisión a las internaciones psiquiátricas (Pirkis, Burgess, Kirk, et al., 2005).

En varios estudios los puntajes totales más altos han estado asociados a pacientes con diagnósticos de alcohol y drogas, psicosis y trastornos bipolares (Pirkis, Burgess, Kirk, et al., 2005).

Con respecto a la evidencia sobre la fiabilidad de la HoNOS, según la revisión de Pirkis, Burgess, Kirk, et al., (2005) los estudios sobre la consistencia interna reportan un alfa de Cronbach entre .59 y .76, que para una medida de este tipo indicaría una consistencia interna moderadamente alta, reafirmando su uso como medida global de severidad de los trastornos mentales (Wing et al., 1998; Sergill, 1999; Eagar, 2005, citados en Pirkis, Burgess, Kirk, et al., 2005, p 80). Por otra parte, cada ítem de la HoNOS pretende aportar algo distinto por tanto los estudios muestran baja redundancia entre ítems (Pirkis, Burgess, Kirk, et al., 2005).

En relación con la evidencia sobre la confiabilidad test- retest, si bien existen pocos estudios al respecto, según Deady (2009), los estudios han reportado unos valores del coeficiente de fiabilidad muy variables: entre .40 (baja estabilidad) y .80 (alta).

En relación a la confiabilidad interjueces, excepto en algunos ítems en particular los distintos estudios han encontrado en general un acuerdo global entre pares de evaluadores, con valores de bajo a moderado (kappa entre .21-.60). (Deady, 2009; Pirkis, Burgess, Kirk, et al., 2005).

La sensibilidad de la HoNOS a registrar el cambio ha sido evaluada en varios estudios examinando la magnitud y dirección del movimiento del puntaje global o de los ítems a lo largo del tiempo utilizando diferentes criterios. Uno de los métodos más simples utilizados ha sido su estudio en determinados ámbitos con la hipótesis de que disminuirá la severidad a medida que se va acercando el final del tratamiento (Pirkis, Burgess, Kirk, et al., 2005).

Las revisiones de Pirkis, Burgess, Kirk et al. (2005) y de Delaffon et al. (2012) coinciden en señalar que la HoNOS en general ha mostrado ser sensible al cambio, observándose un mayor grado de cambio en los pacientes internados en relación a los ambulatorios, así como en los pacientes que presentan puntajes globales más altos al inicio de un tratamiento. En Australia, Burgess, Pirkis, y Coombs (2006) analizaron los datos de 21911 aplicaciones de HoNOS en pacientes internados en estado agudo, y 14699 de pacientes ambulatorios, realizadas en diferentes ocasiones (ingreso- alta, ingreso-control, control-control) entre los años 2000 a 2005. En el estudio observaron que los pacientes que mas cambio mostraron en su puntaje global fueron los del grupo internado evaluados al ingreso y al alta, que presentaron un cambio en la puntuación total en HoNOS de 7.3 y un tamaño de efecto grande (d de Cohen = 1.00).

El estudio de Kisely et al. (2007) mencionado anteriormente, también analizó la sensibilidad de la HoNOS en la práctica clínica en tres sitios de Canadá a través de su aplicación a 485 usuarios internados y ambulatorios durante cuatro meses en diferentes tiempos: en los pacientes internados al ingreso y al egreso; y en los ambulatorios al inicio y en sucesivas consultas de control. Analizaron el cambio en los puntajes globales y en cada ítem, hallando para ambos grupos diferencias significativas en la disminución del puntaje en

HoNOS. En el grupo de pacientes internado el tamaño de efecto observado fue mayor (d de Cohen = 1.00) en comparación con el grupo de pacientes ambulatorio (d de Cohen = .32).

En otro estudio longitudinal realizado en Canadá por Kisely et al. (2010) con una muestra mayor ($n = 4305$) y un diseño que intentó parecerse lo máximo posible a la práctica clínica habitual, se encontraron también cambios significativos en el puntaje global de la HoNOS, así como en 9 de los 12 ítems. Las excepciones fueron los ítems 5,11 y 12.

Para evaluar el cambio fiable, otros estudios han utilizado el cálculo del Índice de Cambio fiable (ICF). El estudio de Parabiaghi et al. (2005), que evaluó a 4759 usuarios de servicios de salud mental ambulatorios de Italia en al menos dos tiempos con un intervalo de tiempo de siete meses, con su cálculo con un ICF de 8, observó que el 91.6% de la muestra se mantuvo estable y un 5.6% mejoró fiablemente.

El cálculo de cambio clínico significativo (CCS) requiere, tal como se dijo, de datos normativos; esto presenta un problema para la HoNOS, que no posee datos con los cuales realizar comparaciones o considerar puntos de corte clínicos (Jacobs, 2009).

Frente a este problema algunos trabajos (por ejemplo Audin, et al., 2001; Kisely et al., 2010; Parabiaghi et al., 2005) han formulado y ensayado diversas alternativas. Por ejemplo, el estudio llevado adelante por Kisely et al. (2010) mencionado anteriormente, estableció como criterio para el punto de corte los ítems que presentaron puntajes que indicarían en la HoNOS presencia de sintomatología (puntajes de 2, 3 y 4) de los puntajes que indicarían ausencia o mínima sintomatología (puntajes de 0 y 1 respectivamente).

Viabilidad. La HoNOS se diseñó específicamente con el objetivo de ser utilizada de manera rutinaria por los clínicos en la práctica habitual en los servicios de salud mental de

adultos (Wing et al., 1998). Si bien algunos estudios de validación de HoNOS han sido realizados en contextos de investigación, desde sus inicios se advirtió la importancia de estudiar las propiedades de la HoNOS en la práctica habitual, y obtener evidencia de que puede brindar información útil y relevante cuando es usada por los propios clínicos (Bebbington et al., 1999).

En los últimos años Kisely en Canadá (Kisely et al., 2007; Kisely et al., 2010), Lovaglio y Monzani en Italia (2011a, 2011b), Koch et al. en Suiza (2013) han llevado adelante estudios sobre las propiedades de HoNOS en el contexto de la práctica clínica habitual, con diseños que intentan alterarla lo mínimo posible, reportando todos ellos adecuada viabilidad. Distintos estudios y revisiones (por ejemplo, Pirkis, Burgess, Kirk et al., 2005; Trauer et al., 1999; Kisely et al., 2007) permiten concluir que esta propiedad es adecuada y aceptada para su uso con diferentes poblaciones.

La evidencia sobre el uso de la HoNOS es abundante (Burgess, Coombs, Clarke, Dickson, y Pirkis, 2012). Sin embargo, la escala ha sido objeto de algunos cuestionamientos por parte de algunos autores (p.e. Slade, 2012; Trauer y Buckingham, 2006), así como de algunos clínicos, tal como fue observado en los estudios de Jacobs (2009) y de Jacobs y Moran (2010). Algunos de los cuestionamientos se basan en algunas de las dificultades para otorgar puntuación y/o para la interpretación de los datos de la HoNOS, reportadas a partir de experiencias con el uso de la escala y en sesiones de entrenamiento (Trauer y Buckingham, 2006). Algunas de las mencionadas por Trauer y Buckingham (2006) se refieren a la dificultad en otorgar puntuación en algunos ítems (especialmente el 8 y los sociales-10- 12- por lo comentado anteriormente); plantear criterios temporales distintos entre los ítems 1-9 y

los ítems 10-12 y la existencia de diferencias en los procedimientos empleados cuando hay ítems sin puntuar.

Estas dificultades, sumadas al debate sobre su estructura factorial, han llevado a proponer desde pequeñas modificaciones a nivel de las instrucciones de la escala hasta importantes modificaciones. Lovaglio y Monzani (2011a) incluso proponen para la versión italiana una versión reducida de la escala con seis ítems.

Si bien el gobierno del Reino Unido tiene el copyright de la HoNOS, la escala es de dominio público, y no se plantean restricciones a las modificaciones que puedan hacerse para adaptarla a los requerimientos locales (Trauer y Buckingham, 2006). Sin embargo, hacer modificaciones de la escala original conduce a ventajas y desventajas. La ventaja principal es que podrían llegar a corregirse algunos problemas, pero la desventaja es que si los cambios son muy importantes se termina creando un nuevo instrumento que no permite las comparaciones internacionales y que puede no llegar a ser comparable a la versión original, ni aprovechable en cuanto a la experiencia acumulada (Trauer y Buckingham, 2006).

Otro cuestionamiento ha tenido que ver con la utilidad de la información de la HoNOS a nivel de los clínicos para la asistencia individual (Slade, 2012). Mientras algunos la han cuestionado, otros son más positivos y la consideran abarcativa, fácil y útil (Pirkis, Burgess, Coombs, et al., 2005). La revisión de Coombs et al., (2011) describe la forma en que los servicios de salud mental y los clínicos de Australia y Nueva Zelanda están utilizando los datos de ROA, con la familia de la HoNOS como su componente central (Coombs et al., 2011). Estos autores demostraron que se está haciendo un uso clínico de los datos de ROA para mejorar los tratamientos, a través de apoyo en la planificación y la toma de decisiones clínicas, aumentando el compromiso del usuario con el tratamiento (Coombs et al., 2011).

Concluyen que es probable que el problema no sea la HoNOS porque no hay una medida estandarizada perfecta (Coombs y McKay, 2012). “Como todos los instrumentos de medición globales, la HoNOS representa un compromiso entre la complejidad y la practicidad” (Trauer y Buckingham, 2006, p. 7). La HoNOS Intenta proveer una "fotografía" relevante en todos los grupos de usuarios adultos de salud mental, cubriendo un amplio espectro de fenómenos clínicos y sociales (Trauer y Buckingham, 2006).

Como afirman Coombs y McKay (2012) el gran uso que se está haciendo de la HoNOS en Australia y Nueva Zelanda, si bien no significa necesariamente que el instrumento sea ideal, sugiere fuertemente que puede ser útil para muchos propósitos. La extensa literatura y la experiencia demuestran que “la HoNOS resulta una medida global válida, fiable y útil de los cambios en la salud y funcionamiento de los usuarios a medida que progresan a través del tratamiento” (Trauer y Buckingham, 2006, p. 7)

5. Modelo de Rasch

Desde comienzos del siglo XX, La Teoría Clásica de los Tests (TCT) ha sido el principal modelo psicométrico empleado en la construcción y análisis de los tests. Sin embargo, sus limitaciones han llevado a la utilización de modelos de medición basados en la teoría de respuesta al ítem (TRI) que resuelven algunos de los problemas de la TCT, y presentan importantes ventajas (Wolfe y Smith, 2007a, 2007b).

En el marco de la TRI, en 1960 el matemático Georg Rasch propuso un modelo de medida alternativo que permite construir pruebas más adecuadas y eficientes, derivando en una creciente aplicación en el ámbito de la evaluación psicológica (Prieto y Velasco, 2003).

En el campo de la salud mental, se puede observar en los últimos años el uso del modelo de Rasch en una variedad de aplicaciones entre ellas, para analizar y desarrollar instrumentos de resultados (véase Elliott, Fox, Beltyukova, Stone, y Gunderson, 2006; Mavranzouli, Brazier, Young, y Barkham, 2011) así como para evaluar el cambio en estudios longitudinales (por ejemplo Pastor y Beretvas, 2006).

5.1. Características del Modelo de Rasch

El modelo propuesto por Georg Rasch (1960) se fundamenta en dos supuestos principales: el atributo que se pretende medir puede representarse en una única dimensión donde se situarían conjuntamente las personas y los ítems; el nivel de la persona en el atributo y la dificultad del ítem son los únicos determinantes de la probabilidad de que la respuesta sea correcta (Prieto y Delgado, 2003).

La formulación más conocida se deriva de la predicción de la probabilidad de una respuesta al ítem (resolverlo correctamente, acertarlo) a partir de la diferencia entre el nivel de la persona en el atributo medido y el nivel de dificultad del ítem (Prieto y Delgado, 2003). La fórmula para modelizar esta relación se describe en la expresión (1).

$$\ln (P_{is} / 1 - P_{is}) = (B_s - D_i) \quad (1)$$

Esta fórmula indica que el cociente entre la probabilidad de una respuesta correcta y la probabilidad de una respuesta incorrecta a un ítem ($P_{is} / 1 - P_{is}$), es una función de la diferencia en el atributo entre el nivel de la persona (B_s) y el nivel del ítem (D_i).

Al igual que en casi todos los modelos de la TRI, la unidimensionalidad es un supuesto central del modelo de Rasch. Este principio indica que la variable que se pretende

medir se refiere a un único atributo en el que varían las personas y los ítems, de forma que sus diferencias son claramente interpretables (Bond y Fox, 2007).

Para la mayoría de los constructos psicológicos, la dimensionalidad estricta nunca es alcanzada (Zickar y Broadfoot, 2009). De acuerdo con Linacre (2012) lo importante no es comprobar si los datos son totalmente unidimensionales, porque en realidad nunca lo son, sino que se trata de determinar si la falta de unidimensionalidad en los datos es suficientemente grande como para amenazar la validez de los resultados. Para valorar esto existen distintos criterios. Reckase (1979) sugiere que siempre que el primer factor explique al menos el 20% de la varianza de la escala y no haya un segundo factor dominante, se puede considerar que los datos son básicamente unidimensionales.

Como en el resto de los modelos TRI y a diferencia de la TCT, el modelo posee la propiedad denominada *medición conjunta* porque los parámetros, tanto de las personas como de los ítems, se expresan en las mismas unidades y se localizan en el mismo continuo (Prieto y Delgado, 2003). Estos valores pueden expresarse en distintas métricas, la más utilizada es la *escala logit*. Una escala logit (unidad log odds) es una escala de intervalo en la que las unidades en las que se miden las distancias entre las localizaciones en el mapa de la persona-ítem son invariantes a lo largo de la variable y tienen el mismo significado. De esta forma, a diferencias iguales entre una persona y un ítem le corresponden la misma probabilidad de una respuesta correcta. Esta propiedad es importante para los estudios de evaluación de cambio terapéutico porque se garantiza la invarianza de las puntuaciones diferenciales a lo largo del continuo.

Aunque la escala logit puede adoptar valores entre más y menos infinito, en la gran mayoría de los casos, si la dificultad de la prueba es apropiada para el nivel de las personas,

se sitúa en el rango ± 5 . La localización del punto 0 de la escala es arbitraria, y se sitúa convencionalmente en la tradición de Rasch en la dificultad media de los ítems (Prieto y Velasco, 2003).

La métrica de intervalo se deriva de la propiedad denominada objetividad específica, por la que las diferencias entre las personas en un atributo son independientes de los ítems utilizados, y las diferencias entre los ítems no dependen de las personas que se utilizan para estimarla (Andrich, 1988).

El análisis Rasch proporciona índices para valorar el ajuste de personas y de ítems de acuerdo a los requerimientos del modelo (Bond y Fox 2007). Para cuantificar el ajuste se emplean generalmente dos estadísticos principales: *Infit* y *Outfit* (Linacre, 2002). Estos estadísticos son medias de residuos (diferencias entre las puntuaciones observadas y las estimadas desde el modelo). El estadístico *Outfit* es la media no ponderada de los residuos cuadráticos estandarizados de una persona o un ítem. *Outfit* es muy sensible a la presencia de *outliers* (valores muy inesperados). *Infit*, por el contrario, es un estadístico de ajuste más robusto que representa la media de los residuos cuadráticos estandarizados ponderados con su varianza (Prieto y Delgado, 2003; Prieto y Velasco, 2003).

Aunque existen distintos criterios para valorar su magnitud, convencionalmente se suele considerar que la oscilación entre .5 y 1.5 indica un ajuste razonable al modelo (Linacre, 2012). Los valores situados entre 1.5 y 2.0 manifiestan un desajuste moderado con escasas consecuencias para la validez de las medidas. Los índices inferiores a .5 pueden sugerir sobreajuste o redundancia, y los valores superiores a 2.0 revelan un severo desajuste.

Desde el modelo de Rasch, la fiabilidad de cada medida individual se puede estimar mediante el error estándar (desviación típica de los errores de los estimadores del parámetro)

y la fiabilidad promedio de un grupo de medidas (de personas o de ítems) mediante estadísticos denominados *Person Separation Reliability (PSR)* e *Item Separation Reliability (ISR)*.

Estos estadísticos, que pueden oscilar entre 0 y 1, son similares al coeficiente de fiabilidad clásico (cociente entre la varianza verdadera y la varianza observada).

5.2. El Modelo de Escalas de Calificación

Debido a que los ítems de la escala HoNOS son politómicos (hay más de dos categorías de respuesta), hemos utilizado en este trabajo el Modelo de Escalas de Calificación (MEC; Andrich, 1978). El MEC es una extensión del modelo dicotómico de Rasch que se aplica a ítems politómicos, que permite el análisis empírico de la calidad métrica de las categorías de respuesta del instrumento.

Este modelo se describe en la expresión (2),

$$\ln (P_{nik} / P_{ni (k-1)}) = B_n - D_i - F_k \quad (2)$$

donde P_{nik} es la probabilidad de que la persona n , al ser puntuada en el ítem i , reciba la categoría de respuesta k ; $P_{ni (k-1)}$ es la probabilidad de que la persona n , al ser puntuada en el ítem i , reciba una categoría de respuesta inmediatamente inferior a k ; B_n es el parámetro de la persona n en la variable medida. D_i es la dificultad del ítem i ; F_k es el valor en la variable latente en el que las categorías k y $k-1$ son equiprobables. F_k es denominado paso y es constante entre los ítems

Linacre (2002) publicó un conjunto de recomendaciones para lograr unas categorías adecuadas. Como análisis previo, se debe constatar que las correlaciones ítem-test (biserial-

puntuales) son positivas y moderadamente altas. Las correlaciones negativas pueden revelar errores de codificación y las muy bajas podrían indicar multidimensionalidad.

La primera recomendación indica que deberán existir por lo menos 10 observaciones en cada categoría de respuesta. Una baja frecuencia conducirá a una estimación imprecisa y potencialmente inestable del paso entre las categorías sucesivas.

La segunda recomendación se refiere a la existencia de una distribución regular de las observaciones. La situación óptima es la de una distribución uniforme por las categorías. Otro tipo de distribuciones también son aceptables: una distribución unimodal (con altas frecuencias en categorías centrales o extremas) o bimodal (con altas frecuencias en las categorías extremas).

La tercera, se refiere al incremento monótonico de los promedios de las medidas de las personas que eligen las categorías. En general, las observaciones en las categorías más altas/más difíciles deben reflejar niveles superiores en el atributo, implicando así que la media de la medida para cada conjunto de observaciones empíricas, debe aumentar monótonicamente con las categorías.

La cuarta recomendación es que las categorías no deben presentar un desajuste elevado con el modelo (los valores Outfit deben ser inferiores a 2.0).

La quinta recomendación es que se ha de observar la existencia de un incremento monótonico de los pasos de calibración, que indica que los pasos entre las categorías sucesivas habrán de estar ordenados monótonicamente. El desorden de los pasos de calibración ocurre muchas veces cuando las frecuencias de las categorías siguen un comportamiento irregular, como por ejemplo cuando una categoría inferior es escogida por sujetos con habilidades medias o superiores (Linacre, 2002).

II. ESTUDIOS EMPÍRICOS

II. ESTUDIOS EMPÍRICOS

Introducción

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 (Uriarte et al., 1999). El RCP envió además, el manual y el material disponible para poder desarrollar el entrenamiento y el estudio de la escala HoNOS.

También se presentó el proyecto de investigación al Comité de ética en investigación de la Facultad de Psicología de la Universidad de la República del Uruguay, y se obtuvo la aprobación.

Uruguay cuenta con una población de 3.368.595 habitantes. Se encuentra dividido en 19 Departamentos, casi el 40% de su población reside en Montevideo, la capital del país.

Desde el 2007 rige en el país un Sistema Nacional Integrado de Salud, integrado por prestadores públicos y privados. El prestador principal del sector público es la Administración de Servicios de Salud del Estado (ASSE) que posee alcance nacional y cubría en el 2012 a 310000 usuarios de salud mental.

El estudio fue llevado a cabo en la red de servicios de ASSE. Participaron del estudio profesionales y usuarios atendidos por 4 Equipos Comunitarios de Salud Mental (ECSM) de 4 policlínicas ambulatorias de distintas zonas de Montevideo, y 3 salas de internación psiquiátrica del Hospital Vilardebó (HV) con pacientes no judiciales.

Los ECSM son 45 en todo el país, distribuidos 17 en Montevideo y 28 en el resto de los Departamentos. Al momento de recolectar los datos (2012) ASSE tenía asignado en Montevideo la atención de 79000 usuarios de salud mental. Los ECSM en general, atienden a

todos los pacientes que consultan ambulatoriamente por problemas de salud mental dentro de su área de cobertura.

El HV es el hospital psiquiátrico nacional de referencia y atiende a usuarios de Montevideo y del resto del país. Está ubicado en la capital del país, cuenta con 292 camas y asiste a personas con trastornos mentales mayores de 15 años. Existen algunas salas especiales para pacientes ingresados por disposición judicial.

El objetivo principal de este trabajo, que será detallado más adelante, es llevar a cabo un estudio de validación de la HoNOS en servicios públicos de salud mental de Uruguay.

En una etapa previa al estudio de validación se realizó un estudio piloto que se describe a continuación.

Estudio piloto de la HoNOS

1. Objetivos

El objetivo general fue analizar la adecuación de la HoNOS y de un cuestionario socio-demográfico y clínico complementario a la Escala HoNOS (CCH), y su necesidad de adaptación para su utilización en el estudio de validación. Los objetivos específicos incluyeron analizar la HoNOS y el CCH respecto a: (1) la relevancia, (2) la comprensión de las instrucciones e ítems, (3) el formato, (4) la necesidad de agregar, eliminar y/o modificar ítems.

2. Metodología

2.1. Participantes. Participaron 8 clínicos: 5 psicólogos y 3 psiquiatras con una media de 18 años de experiencia de trabajo con pacientes con trastornos mentales. El 100% de los clínicos cuenta con experiencia clínica de trabajo en instituciones públicas y el 38% (3) de ellos posee además experiencia en instituciones privadas.

2.2 Instrumentos.

2.2.1. *Health of the Nation Outcome Scales (HoNOS; Wing et al., 1998).*

Se empleó la versión al castellano de Uriarte et al. (1999) cuyas características han sido descritas en la sección cuatro de la revisión teórica.

2.2.2. *Ficha de información complementaria a la Escala HoNOS (CCH)*

La ficha consistió en un cuestionario socio-demográfico y clínico complementario a la Escala HoNOS (CCH), que se diseñó con el propósito de recoger datos socio-demográficos y

clínicos básicos y estandarizados sobre los usuarios y el tratamiento, que permitieran analizar la información proveniente de la aplicación de la Escala HoNOS (ver Apéndice B).

Su diseño fue necesario debido a que si bien la HoNOS incluye algunos datos complementarios (*“HoNOS Chart Background information”*), para su análisis se requiere además de otros datos que habitualmente se extraen de la historia clínica del usuario, pero los servicios de salud mental públicos de Uruguay no poseían un sistema de recolección de datos estandarizados que permitieran su extracción. En Uruguay, a excepción de los datos específicos para las nuevas prestaciones en salud mental (MSP, 2011), no existe a nivel nacional un conjunto mínimo de datos obligatorios sociodemográficos y clínicos de salud mental recolectados por todos los prestadores de salud (Del Castillo, Dogmanas, y Villar, 2011), como consecuencia, los diferentes servicios de salud recogen datos distintos y de diversa calidad.

La construcción del cuestionario se llevó a cabo incluyendo los ítems complementarios que incluye la escala original *“HoNOS Chart Background information”* y los ítems más frecuentes con los que se analiza la HoNOS identificados a partir de la revisión de la literatura.

Las variables incluidas en la versión del CCH sometida al juicio de los expertos fueron: sexo, edad, ciudad de residencia, nivel educativo alcanzado, actividad laboral, estado civil, diagnóstico según la Clasificación Internacional de Enfermedades (CIE-10) y duración del tratamiento.

2.2.3. Cuestionario para la adaptación de la HoNOS en Uruguay. Se incluyeron preguntas para valorar: las instrucciones de la escala (claridad), formato general (extensión,

orden, opciones de respuesta), relevancia y claridad de la redacción de los ítems de la HoNOS (ver Apéndice C).

Las respuestas fueron tanto cerradas como abiertas. Para puntuar la adecuación de las instrucciones, la relevancia y la claridad de los ítems se incluyeron 5 categorías de respuesta, siendo 1 “muy poco” y 5 “total”. Para valorar el formato, las opciones del 1 al 5 correspondieron a: 1 “inadecuada” y 5 “totalmente adecuada”. Las respuestas abiertas se incluyeron para permitir a los participantes añadir sugerencias.

2.2.4. Cuestionario sobre la ficha complementaria a la Escala HoNOS Uruguay

Se incluyeron preguntas referidas al formato del cuestionario de datos socio-demográficos y clínico complementario a la Escala HoNOS, (extensión, orden y número de opciones de respuestas) (ver Apéndice D).

Las categorías de respuesta fueron tanto cerradas como abiertas. Las cerradas fueron en algunos ítems “sí”, y “no”, mientras que en otros, se ofreció 5 opciones de respuesta, siendo el 1 “inadecuada” y el 5 “totalmente adecuada”. Las abiertas se incluyeron con el objetivo de permitir a los participantes añadir explicaciones y sugerencias respecto a eliminar, adicionar o modificar ítems.

2. 3. Procedimiento. El procedimiento incluyó dos etapas: en primer lugar se recogió la valoración individual de los clínicos a través de los cuestionarios y en segundo lugar, se utilizaron dos sesiones grupales de consenso entre expertos. En total se mantuvo tres sesiones con el grupo. En la primera, se explicó el propósito y las características del estudio, se presentaron todos los materiales y se acordó la metodología para las próximas reuniones. Se solicitó a los expertos que para las siguientes reuniones analizaran los materiales y realizaran la evaluación individualmente. Posteriormente, se realizaron dos reuniones grupales de tres

horas de duración cada una con el objetivo de intentar establecer consensos. Las sesiones consistieron en recorrer cada ítem de la HoNOS y del CCH, centrando la discusión en los ítems que presentaron menor puntaje y propuestas de modificación en la evaluación individual.

El rol del investigador en estas sesiones fue el de coordinar y favorecer la discusión con el objetivo de lograr establecer consensos.

Las sesiones fueron grabadas.

3. Resultados

3.1. HoNOS.

Instrucciones. La Claridad de las instrucciones obtuvo una puntuación media de 3.5.

Cinco clínicos plantearon en la evaluación individual que deberían realizarse modificaciones a la redacción de las instrucciones. En la sesión grupal, se manifestaron dificultades para comprender “la lógica” de evaluación propuesta por la HoNOS sólo con los materiales entregados.

No se arribó a ninguna sugerencia concreta para modificar la redacción de las instrucciones.

Formato. La extensión y el orden de la escala obtuvieron en la evaluación individual, puntajes medios de 3.25 en extensión, 3.50 en orden y 3.50 en número de opciones de respuesta.

Relevancia y contenido de los ítems. La relevancia de la mayoría de los ítems de la escala (8 de los 12) fueron valorados con puntajes de 4 ó 5 por el 100% de los expertos.

Mientras que el ítem el 9 (problemas con las relaciones) presentó la mayor puntuación media en relevancia (4.75) y claridad (4.50), el ítem 8 (Otros problemas mentales o conductuales) obtuvo la menor en relevancia (3.12) y claridad (3.25). Además el ítem 8 fue el más discutido en la sesión grupal. Se planteó que este ítem presenta dificultad para otorgar una puntuación por el requerimiento de seleccionar solo un solo problema (entre 9 opciones) y además, opinaron que el criterio establecido en el glosario no resultaba claro.

Los expertos plantearon opciones alternativas que se discutieron en el grupo: (1) elegir todos los pertinentes, (2) elegir dos problemas, (3) elegir dos o más problemas, (4) puntuar de acuerdo a todos los problemas en general. Finalmente el consenso al que arribó el grupo fue proponer que se eliminara del glosario la palabra “únicamente” y que se sustituyera la indicación de puntuar “el problema clínico más severo” por: “el problema clínico determinante”.

En la Tabla 4 se muestran las puntuaciones sobre relevancia y claridad para todos los ítems.

Desde el punto de vista lingüístico se estableció un consenso sobre la propuesta de modificar dos palabras del glosario en 2 de los ítems para su adaptación a Uruguay. En el ítem 1 del glosario la palabra “enseres” por “utensilios”, y en el ítem 12 “paro” por “subsidio.”

Tabla 4

Valores promedio de las respuestas de los expertos sobre la relevancia y claridad de los ítems de la HoNOS

Ítem	Relevancia	Claridad
	M (DT)	M (DT)
Conducta hiperactiva	4.38 (.52)	4.38 (.52)
Autolesiones	3.88 (1.64)	4.13 (.83)
Alcohol o drogas	4.25 (.46)	3.75 (1.67)
Problemas cognitivos	4.13 (.64)	3.88 (.35)
Enfermedad física	4.13 (.83)	4.38 (.74)
Ideas delirantes y alucinaciones	4.13 (.35)	4.50 (1.07)
Humor depresivo	3.88 (1.25)	4.50 (.54)
Otros problemas mentales	3.13 (1.36)	3.25 (.71)
Relaciones	4.76 (.46)	4.50 (.76)
Actividades de la vida cotidiana	4.50 (.53)	4.13 (.64)
Condiciones de vida	3.50 (1.51)	4.13 (.84)
Ocupación	3.88 (1.64)	4.00 (.54)

Nota. Opciones de respuesta: 1 = muy poco. 2 = poco. 3 = regular. 4 = bastante. 5 = total. M = media. DT = desviación típica.

3.2. Ficha de información complementaria a la Escala HoNOS (CCH). En la valoración individual el Formato del CCH fue valorado con puntajes de 3.12 en extensión, 3.25 en orden y 3.12 en opciones de respuesta.

En la discusión grupal, respecto a las modificaciones propuestas para el CCH, se consensuó la propuesta de incluir la identificación del profesional que realiza la evaluación agregando el nombre y la firma con el argumento de lograr mayor compromiso con la evaluación por parte del evaluador.

Otro aspecto en el que se estableció un consenso fue en el de agregar un ítem que contuviera información respecto al tipo de tratamiento recibido por el usuario, que ofreciera además distintas opciones de respuesta. Se propuso la redacción del ítem de la siguiente manera: *“Tratamiento (seleccionar los que correspondan): farmacológico/biológico, psicoterapéutico, social, otro.”*

Otra de las sugerencias fue incluir un ítem que recogiera la existencia de otras variables intervinientes que pudieran explicar los resultados observados (ejemplo: situaciones traumáticas o adaptativas de cambio, pérdidas, etc.). Se propuso la redacción del ítem de la siguiente manera: *“Eventos relevantes desde la última evaluación.”*

No se realizaron propuestas para la eliminación de ítems.

4. Conclusiones

Se analizó la adecuación y necesidad de adaptación de los instrumentos HoNOS y CCH para el estudio de validación en Uruguay.

En relación a la HoNOS, la valoración de los clínicos en los distintos aspectos evaluados: claridad de las instrucciones, relevancia y claridad de los ítems, y el formato indican que resultó aceptable y adecuada para su uso en Uruguay.

En la discusión grupal respecto a la HoNOS, las modificaciones que propusieron los clínicos se centraron en dos aspectos: la modificación de la redacción del ítem 8 y la adaptación lingüística de dos términos del glosario. La dificultad encontrada por los expertos en el ítem 8 respecto a la restricción de seleccionar solo un problema, coincide con lo reportado por otros estudios según la revisión de Pirkis (2005a), y otros estudios han intentado modificarlo también. En nuestro estudio, el grupo de expertos discutió varias opciones planteando finalmente dos modificaciones del glosario. Respecto a la propuesta de eliminar del glosario la palabra “únicamente”, consideramos que su eliminación no resolvería el problema dado que la dificultad que plantea este ítem está más relacionada al criterio y a la restricción de seleccionar sólo un problema, cuando en la práctica real los pacientes pueden presentar varios problemas simultáneamente. Parece plausible pensar que eliminar esa condición podría generar más confusión para el evaluador porque las opciones de respuesta que ofrece la HoNOS permiten indicar sólo uno. Por otra parte, la sugerencia de sustituir el problema clínico “más severo” por el “determinante”, estaría cambiando el criterio que se plantea para toda la escala que es evaluar según el grado de severidad. Además el término propuesto requeriría ser definido conceptualmente y operacionalizarse.

La sugerencia sobre la adaptación lingüística de dos términos del glosario, son razonables para su adecuación al ser utilizado en Uruguay en el estudio de validación.

Respecto al CCH, la valoración de los clínicos fue positiva en general. En la sesión grupal, se sugirieron tres modificaciones. Las variables “tipo de tratamiento” y “eventos relevantes”, aportarían información para valorar la existencia de otras variables intervinientes para explicar los resultados de los tratamientos, enriqueciendo los análisis posibles con HoNOS.

La sugerencia de identificar (con nombre y firma) al profesional que realiza la evaluación, podría ser importante en vistas a su utilización futura como instrumento en la práctica clínica dentro de los servicios de salud. Sin embargo, en nuestro estudio de validación, para poder cumplir con las normas éticas nacionales de investigación con seres humanos respecto a la protección de datos personales, resultará indispensable proteger el anonimato de los clínicos participantes.

La breve presentación de la HoNOS en la primera sesión grupal y la entrega de materiales de lectura, no resultaron suficientes para comprender todos los aspectos de la escala, por lo que tal como lo recomiendan los autores (Wing et al. 1998), se corrobora la necesidad de realizar una capacitación previa en la HoNOS para su utilización.

Estudio de Validación de la HoNOS

1. Objetivos

1.1. Objetivos generales

Llevar a cabo un estudio de validación de la versión al castellano de la escala HoNOS en la población clínica de servicios públicos de salud mental de Uruguay

1.2. Objetivos específicos

1.2.1. Analizar las propiedades psicométricas de la escala HoNOS utilizando la TCT y el Modelo de las Escalas de Calificación. Este análisis implica:

- a) poner a prueba el sistema de categorías de respuesta de la HoNOS.*
- b) analizar el ajuste de las personas e ítems al modelo de Rasch.*
- c) examinar la fiabilidad de las puntuaciones de la HoNOS.*
- d) obtener evidencias sobre la validez discriminante de la HoNOS.*
- e) analizar la sensibilidad al cambio de la HoNOS.*

1.2.2. Obtener evidencias sobre la viabilidad de HoNOS como instrumento de evaluación rutinaria de resultados de los servicios de salud mental de Uruguay.

2. Metodología

2.1. Participantes. La muestra está integrada por 402 usuarios de los servicios de salud mental descritos; 285 (71%) se encontraban recibiendo tratamiento en atención ambulatoria y 117 (29%) en régimen de internación psiquiátrica.

La edad media del grupo ambulatorio es de 44.42 (DT = 15.27) y la del grupo internado de 39.28 (DT = 13.83) no resultando la diferencia estadísticamente significativa (p

> .05). El número de hombres y mujeres en ambos grupos no es similar; hay más hombres en el grupo internado y más mujeres en el grupo ambulatorio. La Tabla 5 presenta las características sociodemográficas de los grupos.

La Tabla 6 describe los diagnósticos principales en ambos grupos de pacientes. En el grupo ambulatorio el diagnóstico principal más frecuente es el Trastorno del estado de ánimo (33%), mientras que en la mayoría de los usuarios del grupo de internados el más frecuente es la Esquizofrenia y trastornos delirantes (54.7%).

En la recolección de datos participaron 22 clínicos: 14 psiquiatras (7 de los ECSM y 7 del HV) y 8 psicólogos (5 de los ECSM y 3 del HV) con una media de 18.09 (DT = 8.12) de años de experiencia profesional.

Tabla 5

Características sociodemográficas de la muestra

	Grupo de estudio		
	Ambulatorio (n = 285)	Internado (n = 117)	
Edad			
- Media	44.4 (DT = 15.2)	39.3 (DT = 13.8)	T -3.139
Género			
- Hombres	87	71	χ^2
- Mujeres	198	46	31.623 *
Nivel educativo (n1)			
- Primaria incompleta	.7 %	1.9 %	
- Primaria completa	35.2 %	43.0 %	
- Secundaria	48.8 %	38.3 %	
- UTU	6.8 %	12.1 %	
- Universitario incompleto	5.7 %	3.7 %	
- Universitario completo	2.8 %	.9 %	
Situación laboral (n2)			
- Trabaja	35.2 %	9.7 %	
- No trabaja	64.8 %	90.3 %	

Tabla 5 (Cont.)

	Grupo de estudio	
	Ambulatorio	Internado
	(<i>n</i> = 285)	(<i>n</i> = 117)
Estado civil (<i>n</i> 3)		
- Soltero	35.7 %	65.5 %
- Casado	22.6 %	8.0 %
- Separado/divorciado/viudo	32.5 %	19.5 %
- Unión libre	9.2 %	7.1 %
Lugar de residencia (<i>n</i> 4)		
- Montevideo	95.4 %	75.4 %
- Interior	4.6 %	24.6 %
Lugar de nacimiento (<i>n</i> 5)		
- Montevideo	73.9 %	73.9 %
- Interior	24.4 %	25.2 %
- Extranjero	1.8 %	.9 %
Primera vez atención en SSM (<i>n</i> 6)		
- Si	34.5 %	10.7 %
- No	65.5 %	89.3 %

Nota. * $p \leq .01$. *n*1 = 388; *n*2 = 397; *n*3 = 398; *n*4 = 396; *n*5 = 394.; *n*6 = 390; SM = Servicios de Salud Mental.

Tabla 6

Frecuencia de pacientes por grupo y diagnóstico principal

	Grupo de estudio	
	Ambulatorio (<i>n</i> = 285)	Internado (<i>n</i> = 114)
Demencia	4	4
Drogas / Alcohol / Abuso de sustancias	18	8
Esquizofrenia / Trastornos delirantes	36	64
Trastornos del estado de ánimo	94	16
Neurosis / Ansiedad	62	1
Trastornos del sueño, alimentación, etc.	7	0
Trastornos de personalidad	43	8
Otros	21	13

2.2. Instrumentos.

2.2.1. Escala HoNOS. Se empleó la versión al castellano de Uriarte et al. (1999) cuyas características han sido descritas en la sección cuatro de la revisión teórica.

A partir de las propuestas realizadas por los expertos en el estudio piloto, se realizaron dos modificaciones a la versión al castellano de Uriarte et al. (1999) para su adaptación lingüística a Uruguay. Se modificaron dos palabras del glosario en dos ítems; en el ítem 1 se

sustituyó la palabra “enseres” por “*utensilios*”; y en el ítem 12 la palabra “*paro*” por “*subsidio*.”

2.2.2. Hoja de registro de HoNOS. A partir de los resultados del estudio piloto que analizó la ficha de información complementaria a la Escala HoNOS (CCH), se diseñó una nueva versión. Se incluyeron en la versión aplicada en el estudio de validación, las dos modificaciones propuestas en el estudio piloto (tratamiento y eventos relevantes).

El CCH utilizado incluyó datos socio-demográficos, clínicos y relacionados a la aplicación de HoNOS. Los datos socio-demográficos del usuario relevados fueron: fecha de nacimiento, sexo, nivel educativo, trabajo, ciudad de residencia, ciudad natal, estado civil, convivencia. Los datos clínicos incluyeron: el diagnóstico primario según los grupos principales de la CIE-10 (Organización Mundial de la Salud [OMS], 1992): Demencia (F00-09); Drogas/ alcohol / Abuso de sustancias (F10-19); Esquizofrenia /trastornos delirantes (F20-29);Trastornos del estado de ánimo (F30-39); Neurosis, ansiedad (F40-48); Trastornos del sueño, alimentación, etc (F50-59);Trastornos de personalidad (F60-69); Otros, el tiempo de tratamiento; datos del tratamiento indicado (farmacológico, psicoterapéutico, social, otro) y otros eventos relevantes desde última aplicación. La información sobre la aplicación de la Escala incluyó: centro de salud, fecha de aplicación, profesión del evaluador, ámbito (tipo de servicio) y período de aplicación (ver Apéndice E).

2.2.3. Cuestionario de Viabilidad de HoNOS. Debido a la inexistencia de un instrumento psicométrico validado para evaluar la viabilidad de HoNOS al momento de realizar el presente estudio; a partir de las definiciones de Andrews et al. (1994) y Pirkis, Burgess, Kirk, et al. (2005), se desarrolló un breve cuestionario para recoger la valoración de los clínicos. Los contenidos incluidos han sido investigados de distinta manera, en los

estudios de Fonagy et al. (2005), Jacobs (2009) y Stedman et al. (1997), sobre la viabilidad de HoNOS.

El cuestionario utilizado evalúa la viabilidad de la HoNOS como medida para evaluar resultados en la práctica clínica, recogiendo a través de cinco ítems y una pregunta lo siguiente: (1) la utilidad, (2) relevancia del contenido, (3) facilidad para su aplicación, (4) el tiempo que insume (5) la capacitación que requiere, (6) su probable utilización. Las opciones de respuesta fueron cinco donde 1 es el mínimo y 5 el máximo (Ver Apéndice F).

De acuerdo a lo planteado por Andrews et al. (1994) sobre las características que implica la *viabilidad*, en el cuestionario desarrollado se recoge mediante la valoración de los clínicos, con los ítems 1 (utilidad) y 2 (relevancia) la aplicabilidad; con el ítem 3 (facilidad), la aceptabilidad; y con los ítems 4 (tiempo) y 5 (capacitación) la practicidad. El ítem 6 formulado de la siguiente manera: “*¿En qué medida la utilizaría para evaluar resultados en su práctica clínica?*”, intentó recoger la definición de viabilidad utilizada por Pirkis, Burgess, Kirk, et al. (2005), al definirla como el grado en que un instrumento resulta aceptable para los involucrados.

2.3. Procedimiento. La Dirección de Salud Mental y Poblaciones Vulnerables de ASSE (DSM ASSE) convocó en primer lugar a una reunión con los coordinadores de distintos equipos de salud mental para presentar el estudio e identificar los posibles interesados en participar. Posteriormente, entre los coordinadores de los equipos y la DSM ASSE se identificó a los clínicos que se invitaría a participar, de acuerdo a los criterios de inclusión definidos en el proyecto. Los criterios de inclusión de clínicos al estudio fueron: (1) psiquiatras o psicólogos de adultos, (2) integrantes de los equipos comunitarios de salud mental (ECSM) de Montevideo y/o del Hospital psiquiátrico Vilardebó, (4) trabajar en un

servicio con flujo importante de nuevos ingresos (en el caso de los ECMS), (5) decidir voluntariamente participar en el estudio, (6) participar en la reunión de capacitación requerida para ingresar al estudio.

El criterio para la inclusión de pacientes fue: (1) adultos, (2) con cualquier tipo de trastorno mental y/o diagnóstico, (3) usuarios de los servicios de salud mental participantes en el estudio, (4) pacientes de alguno de los clínicos participantes del estudio.

Una vez designados los participantes, se realizaron siete reuniones de capacitación con una duración aproximada de cuatro horas cada una, en las que se capacitó a 27 profesionales de 4 Policlínicas ambulatorias (ECSM) y 3 Salas del HV. La capacitación se diseñó 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. Consistió en la presentación del estudio, la revisión de la escala, ejercicios y entrega de los instrumentos de recolección de datos y materiales de lectura de apoyo para realizar el estudio.

De los 27 profesionales que participaron en las reuniones de capacitación, se acordó que ingresarían al estudio 22 de ellos. Los cinco que no participaron, se excluyeron por no cumplir con alguno de los criterios de inclusión.

La recolección de datos se realizó entre los meses de noviembre de 2011 a agosto de 2012. Los clínicos fueron ingresando al estudio en distintos momentos, según fueron confirmando su participación y de acuerdo a su disponibilidad.

Como el estudio se diseñó para realizarse en la práctica clínica habitual, intentando alterarla y/o sobrecargarla lo mínimo posible, el tiempo y la frecuencia de recolección de datos se acordó con cada clínico durante el intercambio en las reuniones de capacitación. Con los clínicos del HV, se acordó que, durante el período de participación en el estudio (junio a

agosto 2012), recolectarían datos de todos sus pacientes durante las evaluaciones rutinarias de control; y de los pacientes al momento de ingreso y de alta clínica, o próximos al alta al momento de finalizar el estudio. Se estableció con los profesionales como criterio de alta, el alta clínica, dado que en algunas ocasiones (por ejemplo, problemas de vivienda), el paciente permanece internado más tiempo del que según el criterio clínico se requeriría para la internación psiquiátrica.

Con los clínicos de los ECSM, se acordó que durante su período de participación en el estudio, recolectaran datos de todos los nuevos usuarios al momento de iniciar el tratamiento y en la consulta de control en un lapso aproximado de 90 días desde el inicio. Se determinó este momento para realizar la segunda evaluación tomando como referencia al que establece Australia para la recolección de datos de las consultas de control (Trauer, 2010b). Con aquellos clínicos que manifestaron que no recibían con frecuencia nuevos pacientes, se acordó que aplicarían la HoNOS a todos los pacientes que ya estuvieran en tratamiento en la próxima visita de control.

Al finalizar el período de recolección de datos se solicitó a los clínicos que respondieran el cuestionario de viabilidad de HoNOS.

2.4. Análisis de datos. 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, presentar los resultados con los puntajes directos (la suma de los valores de los 12 ítems), realizar las comparaciones, y correlaciones de las puntuaciones.

El análisis Rasch de las puntuaciones de la HoNOS se realizó mediante el software WINSTEPS 3.74.0 (Linacre, 2012).

Para analizar la unidimensionalidad de los datos se utilizaron dos procedimientos: El Análisis Paralelo (Horn, 1965) en el contexto del Análisis Factorial Exploratorio (AFE) y el Análisis de Componentes Principales de los Residuos, muy usado en la última década en el marco del modelo de Rasch.

Para realizar el Análisis Paralelo (AP) se utilizó se utilizó el programa Factor 7.2 (Lorenzo-Seva y Ferrando, 2006). De acuerdo a Abad, Olea, Ponsoda (2011), “el análisis paralelo es el método más preciso para decidir el número de factores” (p.232). La técnica permite generar H matrices de datos aleatorios de orden igual al de la muestra empírica (con el mismo número de sujetos e ítems), de forma que las variables sean independientes (no existan factores comunes). El procedimiento consiste en obtener el promedio de cada autovalor a través de las H muestras y compararlo con el autovalor de la muestra empírica (Abad, Olea y Ponsoda, 2011). Para realizar el AP con FACTOR se siguió el procedimiento sugerido por Ferrando, y Anguiano-Carrasco (2010) que consiste en las siguientes etapas:

1. Verificar la adecuación de las variables: media, desviación típica y asimetría de las distribuciones de los ítems, cuyo valor debe oscilar entre -1 y +1.
2. Verificar la adecuación de los datos utilizando la Prueba de esfericidad de Barlett (1950), que ha de ser significativa, y el índice KMO de Kaiser que ha de ser $\geq .80$.
3. Extracción de factores mediante el método de Mínimos Cuadrados no Ponderados (Jöreskog, 1977).
4. Determinación del número de factores mediante el método de Análisis Paralelo de Horn (1965): se compara el porcentaje de la varianza explicado por cada autovalor empírico con el porcentaje explicado por el promedio de los autovalores aleatorios de cada factor. Se

seleccionan tantas dimensiones como aquellas en las que el porcentaje empírico sea mayor que el aleatorio (Ferrando y Anguiano-Carrasco, 2010).

Adicionalmente al empleo de un análisis *omnibus* del ajuste al modelo como el realizado con los estadísticos *Infit* y *Outfit*, se ha generalizado en los últimos años la utilización de contrastes específicos del supuesto más importante del modelo: la unidimensionalidad. En el contexto del modelo de Rasch, la unidimensionalidad suele analizarse mediante el Análisis de Componentes Principales de los residuos (Chou & Wang, 2010). Aunque el logro de una unidimensionalidad perfecta es muy difícil en la práctica (Zickar & Broadfoot, 2009), el Análisis de Componentes Principales (ACP) de los residuos permite evaluar si la dimensión extraída con el modelo de Rasch explica una cantidad suficiente de la varianza de las observaciones, y si en la varianza no explicada (residual) existe algún factor o dimensión adicional con suficiente entidad para amenazar la validez de las puntuaciones obtenidas. Suele afirmarse (Chou & Wang, 2010; Smith & Miao, 1994) que los datos son esencialmente unidimensionales si la medida Rasch da cuenta de un porcentaje de la varianza moderadamente alto (al menos un 20% en opinión de Reckase, 1997) y si además el autovalor del primer componente de los residuos es bajo (2.0). El ACP de los residuos se llevó a cabo con el programa Winsteps (Linacre, 2012).

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 puntuaciones obtenidas en la primera evaluación con HoNOS (pre-tratamiento) y 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.

Se llevaron a cabo análisis estadísticos a nivel grupal y a nivel individual. A nivel grupal se contrastó la significación de la diferencia entre las medias de las puntuaciones en el pre-tratamiento y el post- tratamiento mediante 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 calcular el cambio a nivel de las personas, se utilizó el Índice de Cambio Fiable (ICF) con la puntuación clásica. El ICF representa la diferencia mínima entre las puntuaciones pre-post que una persona necesita obtener para rechazar la hipótesis nula de que no se ha producido un cambio. La fórmula para el cálculo del ICF en la TCT de acuerdo al criterio de Jacobson y Truax (1991) es:

$$\text{ICF} = 1.96 * \text{SEdiff} \quad (4)$$

siendo: SEdiff = Error típico de medida de la diferencia $X_1 - X_2$ (pre-test – post-test); 1.96 = el valor de la diferencia en errores típicos que es suficiente para rechazar la hipótesis nula ($\alpha = .05$; contraste bilateral).

La fórmula para la diferencia del error típico de la diferencia es:

$$\text{SE}_{\text{diff}} = \sqrt{2(\text{SE})^2} \quad (5)$$

siendo: $\text{SE} = s_1 \sqrt{1 - r_{xx}}$; s_1 = Desviación típica de pre-test; r_{xx} = coeficiente de fiabilidad del pre-test.

En consecuencia, el cambio de una persona será fiable si la diferencia absoluta entre el pre-tratamiento y el post-tratamiento es mayor que ICF (una diferencia positiva indicaría mejoría y una diferencia negativa revelaría deterioro). ICF es el punto de corte para rechazar la hipótesis nula de que no se ha producido cambio.

Sin embargo, se podría definir la hipótesis nula como la inexistencia de mejoría y contrastarla mediante un contraste unilateral, definiendo en este caso un Índice de Mejoría Fiable (IMF) que sería el punto de corte a partir del cuál es posible afirmar que la diferencia no indica simplemente cambio, sino mejoría fiable.

En este caso, IMF sería igual a:

$$\text{IMF} = 1.64 * \text{SEdiff} \quad (6)$$

Para realizar el análisis Rasch del cambio o la mejoría fiable entre las puntuaciones de la HoNOS en el pre-tratamiento (pre) y el post-tratamiento (post), previamente fue necesario realizar un escalamiento en la misma métrica, debido a que en las calibraciones separadas del pre-tratamiento y el post-tratamiento el punto cero de la escala se sitúa por defecto en la dificultad media de los ítems (Prieto y Velasco, 2003).

El procedimiento para poner en la misma métrica una escala evaluada en distintas situaciones consiste en primer lugar en calibrar los datos por separado, calculando la dificultad de los ítems (Di) para cada uno de los tests utilizados. Después de verificada la invarianza de los parámetros de dificultad de los ítems mediante la correlación de Pearson, se ha de tomar la decisión del lugar en el que se quiere situar el origen de la escala general (el punto cero); se elige uno de los dos test de referencia y se fija la dificultad de los ítems del otro en los valores de aquél (Prieto y Velasco, 2003).

Se ha de notar que en el modelo de Rasch no se asume que el error típico de medida es el mismo para todas las puntuaciones. Es decir, el error estándar es específico de cada punto de la variable. En consecuencia que una diferencia entre dos puntuaciones Rasch sea fiable depende de la magnitud del error estándar de las puntuaciones y esa magnitud cambia a

lo largo de la variable medida. Por ello, en el caso de la escala Rasch, en lugar de un punto de corte para definir ICF e IMF, se usa un contraste de la hipótesis nula para decidir si un sujeto ha cambiado o mejorado en el post-tratamiento

$$\text{El contraste sería igual a } RC = B_1 - B_2 / \sqrt{SE_1^2 + SE_2^2} \quad (7)$$

Siendo B_1 la puntuación de la persona en el pre-test y B_2 la puntuación en el post-test y SE_1 e SE_2 los errores típicos de medida asociados a B_1 y B_2 .

En un contraste bilateral se decide que el *cambio* del sujeto es fiable si el valor absoluto de RC es mayor de 1.96.

En un contraste unilateral se decide que la *mejoría* del sujeto es fiable si el valor algebraico de RC es mayor de 1.64.

Para llevar a cabo una validación de criterio de los índices ICF e IMF, y de los indicadores de cambio y mejoría correspondientes al modelo de Rasch, se analizó la convergencia entre la clasificación llevada a cabo con los puntos de corte de HoNOS y el punto de vista del criterio de los clínicos considerando a los pacientes que estaban en condiciones de ser dados de alta (egresar del hospital). Al compararse los pacientes que muestran mejoría en HoNOS con un criterio como el egreso (alta), se esperaba que la relación fuese alta y que, si no todos, al menos la mayoría de los pacientes considerados de alta por los clínicos presentaran un cambio y/o mejoría fiable.

Para analizar el cuestionario de viabilidad de la HoNOS se calcularon las frecuencias de las respuestas de los clínicos participantes y las puntuaciones medias en cada ítem.

3. Resultados

3.1. Análisis del funcionamiento de las categorías de respuesta de la HoNOS

En primer lugar, se procedió al análisis de las categorías de respuesta de la HoNOS, de acuerdo con los criterios de Linacre (2002).

La primera recomendación de Linacre se cumple adecuadamente dado que el número de observaciones por categoría es elevado (la menor frecuencia es de 122). Por lo que respecta a la segunda recomendación, se observa que la distribución es adecuada (con la mayor frecuencia en la categoría 0). La tercera recomendación se cumple, ya que existe un incremento de los promedios en la escala de las personas que eligen cada categoría. También se cumple la cuarta recomendación ya que no se observan categorías desajustadas, siendo todos los valores de *Outfit* inferiores a 2.0. La quinta recomendación no se cumple ya que los pasos entre las categorías sucesivas no están ordenados monotónicamente. Esto significa que no todas las categorías de respuesta son funcionales. De acuerdo con Linacre (2002) las categorías deben de ser las más probables en un rango suficientemente grande de la variable medida (es aconsejable una distancia de al menos 1.0 *logit* entre los pasos de categorías sucesivas).

Los resultados de los análisis sobre la adecuación métrica de las categorías de respuesta mediante el MEC se presentan en la Tabla 7.

Tabla 7

Calidad psicométrica de las categorías de respuesta de la HoNOS según criterios de Linacre (2002)

Criterio	Indicador	Categoría	HoNOS	
1	Al menos 10 observaciones en cada categoría de respuesta	0	2528 (52%)	
		1	777 (16%)	
		2	862 (18%)	
		3	535 (11%)	
		4	122 (3%)	
2	Distribución regular de las observaciones		Si	
3	Incremento monotónico de los promedios de las medidas de las personas que eligen las categorías	0	-1.51	
		1	-1.00	
		2	-.70	
		3	-.47	
		4	-.23	
4	<i>Outfit</i> inferior a 2.0		Infit	Outfit
		0	.97	.99
		1	.93	.69
		2	.87	.82
		3	1.10	1.41
4	1.18	1.29		
5	Orden monotónico de los pasos	0	----	
		1	-.11	
		2	-1.01	
		3	-.10	
		4	1.22	

En la Figura 1 se muestra el funcionamiento de las cinco categorías de la HoNOS.

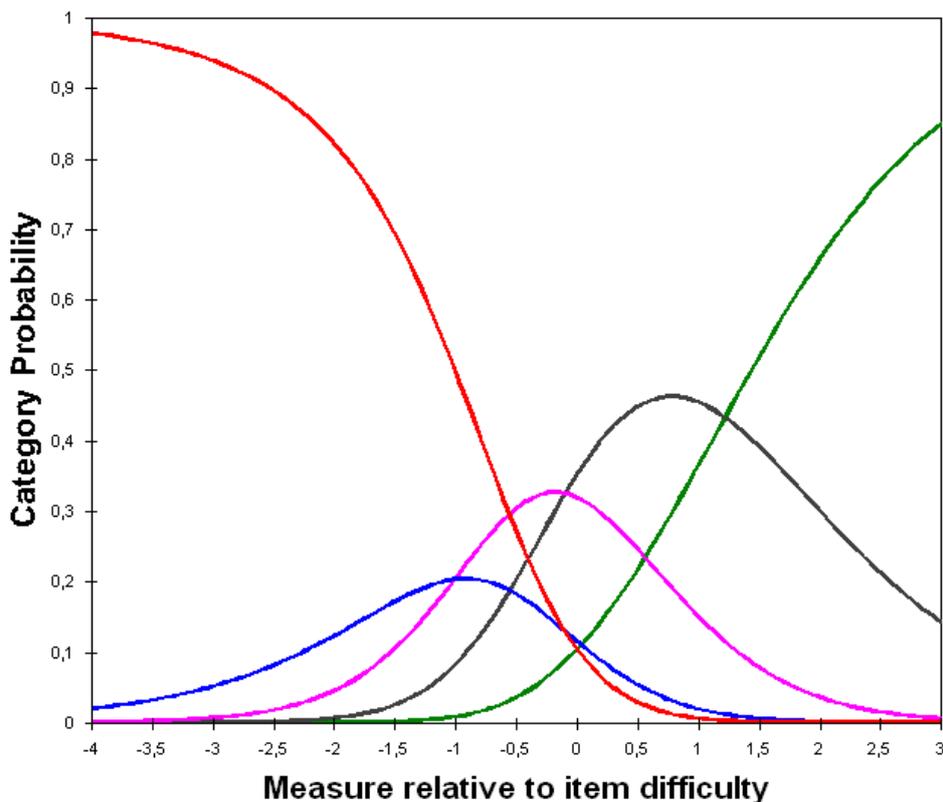


Figura 1. Representación gráfica de las curvas características de las categorías de respuesta de la HoNOS.

3.2. Construcción de un sistema apropiado de categorías de la HoNOS

Después de verificado que el funcionamiento de la HoNOS original no es adecuado, las categorías debieron ser redefinidas mediante una agrupación de categorías adyacentes, lo que llevó a reducir las cinco categorías originales a cuatro. La agrupación a cuatro categorías 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). En la Tabla 8 se muestran los resultados con la nueva agrupación de categorías.

Tabla 8

Calidad psicométrica de las categorías de respuesta de la HoNOS, con agrupación de categorías según criterios de Linacre (2002)

Criterio	Indicador	Categoría	HoNOS	
1	Al menos 10 observaciones en cada categoría de respuesta	0	2528 (52%)	
		1	1639 (34%)	
		2	535 (11%)	
		3	122 (3%)	
2	Distribución regular de las observaciones		Si	
3	Incremento monotónico de los promedios de las medidas de las personas que eligen las categorías	0	-2.20	
		1	-1.21	
		2	-.72	
		3	-.24	
4	<i>Outfit</i> inferior a 2.0		<i>Infit</i>	<i>Outfit</i>
		0	.96	.97
		1	.85	.73
		2	1.10	1.36
		3	1.18	1.35
5	Orden monotónico de los pasos	0	----	
		1	-1.30	
		2	.15	
		3	1.15	

Esta agrupación mostró pasos ordenados monotónicamente, obteniendo un sistema de categorías que cumple con las recomendaciones de Linacre (2002) y presenta una funcionalidad más eficaz (ver Figura 2).

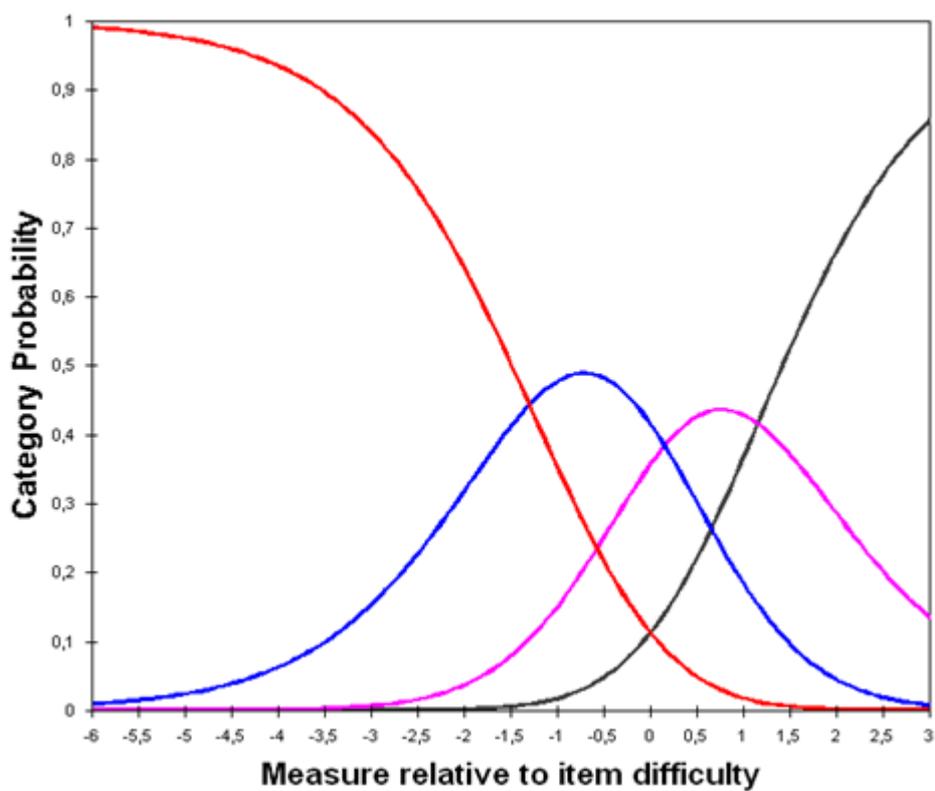


Figura 2. Representación gráfica de las curvas características de las categorías de respuesta de la HoNOS agrupadas (4 categorías).

3.3 Análisis de la dimensionalidad

Respecto a la dimensionalidad de la HoNOS, en el contexto del AFE, el Análisis Paralelo (AP) revela que solamente existe una única dimensión básica, dado que sólo el porcentaje de la varianza explicado por el autovalor del primer factor empírico (32.1%) era 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%).

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 superior al valor mínimamente aceptable según el criterio de Reckase (1997). Además, el autovalor correspondiente al primer componente de los residuos es sólo ligeramente superior a 2 (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.

3.4. Análisis de los valores de las personas e ítems de la HoNOS

Una vez corroborado que con la agrupación de las categorías de respuesta se obtiene un sistema métrico más adecuado, y que las medidas obtenidas con la HoNOS son

fundamentalmente unidimensionales, se presentan los resultados de los análisis con TCT y Rasch de las medidas de las personas y los ítems con la nueva agrupación.

En la Tabla 9 se presenta los estadísticos descriptivos con las puntuaciones clásicas y en logit, y un resumen de los resultados del análisis del ajuste de las personas al modelo de Rasch.

Se observa que el promedio de los estadísticos de ajuste de las personas es razonable: pues los valores de la media y de la desviación típica son respectivamente de 1.0 y .54 (*Infit*) y 1.0 y .66 (*Outfit*). Además, la mayoría de las personas se ajustaron al modelo: sólo 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).

La localización de las personas (en logit) se ubica en un rango de -5.18 a .82 ($M = -1.67$; $DT = .88$) y con las puntuaciones directas entre 0 a 24 ($M = 7.6$, $DT = 4.2$), lo que muestra que, aunque aparece una variabilidad moderadamente alta, el nivel de severidad de las personas de la muestra es bajo. Los errores típicos de medida de las personas van de .36 a 1.84. La fiabilidad global de las personas ($PSR = .59$, α de Cronbach = .69), resulta levemente inferior al valor recomendado para una fiabilidad aceptable (.70).

Tabla 9

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
Max	24	.82	3.0	4.5
α	.69	--	--	--
PSR	--	.59	--	--
% D1	--	10.9	10.9	--
% D2	--	9.2	9.2	--

Nota. X = Puntuación clásica. B = valores Rasch (logits). M = Media. DT = desviación típica. Min = valor mínimo. Max = valor máximo. α = alpha de Cronbach. PSR = Person Separation Reliability (fiabilidad Rasch). % D1 = porcentaje de personas con *Infit* y/o *Outfit* > 1.5 y < 2. % D2 = porcentaje de personas con *Infit* y/o *Outfit* > 2.

La Tabla 10 muestra la localización o nivel de dificultad (en *logit*), el error estándar asociado y los estadísticos de ajuste (*Infit* y *Outfit*) para todos los ítems de HoNOS. Además de los estadísticos de los ítems propios del modelo de Rasch, se incluyen las correlaciones ítem-escala (r_{iX}) con las puntuaciones directas.

En relación al ajuste de los ítems al modelo de Rasch, los valores medios de *Infit* se encuentran en un rango entre .53 a 1.74 y los de *Outfit* entre .52 a 1.91. 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); se ha de subrayar que ningún ítem presenta un desajuste severo con valores superiores a 2.00 (Linacre y Wright, 2000).

La mayoría de los ítems mostró adecuada correlación ítem-escala ($r_{iX} > .30$), excepto los ítems 5 y 7 que presentaron correlaciones de .12 y .24 respectivamente. El análisis de la consistencia interna de la HoNOS a través del α de Cronbach arrojó un valor de .69 (muy próximo a .70, valor considerado como mínimamente aceptable).

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 Item Separation Reliability es muy alta (.99).

Tabla 10

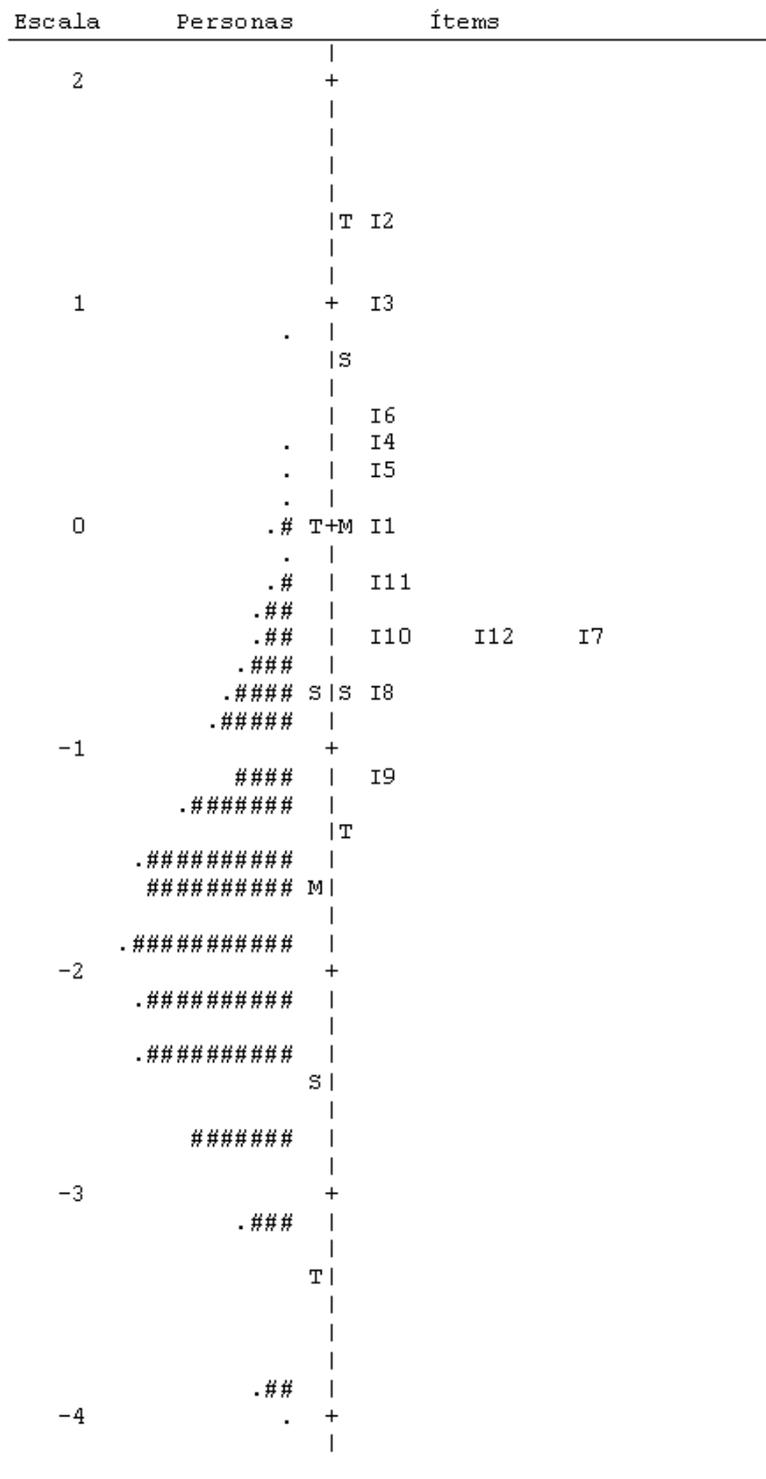
Dificultad, error estándar, estadísticos de ajuste y correlaciones ítem-escala (r_{iX}) de los ítems de HoNOS

Ítem	Dificultad	Error estándar		(r_{iX})	
	(en logit)	(en logit)	Infit	Outfit	
1	.06	.08	.83	.82	.55
2	1.36	.12	1.13	.80	.45
3	1.02	.11	1.57	1.45	.33
4	.35	.09	.99	.97	.43
5	.23	.08	1.74	1.91	.12
6	.50	.09	1.62	1.46	.41
7	-.52	.07	1.22	1.29	.24
8	-.70	.07	1.17	1.21	.41
9	-1.08	.07	.53	.55	.66
10	-.54	.07	.53	.52	.68
11	-.22	.07	.88	.81	.59
12	-.44	.07	.75	.70	.62

En la figura 3 se muestra una representación conjunta de personas-ítems, que permite observar la distribución de todas las personas de la muestra y de los ítems a lo largo de la variable en logit. Puede observarse en la figura 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 mostrando que el nivel de severidad de los sujetos de la muestra es muy bajo, y la HoNOS sólo tiene ítems situados en el rango superior del constructo. En consecuencia, se observa que los ítems de la escala corresponden a indicadores de un grado elevado de severidad, por lo que la escala mide con una menor precisión a las personas con niveles bajos de severidad.

Se observa en el mapa que el ítem 2 “autolesiones no accidentes” resulta el indicador de mayor nivel de gravedad de la HoNOS y el ítem 9 “problemas en las relaciones” el de menor nivel.



Nota: "#" = 4, "." = 1 a 3, M = Media.
 S = M + 1 Desviación Típica.
 T = M + 2 Desviaciones Típicas.

Figura 3. Representación conjunta de personas e ítems de HoNOS.

Al analizar el nivel de las puntuaciones en relación con el sexo, se observa que los hombres de la muestra presentan mayor nivel de severidad que las mujeres. En la Tabla 11 se muestran los estadísticos descriptivos de los valores medios en la medida de los hombres y mujeres, y los resultados obtenidos con un ANOVA tomando como variable dependiente las puntuaciones del modelo de Rasch y clásicas (TCT). Las diferencias resultaron estadísticamente significativas con los valores Rasch ($F = 17.00$, $gl=1$, $p<.001$) y con la TCT ($F=14.30$, $gl=1$ $p<.001$).

Tabla 11

Estadísticos descriptivos y diferencias en logit entre las medias de los hombres y de las mujeres

	Grupo				ANOVA		
	Hombre (158)		Mujer (244)				
	Media	DT	Media	DT	F	g. l.	Sig.
Rasch	-1.45	.80	-1.81	.90	17.00	1	.000
TCT	8.61	4.16	7.02	4.09	14.30	1	.000

Nota. DT = desviación típica. F= F de Snedecor. g. l. = grados de libertad. Sig. = p.

En cuanto a los resultados según grupos diagnósticos, los puntajes más altos son obtenidos por los pacientes de los grupos Demencias, Abuso de sustancias y Esquizofrenias y

Trastornos Delirantes respectivamente, y los más bajos por los del grupo Neurosis/Ansiedad y Trastorno del sueño y Alimentación. En la Tabla 12 se han incluido los estadísticos descriptivos para cada diagnóstico de la muestra.

Tabla 12

*Estadísticos descriptivos en los diferentes grupos diagnósticos **

Grupo	n	Media (X)	DT (X)	Media (B)	DT (B)
Demencia	8	10.75	4.0	-1.05	.65
Drogas / Alcohol / Abuso de sustancias	26	9.77	4.3	-1.23	.76
Esquizofrenia / Trastornos delirantes	100	9.19	4.9	-1.38	.90
Trastorno del estado de ánimo	110	6.98	3.8	-1.82	.93
Neurosis / Ansiedad	63	4.84	2.1	-2.24	.64
Trastorno sueño, alimentación, etc.	7	5.00	1.6	-2.13	.41
Trastorno de personalidad	51	7.92	3.3	-1.56	.70
Otros	34	8.12	4.2	-1.57	.87

Nota. * n = 399. X = Puntuación clásica (suma de los valores de los ítems). B = valores Rasch (logits).

3.5. Validez discriminante de la HoNOS

Para obtener pruebas de criterio, 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 atendidos los pacientes es una información externa asociada a la severidad de los problemas de salud mental de los pacientes: los servicios ambulatorios atienden a pacientes con menores niveles de severidad, mientras que los de internación a pacientes 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 espera que existan diferencias significativas entre los grupos.

La Tabla 13 muestra los estadísticos descriptivos del grupo ambulatorio y del grupo internado y los resultados obtenidos a través del ANOVA. Como puede apreciarse, los pacientes internados presentan puntuaciones más altas que los del grupo ambulatorio y las diferencias son estadísticamente significativas en las puntuaciones Rasch ($F = 48.05$, $gl=1$, $p<.001$) y en las puntuaciones clásicas ($F = 66.49$, $gl=1$, $p<.001$). 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 $.73$).

Es importante notar que 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 = $.61$).

Tabla 13

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

	Grupo				ANOVA		
	Ambulatorio (285)		Internado (117)		F	g. l.	Sig.
	Media	DT	Media	DT			
Rasch	-1.85	.79	-1.22	.92	48.05	1	.000
TCT	6.64	3.29	10.11	5.04	66.49	1	.000

Nota. DT = desviación típica. F= F de Snedecor. g. l. = grados de libertad. Sig. = p.

3.6. Sensibilidad al cambio de la HoNOS

Para estudiar la sensibilidad al cambio se compararon las puntuaciones totales y en cada ítem en la HoNOS de las personas evaluadas en los dos tiempos definidos (pre- y post-tratamiento): en el grupo internado al ingreso y al egreso, y en los pacientes ambulatorios al inicio del tratamiento y en la consulta de control.

De la muestra total, 75 personas presentaron dos evaluaciones con los criterios de inclusión para estudiar la sensibilidad al cambio de la HoNOS. De la muestra evaluada en el pre- y post-tratamiento; 24 son hombres y 51 mujeres con una media de edad de 38.21 (DT =

12.00); 39 pertenecen al grupo ambulatorio y 36 al grupo internado. La Tabla 14 muestra los diagnósticos principales del grupo ambulatorio e internado evaluado en los dos tiempos.

Tabla 14

Diagnósticos principales del grupo ambulatorio e internado evaluado en el pre-tratamiento y post-tratamiento

Diagnóstico	Grupo de estudio	
	Ambulatorio (n = 39)	Internado (n = 36)
Demencia	0	0
Drogas / Alcohol / Abuso de sustancias	3	3
Esquizofrenia / Trastornos delirantes	2	19
Trastorno del estado de ánimo	10	9
Neurosis/ansiedad	13	0
Trastorno sueño, alimentación, etc.	1	0
Trastorno de personalidad	8	2
Otros	2	3

En la Tabla 15 se puede apreciar la diferencia entre las medias de las puntuaciones en el pre-tratamiento (1) y el post-tratamiento (2) con puntuación clásica y valores Rasch. Con la

puntuación clásica se observa un cambio en los pacientes en la puntuación media total de HoNOS de 5.97 entre el pre-tratamiento (9.36, DT = 4.53), y el post-tratamiento (6.2, DT = 3.39). Esta diferencia resultó estadísticamente significativa cuando se contrastó con la prueba *t* de Student para muestras apareadas ($t = 7.66$, $gl = 74$, $p < .001$). El tamaño del efecto de esta diferencia es mediano (d de Cohen = .79). Con el análisis Rasch la diferencia en la puntuación entre el pre-tratamiento y post-tratamiento fue de 1.04 logit y resultó estadísticamente significativa ($t = 10.56$, $gl = 74$, $p < .001$), siendo grande el tamaño del efecto correspondiente (d de Cohen = 1.08).

Tabla 15

Estadísticos descriptivos y diferencia entre las medias de las puntuaciones en el pre y post-tratamiento

Escala	M(t1)	DT(t1)	M(t2)	DT(t2)	t	g.l.	Sig.	<i>d</i>
TCT	9.36	4.53	6.2	3.39	7.66	74	.00	.79
Rasch	-1.18	.79	-2.22	1.10	10.56	74	.00	1.08

Nota. M = Media. DT = Desviación típica. t1 = pre-tratamiento. t2 = post-tratamiento. $t = t$ Student; g. l. = grados de libertad; Sig = significación (p). d = tamaño del efecto (d de Cohen).

Respecto al cambio en los ítems de la HoNOS, con la puntuación directa, se observaron cambios estadísticamente significativos ($p \leq .05$) en 11 de los 12 ítems. La excepción fue el ítem 4 (problemas cognitivos). La Tabla 16 muestra los estadísticos

descriptivos con las puntuaciones directas, el estadístico *t* de Student y el *p* valor para las diferencias entre las medias entre el pre- y post-tratamiento en cada ítem de la HoNOS.

Tabla 16

Estadísticos descriptivos y diferencias entre el pre-tratamiento y el post-tratamiento en los ítems de la HoNOS

Ítem	Dif. Media	DT	t	p
1	.40	.66	5,268	.000
2	.31	.68	3,921	.000
3	.16	.49	2,803	.006
4	.07	.41	1,395	.167
5	.08	.32	2,173	.033
6	.36	.71	4,392	.000
7	.35	.69	4,366	.000
8	.49	.88	4,879	.000
9	.47	.66	6,083	.000
10	.20	.47	3,725	.000
11	.13	.46	2,594	.011
12	.15	.49	2,620	.011

Nota. Dif Media = diferencia entre el pre-tratamiento y post-tratamiento. DT = desviaciones típicas. t = t de Student. p = p valor.

En relación al análisis del cambio en la puntuación media total de HoNOS entre el pre- y post-tratamiento para el grupo ambulatorio e internado, en la Tabla 17 se muestra los estadísticos descriptivos, los contrastes *t* y el tamaño de efecto calculados con la puntuación

clásica y en logit. Como puede apreciarse, en ambos grupos se presentan cambios estadísticamente significativos en la puntuación media total en HoNOS entre el pre y el post-tratamiento, siendo mayor el cambio y el tamaño del efecto en el grupo internado que en el grupo ambulatorio.

Tabla 17

Estadísticos descriptivos y diferencias entre las medias del pre- y post-tratamiento en el grupo ambulatorio y el grupo internado

Escala	Grupo	M (t1)	DT (t1)	M (t2)	DT (t2)	Dif. (t1-t2)	t	p	d
TCT	Ambulatorio	7.74	3.17	6.23	2.86	1.51	5.153	.000	.50
	Internado	11.11	5.15	6.17	3.93	4.94	7.190	.000	1.08
Rasch	Ambulatorio	-1.42	.670	-2.14	.942	.724	8.866	.000	0.88
	Internado	-.93	.834	-2.30	1.27	1.38	8.176	.000	1.28

Nota. M = Media. DT = Desviación típica. t1 = pre-tratamiento. t2 = post-tratamiento.

Dif.(t1-t2) diferencia pre-tratamiento- post-tratamiento. t = t Student. p = p valor. d = d de Cohen.

En el análisis a nivel individual del cambio en la HoNOS con la puntuación directa, se observa que con el ICF obtenido, 8 personas (10.7% de la muestra) presentan un cambio fiable (cambio superior a 7.3) y con el IMF, 11 personas que corresponden al 14.7% de la muestra presentan una mejoría fiable (mejoría superior a 6.07). Todas las personas que

presentan un cambio fiable y la casi totalidad (10 de las 11) de las personas que presentaron una mejoría fiable pertenecen al grupo internado.

Para el análisis del cambio en las puntuaciones Rasch, en primer lugar se analizó la invarianza de los ítems entre el pre-tratamiento y el post-tratamiento. La correlación entre los parámetros de dificultad de los ítems en el pre y el post fue muy alta (.90), lo cuál sugiere una aceptable invarianza. No obstante un contraste de significación t indicó que la dificultad de 5 ítems se modificó significativamente entre el pre-tratamiento y el post-tratamiento. Los ítems que no variaron fiablemente en su dificultad entre el pre y el post fueron: 1, 3, 4, 5, 7, 8 y 9. En consecuencia, se calibró la variable post-tratamiento en la misma métrica que la variable pre-tratamiento fijando el parámetro de dificultad de los ítems invariantes y dejando fluctuar la dificultad del resto de los ítems en la escala post-tratamiento. En la Tabla 18 se presentan los estadísticos descriptivos del pre- y post-tratamiento en la HoNOS y los resultados de la calibración.

Tabla 18

Estadísticos descriptivos de las puntuaciones de los sujetos en el pre-tratamiento y post-tratamiento y datos de anclaje

Ítem	Di pre	SE pre	Di post	SE post	Di Post (anclada en pre)
1	.00	.16	.57	.26	.00
2	.68	.20	1.91	.37	1.90
3	.93	.21	1.35	.31	.93
4	.72	.20	.45	.25	.72
5	1.08	.23	1.09	.29	1.08
6	.28	.17	1.09	.29	1.08
7	-.28	.15	-.31	.22	-.28
8	-.80	.14	-.90	.19	-.80
9	-1.08	.14	-1.47	.17	-1.08
10	-.50	.14	-1.15	.18	-1.06
11	-.46	.15	-1.25	.18	-1.15
12	-.56	.14	-1.38	.18	-1.27

Nota. Di pre = parámetro de dificultad del ítem del pre-tratamiento. SE pre = error estándar del ítem del pre-tratamiento. Di post = parámetro de dificultad del ítem del post-tratamiento (calibración por separado). SE post = error estándar del ítem del post-tratamiento (calibración por separado). Di Post = parámetro de dificultad de los ítems del post-tratamiento después de colocarse en la misma métrica.

Una vez realizado el escalamiento se procedió al análisis del cambio individual en valores Rasch. Mediante la Razón Crítica (ecuación 7) se determinó que 15 personas (20 %) presentaron un cambio fiable y que 22 personas (29.3%) presentaron una mejoría fiable.

Cuando se analiza el cambio por grupo de pacientes, se observa que la mayoría (86.7%) de las personas que presentan cambio fiable y la mayoría (81.8%) con mejoría fiable, pertenecen al grupo de pacientes internado.

El análisis de la convergencia entre la clasificación con los puntos de corte de la HoNOS medidos con los índices ICF, IMF, y las RC, y el punto de vista del criterio de los clínicos considerando a los pacientes que estaban en condiciones de ser dados de alta (egresar del hospital), muestra unos índices de *sensibilidad* (proporción de egresados que presentaban indicadores de cambio o mejoría) entre bajos y moderadamente bajos. Como puede apreciarse en la tabla 19, la relación más alta se observó con el indicador de $Mejoría_{Rasch}$ con el que el 46.7% de los pacientes egresados presentaron un índice de mejoría fiable. Con los demás índices (ICF, $Cambio_{Rasch}$, IMF) el porcentaje de pacientes en situación de egreso que presentan cambio o mejoría fiable es menor (13.3%, 20% y 30% respectivamente).

Tabla 19

Análisis del cambio individual en HoNOS entre el pre-tratamiento y el post-tratamiento medidos con los índices ICF_{TCT} , IMF_{TCT} , $Cambio_{Rasch}$ y $Mejoría_{Rasch}$

Estadístico	Valor	N_T	$\%_T$	N_I	$\%_I$	Sensibilidad _I
ICF_{TCT}	7.3	8	10.7	8	22.2	13.3
IMF_{TCT}	6.07	11	14.7	10	27.8	20.0
$Cambio_{Rasch}$	-	15	20	13	36.1	30.0
$Mejoría_{Rasch}$	-	22	29.3	18	50.0	46.7

Nota. Valor= Punto de corte de ICF_{TCT} o IMF_{TCT} : las personas con una diferencia mayor cambiaron o mejoraron fiablemente; N_T = número de casos con cambio o mejoría fiable en la muestra total; $\%_T$ = porcentaje de casos con cambio o mejoría fiable en la muestra total (N=75); N_I = número de casos con cambio o mejoría fiable en el grupo internado; $\%_I$ = porcentaje de casos con cambio o mejoría fiable en la muestra del grupo internado (N=36); Sensibilidad_I = Porcentaje de los pacientes del grupo internado egresados que presentaron un cambio o mejoría fiable según HoNOS.

3.7. Viabilidad de la HoNOS

El cuestionario de viabilidad fue respondido por 21 de los 22 clínicos participantes en el estudio.

Las frecuencias y medias de las respuestas a las preguntas del cuestionario se presentan en la Tabla 20. Como se puede observar, el ítem que presenta una puntuación media más baja es la probable utilización de la HoNOS (3.19, DT=1.12) y el ítem sobre el requerimiento de capacitación (5) presenta la puntuación media más alta (3.90, DT = 1.04).

En relación a los ítems que recogen los aspectos relacionados a la aplicabilidad (utilidad y relevancia) y a la aceptabilidad (facilidad) la valoración de los clínicos es de moderada a alta. Al observar la frecuencia de las categorías de respuestas; la mayoría de los clínicos (71.4 %) seleccionaron las opciones de respuesta 4 o 5 (bastante o total) para valorar la relevancia de la HoNOS.

Los ítems que recogieron la valoración sobre la practicidad de la HoNOS (ítem 4: tiempo e ítem 5: capacitación) presentaron una puntuación alta y moderadamente alta con una media de 3.43 y 3.90 respectivamente.

Respecto a la pregunta que intentó evaluar el grado en que utilizarían la HoNOS para evaluar los resultados en su práctica clínica; 9 clínicos respondieron que con bastante o total probabilidad (opciones 4 y 5 respectivamente) la utilizarían, mientras que en el otro extremo, 2 respondieron que con ninguna probabilidad.

Tabla 20

Frecuencias, medias y desviaciones típicas de las respuestas de los clínicos en el cuestionario sobre viabilidad de la HoNOS

Contenido	Frecuencia de repuestas*					Media (DT)
	1	2	3	4	5	
Utilidad	0	4	8	7	2	3.33 (.91)
Relevancia	0	2	7	11	1	3.52 (.75)
Facilidad para aplicar	0	2	4	11	4	3.81 (.87)
Tiempo requerido	0	3	9	6	3	3.43 (.93)
Capacitación requerida	0	2	6	5	8	3.90 (1.04)
Probable utilización	2	3	7	7	2	3.19 (1.12)

Nota. * n = 21. 1 = ninguna / nada. 2 = algo. 3 = regular. 4 = bastante. 5 = máximo / total.

4. Discusión

El objetivo principal fue llevar a cabo un estudio de validación de la versión al castellano de la escala HoNOS (Wing et al., 1996) utilizando la Teoría Clásica de los Tests y el Modelo de Escalas de Calificación (MEC). La muestra incluyó 402 usuarios adultos de

servicios de salud mental de Uruguay, 285 atendidos en servicios ambulatorios y 117 en régimen de internación psiquiátrica.

El estudio analizó las propiedades de la HoNOS en cuanto a su validez, fiabilidad, sensibilidad al cambio y viabilidad. Dado que la HoNOS es un instrumento desarrollado para ser utilizado por los clínicos en la práctica clínica habitual de los servicios de salud, la recolección de datos se llevó a cabo en las condiciones habituales de trabajo, intentando alterarla lo mínimo posible.

El análisis mediante el MEC de las categorías de respuesta determinó que las cinco categorías del sistema original de HoNOS no cumplen con todas las recomendaciones de Linacre (2002), dado que los pasos entre las categorías adyacentes no están ordenados monótonicamente. Este dato indica que alguna de las categorías (en concreto, la número 2) es inadecuada, puesto que no es la de más probable elección en ningún intervalo de la variable medida. En consecuencia, las categorías necesitaron ser redefinidas mediante una agrupación de categorías adyacentes. Hemos buscado una solución posible de mejora reduciendo las cinco categorías iniciales a cuatro. Después del reanálisis, se observó que con la agrupación de categorías de respuestas de la HoNOS, se cumple con los criterios de calidad propuestos por Linacre (2002).

El supuesto de unidimensionalidad se analizó mediante un Análisis Factorial Exploratorio de los datos en el que se estimó el número de dimensiones mediante el Análisis Paralelo. En el contexto del Modelo de Rasch se utilizó el Análisis de Componentes Principales de los Residuos. Mediante ambas técnicas se pudo concluir que los datos se adecuaron suficientemente a la dimensionalidad como para permitir el análisis mediante el modelo de Rasch y no amenazar la validez de los resultados.

La discusión sobre la estructura factorial de la HoNOS sigue siendo un tema recurrente en la literatura dado que los resultados de los distintos estudios han sido contradictorios (Lovaglio y Monzani, 2011a; Speak y Hay, 2012).

La consistencia interna de HoNOS, cuantificada mediante el coeficiente alpha de Cronbach, si bien no fue óptima, (el valor recomendado para considerarlo adecuado es .70), resultó aceptable (.69). Este resultado fue esperable dado 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, Burgess, Kirk, et al., 2005). Los ítems 5 y 7 presentaron menores índices de discriminación clásica (correlación ítem-test) que el resto, por lo que decrementan ligeramente la consistencia interna de la escala. A pesar de este ligero decremento, no es aconsejable eliminar estos ítems, dado que no presentan un desajuste severo con el modelo de Rasch. Además, cuántas más y mayores modificaciones se realicen, el instrumento más se apartará de la versión original de la HoNOS hacia la creación de un instrumento distinto, lo que conllevaría la desventaja de no permitir las comparaciones internacionales, y capitalizar la importante experiencia acumulada en ella (Trauer y Buckingham, 2006).

Con el análisis Rasch de los estadísticos de ajuste, se identificaron 3 ítems con un desajuste moderado (ítem 3 - consumo problemático de alcohol o drogas, ítem 5 - enfermedad física, e ítem 6 – ideas delirantes y alucinaciones); aunque ningún ítem de la HoNOS presentó un desajuste severo. En cuanto a la fiabilidad global de los ítems, evaluada mediante el estadístico Item Separation Reliability, los resultados proporcionaron índices muy apropiados para los ítems (.99).

En relación a las personas, los promedios de los estadísticos de ajuste de las personas presentaron valores de Infit y Outfit adecuados, menos del 10% de la muestra presentó un

desajuste severo, pero una fiabilidad global (Person separation reliability = .66), levemente inferior al recomendable para una estimación precisa. 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 era mayor (Person Separation Reliability = .75, α de Cronbach = .71) que en el grupo ambulatorio (Person separation reliability = .54, α de Cronbach = .61).

Algunos estudios sugieren que la capacitación y la experiencia en el uso de la escala afectan la fiabilidad (Turner, 2004). Los clínicos de nuestro estudio no tenían experiencia previa en el uso de HoNOS, dado que comenzaron a utilizar la escala inmediatamente después de la capacitación. Futuros estudios deberían evaluar si con una capacitación más extensa y con el reiterado uso de la escala aumenta la fiabilidad.

También serán necesarios estudios que generen evidencia de la fiabilidad de la HoNOS midiendo su estabilidad temporal (test-retest), y la correlación entre las puntuaciones de distintos evaluadores (fiabilidad inter-jueces). Este tipo de estudios resultan más difíciles de realizar cuando se utiliza un diseño que intenta mantener lo máximo posible la práctica habitual de los clínicos, a la vez que no sobrecargarlos con actividades extraordinarias.

En cuanto a la distribución de los parámetros de las personas, pocos pacientes mostraron un nivel alto de severidad de trastorno mental, mientras que la mayoría de los pacientes mostraron bajos niveles. 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 Comunitarios de Salud Mental (ECSM), 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 atienden los ECSM de 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.

La medición conjunta de personas e ítems que permitió la utilización del modelo de Rasch, mostró que el promedio de severidad de los ítems fue superior al nivel presentado por la muestra; y no se observó un subconjunto de ítems para la medición de los niveles más bajos de severidad de los trastornos mentales. 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 podría tener importantes implicaciones prácticas para la utilización de la HoNOS en los servicios de salud mental de Uruguay, 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 del país, se reservara para pacientes con trastornos mentales más severos, y se

utilizara otro instrumento de resultados para medir a los pacientes con trastornos mentales más leves.

En cuanto a las diferencias por sexo, en nuestro estudio la gravedad de los problemas medidos con la HoNOS fue significativamente mayor en la muestra de hombres (158) respecto a la muestra de mujeres (244). En relación a los diagnósticos, los puntajes más altos en la HoNOS fueron obtenidos por los pacientes con diagnósticos de los grupos correspondientes a Demencias, Abuso de sustancias, y Esquizofrenias y otros trastornos delirantes respectivamente. Este resultado coincide con los hallazgos de otros estudios, en los cuales los puntajes más altos en HoNOS se asociaron a los pacientes con diagnósticos de consumo de drogas y alcohol y a pacientes psicóticos (Pirkis et al., 2005a).

Los resultados obtenidos en nuestro estudio mostraron buena validez discriminante de la HoNOS al hallar diferencias estadísticamente significativas entre los puntajes obtenidos por el grupo internado y el grupo ambulatorio, con un alto tamaño de efecto. Este resultado fue el esperado dado que los pacientes internados poseen mayores niveles de gravedad y se pretende que la HoNOS sea un instrumento eficaz para discriminar a sujetos con diferentes niveles de severidad de los problemas. Resultados similares han sido reportados en otros estudios internacionales (Pirkis, Burgess, Kirk, et al., 2005).

Futuras investigaciones deberían intentar obtener evidencia de la validez convergente de la versión al castellano de la HoNOS con otras medidas consolidadas para evaluar los resultados de los servicios salud mental. En nuestro estudio esto no fue posible porque el único instrumento psicométrico con estudios de validación en Uruguay es el SCL-90 (Najson, 2008) que es una escala de auto-informe, y existe suficiente evidencia que apoya la baja concordancia entre los resultados cuando se usan fuentes de información distintas (Pirkis,

Burgess, Kirk, et al., 2005; Trauer, 2010b). Sería importante poder evaluar la correlación de la HoNOS con alguna medida estandarizada que recogiera el juicio clínico.

Otro de los objetivos específicos del estudio fue analizar la sensibilidad al cambio de la HoNOS, para lo cual comparamos el cambio en la muestra de 75 pacientes que fue evaluada en dos tiempos. El grupo de pacientes internados fue evaluado al inicio y al alta (o próxima a esta) y el grupo ambulatorio al inicio del tratamiento y en la consulta de control.

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. Se observaron cambios estadísticamente significativos entre las medias de las puntuaciones totales del pre-tratamiento y el post-tratamiento con la puntuación clásica y con los valores Rasch, con un tamaño del efecto mediano y grande respectivamente. Respecto al cambio en los ítems, la diferencia entre las medias de las puntuaciones del pre-tratamiento y el post-tratamiento con la puntuación clásica, resultó estadísticamente significativa en 11 de los 12 ítems de la HONOS (la excepción fue el ítem 4 -problemas cognitivos).

Para realizar el análisis Rasch del cambio fiable previamente fue necesario situar los parámetros de localización de los ítems entre el pre-tratamiento y el post-tratamiento en una métrica común (Prieto y Velasco, 2003). Las calibraciones por separado de ambos tests mostraron que 5 ítems cambiaron en su dificultad en cantidades ligeramente superiores a .5 logit, y dado que la condición esencial para el escalamiento común es la invarianza de la localización de los ítems, se calibró la variable post –tratamiento en la misma métrica que la variable pre-tratamiento asignando en el post-test los parámetros de dificultad de los ítems invariantes en el pre-test. Esta calibración del post-test utilizando como anclaje los ítems

invariantes, mejora la precisión de la medición del cambio, y resulta de gran aporte para los estudios de resultados.

Cuando se realizó el análisis individual mediante el ICF y el IMF para observar el porcentaje de pacientes que cambió y presentó mejoría fiable luego del tratamiento, con la puntuación directa en HoNOS se observó que el 10.7% de la muestra presentó un cambio fiable, y con el IMF el 14.7% de la muestra presentó mejoría fiable. En el análisis Rasch, aumentó la capacidad de la HoNOS para registrar el cambio observándose un cambio fiable y una mejoría fiable en el 20% y 29.3% de la muestra respectivamente.

Una explicación sobre el modesto resultado obtenido con el cálculo del ICF e IMF, podría ir en la misma dirección a la que apuntaron otros estudios; la del efecto techo: los pacientes que presentan puntajes bajos en la HoNOS tienen menos oportunidades de poder mostrar un cambio que los pacientes que presentan puntajes iniciales más altos (Pirkis, Burgess, Kirk, et al., 2005); y la muestra de este estudio presentó bajos niveles de severidad, por tanto un bajo porcentaje de pacientes de la muestra tenía la posibilidad de presentar cambio y mejoría fiable.

Los cambios observados entre el pre-tratamiento y el post-tratamiento con todos los métodos de análisis utilizados (contraste de la *t* de Student, el tamaño del efecto, el cálculo del Índice de Cambio Fiable, y del Índice de Mejoría Fiable) fueron mayores en los pacientes internados que en los pacientes ambulatorios. Resultados similares obtuvieron otros estudios y fueron atribuidos a lo mencionado anteriormente; los pacientes internados en general presentan puntajes iniciales más altos que los pacientes ambulatorios, por lo cual, tienen más posibilidades de mostrar cambios en la HoNOS (Pirkis, Burgess, Kirk, et al., 2005; Delaffon et al., 2012). Además, forma parte de la complejidad que presenta la medición de resultados

en los servicios de salud ambulatorios que brindan tratamiento a personas que poseen trastornos mentales crónicos, el hecho de que en muchos casos una de las características de estos pacientes es la falta de cambio (Trauer, 2010b).

Por otra parte, en este estudio los pacientes internados fueron evaluados al inicio (ingreso) y al finalizar el tratamiento (alta), mientras que los pacientes de los servicios ambulatorios fueron evaluados al inicio y durante el tratamiento (en la consulta de control). A pesar de que los tratamientos en pacientes ambulatorios pueden durar años, o en pacientes crónicos incluso toda la vida, sería importante que futuros estudios evalúen el desempeño de la HoNOS en los pacientes ambulatorios al inicio y al finalizar el tratamiento, o al menos con períodos de tiempo más extensos entre las evaluaciones.

Con el propósito de obtener evidencia de validación de criterio de los índices ICF_{TCT} , IMF_{TCT} , $Cambio_{Rasch}$, y $Mejoría_{Rasch}$, se analizó la convergencia entre la clasificación llevada a cabo con los puntos de corte de HoNOS y el punto de vista del criterio de los clínicos considerando a los pacientes que estaban en condiciones de ser dados de alta (egresar del hospital). Se esperaba que la relación fuese alta y que, si no todos, al menos la mayoría de los pacientes considerados de alta por los clínicos presentaran un cambio y/o mejoría fiable. Contrariamente a lo esperado, se observaron índices de sensibilidad entre bajos y moderadamente bajos; la relación más alta se observó con el $Mejoría_{Rasch}$ donde el 46.7% de los pacientes egresados presentó un Índice de Mejoría Fiable.

El hecho de que nuestro estudio se llevara a cabo en el principal hospital psiquiátrico nacional, que es un hospital psiquiátrico monovalente donde el foco de atención son los problemas mentales, nos permite hipotetizar que, tal vez, las expectativas terapéuticas y el criterio de alta de los clínicos del estudio, no coinciden con lo que propone la HoNOS como

criterio para evaluar la efectividad de los tratamientos. La HoNOS se desarrolló para medir los resultados de los servicios de salud enmarcada en una estrategia nacional de salud que propone la mejoría de las personas con trastornos mentales en los distintos aspectos; físicos, mentales y sociales, teniendo todos ellos el mismo peso en la puntuación global de la HoNOS. Además, la población que se asiste en el Hospital, pertenece al sector de la población más vulnerable del país, donde es frecuente que se encuentren personas con problemas sociales, por ejemplo de vivienda graves, que no encuentran rápida ni sencilla solución. Complementariamente a esto, podemos comentar que durante la planificación del procedimiento para recolectar los datos del estudio, algunos de los clínicos plantearon si el criterio de egreso hospitalario, debía ser el egreso en sentido estricto, o el egreso según el criterio del clínico, dado que consideraban que existían aspectos “ajenos” al tratamiento clínico que en ocasiones llevaban a retrasar que se concretara el alta.

Es preciso notar que esta evidencia es limitada por el pequeño tamaño de la muestra utilizada y serían necesarios más estudios con una muestra de mayor tamaño para observar si los resultados se replican.

Otro objetivo específico del estudio de validación fue analizar la viabilidad de la HoNOS como medida para evaluar los resultados en la práctica clínica, para lo que se empleó un breve cuestionario que recogió la valoración de 21 de los 22 clínicos que participaron en el estudio recolectando los datos de sus pacientes.

Se construyó un breve cuestionario para recoger la siguiente información: (1) utilidad, (2) relevancia del contenido, (3) facilidad para su aplicación, (4) tiempo que insume (5) capacitación que requiere, (6) su probable utilización. Los ítems intentaron relevar los aspectos relativos a los criterios de: aplicabilidad, aceptabilidad y practicidad propuestos por

Andrews et al. (1994) y el grado de aceptación de acuerdo al criterio de Pirkis, Burgess, Kirk, et al., (2005).

Los datos obtenidos mostraron que la HoNOS fue vista por los clínicos como una medida que presenta de moderada a alta utilidad, relevancia y facilidad. Por otra parte los ítems que recogieron la valoración sobre el requerimiento de tiempo y capacitación, que apuntaron a evaluar la practicidad de la HoNOS, presentaron puntuaciones moderadamente altas. Todos estos aspectos han sido considerados clave por distintos estudios y revisiones para asegurar la viabilidad de utilizar una medida de evaluación rutinaria de resultados en los servicios de salud mental (Andrews et al., 1994; Pirkis, Burgess, Kirk, et al., 2005). Por tanto, estos datos sugieren una adecuada viabilidad de la HoNOS para la evaluación rutinaria de los resultados de los servicios de salud mental de Uruguay. Además la HoNOS cumple con otros requisitos importantes para la viabilidad de una medida de ROA dado que es breve y su uso no implica un costo económico. Sin embargo, la pregunta sobre el grado en que utilizarían la HoNOS para evaluar los resultados en su práctica clínica obtuvo la puntuación media más baja por parte de los clínicos, y se observó una dispersión en las frecuencias de respuestas que reflejó un diverso grado de aceptación de la escala.

Si bien no tenemos datos en este estudio para valorar el peso de cada uno de los ítems del cuestionario en la viabilidad de la HoNOS, una hipótesis plausible para explicar este resultado, en algún sentido contradictorio, es que mientras el resto de los ítems (1 al 5) implicaron la valoración de la escala en sí misma, la pregunta sobre su utilización futura requiere además valorar la posibilidad de incorporar una nueva práctica como lo es la evaluación de resultados con instrumentos psicométricos. Diversos autores han señalado que para lograr la evaluación rutinaria de los resultados de los servicios de salud mental, no es

suficiente con contar con el instrumento adecuado si no se consideran otros aspectos relativos a la implementación (Milne, Reichelt, y Wood, 2001). Existe abundante evidencia que demuestra que una de la barreras más importantes para implementar la evaluación rutinaria de resultados mediante la utilización de instrumentos psicométricos como la HoNOS, es la falta de convencimiento por parte de los clínicos sobre su utilidad y relevancia para la práctica clínica (Fonagy et al., 2005; Jacobs, 2009; Meehan et al. 2006). Diversos autores han sugerido que la implementación de ROA genera resistencia en los clínicos por el cambio cultural que implica (Coombs y McKay, 2012; Slade, 2012; Jacobs, 2009).

La mayor aportación de esta investigación consiste en permitir que la investigación futura en Uruguay utilice la HoNOS como una herramienta para la evaluación de los resultados de los servicios de salud mental. Esto es de gran importancia, ya que es urgente empezar a introducir instrumentos estandarizados y validados que permitan generar información sobre los resultados que obtienen los servicios de salud mental con sus tratamientos, y mejorar así la distribución productiva de los recursos y la calidad de la atención de las personas con problemas de salud mental. Sin embargo, para lograr que efectivamente se utilice la HoNOS, se debe prever que será necesario implementar estrategias para que los clínicos estén motivados y comprometidos, y para esto, deben poder ver los beneficios. Algunas de las recomendaciones para lograrlo han sido el intercambio (feedback) con los clínicos sobre los resultados obtenidos, brindar el tiempo necesario para la recolección de datos y la infraestructura necesaria. Simultáneamente, se requerirá trabajar en el cambio hacia una cultura que incorpore la medición de resultados con instrumentos estandarizados como una práctica de rutina.

Por último, incluimos algunas consideraciones respecto al cuestionario socio-demográfico y clínico complementario a la Escala HoNOS (CCH) que fue desarrollado para esta investigación y se utilizó en el estudio de validación. La versión utilizada del CCH fue desarrollada a partir de la revisión de la literatura y de los resultados obtenidos en el estudio piloto. Luego del estudio piloto, se modificó la versión inicial del CCH, incorporando algunas de las sugerencias de los expertos, entre ellas la inclusión de otras variables consideradas de interés para poder evaluar los resultados del tratamiento. Concretamente, las variables que se incluyeron fueron: “el tipo de tratamiento indicado”, y la variable “eventos relevantes desde la última aplicación de HoNOS.”

A pesar de que estos datos resultan de importancia para analizar los resultados, sobre todo a nivel del análisis de las diferencias individuales entre los pacientes; estos datos no pudieron ser analizados en nuestra investigación, porque los clínicos utilizaron diferentes criterios para completarlos. Las dificultades fueron básicamente dos; en primer lugar, que al no proponer categorías de respuesta cerradas, los clínicos completaron la información tomando distintos criterios que hicieron imposible la comparación (por ejemplo, en algunos casos incluyeron el tipo o el grupo de medicamento, en otros casos la marca comercial, en otros la dosis, etc.). La segunda dificultad encontrada fue que para relevar este aspecto se propuso seleccionar “el tratamiento indicado”, pero pudimos observar en las reuniones de capacitación que a los clínicos les generó la duda sobre qué debían consignar sobre los pacientes que iniciaban un tratamiento: si era el tratamiento que el clínico (del estudio) había indicado luego de la primera consulta, o el tratamiento que ya venía realizando el paciente previo a la primera consulta, o incluso el tratamiento que efectivamente estaba realizando el paciente (que podía no coincidir con lo que tenía indicado). Sumado a esto, los datos

obtenidos no fueron consistentes, por lo que decidimos dejarlos fuera del análisis de este estudio. Tampoco se pudo analizar la variable relativa a con quién vive la persona, porque no resultaron funcionales las opciones de respuesta que ofreció el cuestionario. En estudios futuros deberá revisarse la forma de incluir otras variables intervinientes como los tipos de tratamiento para evaluar los resultados que se obtienen con HoNOS. Además, será necesario modificar la variable con quien vive incluyendo más alternativas. Por último parece conveniente que el CCH se acompañe de un instructivo en el que se detalle la información que pretende relevar cada ítem del cuestionario y la forma en que deben completarse los datos requeridos para asegurar la calidad de los datos.

III. CONCLUSIONES

III. CONCLUSIONES

The Health of the Nation Outcome Scales (HoNOS) es uno de los instrumentos más utilizados a nivel internacional para evaluar los resultados de los servicios de salud mental como práctica de rutina (ROA). La HoNOS fue traducida a diferentes idiomas y sus propiedades psicométricas se estudiaron en diversos países. No hay ningún estudio empírico sobre las propiedades psicométricas de la versión al castellano (Uriarte et al., 1999). Por otra parte, no hay ningún instrumento de ROA con estudios de validación en Uruguay, que permita medir con ciertas garantías los resultados de los servicios de salud mental.

Por lo tanto, se pretendió con este trabajo generar evidencia sobre la validez, fiabilidad, sensibilidad al cambio y viabilidad de la versión al castellano de la HoNOS en servicios de salud mental públicos de Uruguay. Además, la investigación fue llevada a cabo en el contexto de la práctica habitual de los servicios de salud mental, para obtener evidencia sobre sus propiedades cuando se usa en el contexto para el que la HoNOS fue creada, y en el cual se pretende que sea utilizada.

Se han analizado las propiedades psicométricas de la HoNOS mediante la TCT y un modelo TRI, el Modelo de Escalas de Calificación (MEC). Este modelo facilita llevar a cabo ciertas mejoras de los instrumentos de medida, entre las que destacan la determinación empírica de un sistema adecuado de las categorías de respuesta, la medición conjunta de la variable latente en las personas y los ítems, el análisis de la precisión de las estimaciones de los parámetros en el atributo latente de las personas y los ítems, la cuantificación del grado de severidad de los indicadores del constructo medido y en los estudios longitudinales una métrica común para medir el cambio.

En una etapa previa al estudio de validación, se realizó un estudio piloto con el objetivo de someter al juicio de expertos la adecuación y necesidad de adaptar la HoNOS y el cuestionario socio-demográfico y clínico complementario a la HoNOS (CCH), para su aplicación en los servicios de salud mental de Uruguay. Fue necesario desarrollar el CCH porque para analizar los resultados de los tratamientos con la HoNOS, se requiere de datos complementarios que habitualmente se extraen de la historia clínica del usuario; pero los servicios de salud mental públicos de Uruguay no poseen un sistema de recolección de datos estandarizados que permita esta extracción.

Los expertos que participaron en el estudio piloto consideraron que ambos instrumentos (HoNOS y el CCH) eran adecuados y recogían en general aspectos relevantes. Sin embargo, propusieron algunas modificaciones (menores en el caso de la HoNOS), para su utilización en Uruguay; algunas de las cuales se incorporaron en la presente investigación. En el caso de la HoNOS la modificación implicó una adaptación lingüística de dos palabras en dos ítems del glosario. En el caso del CCH, resultó relevante la sugerencia de los expertos sobre incluir en la recolección de datos otras variables intervinientes que pudieran incidir en el análisis de los resultados de los efectos de los tratamientos. Sin embargo, en el presente estudio no fue posible analizar esos datos porque no presentaron la calidad suficiente. Para asegurar la calidad de estos datos, se requerirá que en estudios futuros se desarrolle un instructivo que acompañe al CCH.

Respecto a los resultados encontrados en el estudio de validación de la HoNOS, el análisis de las categorías de respuesta ha permitido observar que el sistema original de cinco categorías presenta deficiencias desde el punto de vista psicométrico, dado que los datos incumplen una de las reglas propuestas por Linacre (2002). Se observó un incremento no

monotónico de los pasos de calibración, lo que significa que los pasos de transición no están ordenados de forma creciente, mostrando un comportamiento irregular y desordenado. Este aspecto indica que algunas categorías no serán las más probables en ningún rango del continuo. Para mejorar el funcionamiento de las categorías de respuesta de la HoNOS se propuso una agregación con cuatro categorías y el reanálisis de los datos mostró que cumplieron todos los criterios de calidad propuestos por Linacre (2002).

Como en la casi totalidad de los modelos TRI, el Modelo de Escalas de Calificación asume que la probabilidad de las respuestas de las personas a los ítems son una función de la diferencia en una dimensión latente entre el nivel de la persona y el grado de severidad del ítem. Por tanto, uno de los supuestos básicos del modelo es la unidimensionalidad. Los datos de la HoNOS se adecuaron suficientemente a la dimensionalidad como para permitir el análisis mediante el modelo de Rasch.

En el contraste omnibus del ajuste entre las respuestas observadas y las esperadas a partir de los supuestos (los estadísticos Outfit e Infit), se ha observado que un porcentaje de personas menor al 10 % desajusta de manera severa, por lo que se ha de concluir que el ajuste es razonable; y ninguno de los ítems desajusta de forma severa. Se puede asumir, en consecuencia, que las respuestas de la HoNOS pueden explicarse fundamentalmente por una dimensión latente.

La fiabilidad de las medidas de los ítems cuantificada por el estadístico derivado del MEC (Item Separation Reliability) es elevada, y la de las personas se encuentra ligeramente por debajo del valor recomendado cuando es estimada con el estadístico Person Separation Reliability (.65), mejorando cuando es medida desde la Teoría Clásica de los Tests (alfa de

Cronbach= .69). Este resultado estuvo dentro del rango que reportaron la mayoría de los estudios previos sobre la HoNOS en diversos idiomas y países.

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. Este resultado fue explicado por el bajo nivel de severidad presentado por la mayoría de los pacientes de nuestra muestra. Dado que en la escala la mayor parte de los ítems son indicadores de trastornos graves, la precisión de las medidas tiende a ser menor en los rangos del continuo correspondientes a una menor gravedad de los trastornos mentales. Nuestros datos convergen con las evidencias de otros estudios en los que se observó que la precisión de la HoNOS es mayor para medir a los pacientes con trastornos mentales más graves.

En lo referido a la validez de criterio, se ha observado que los promedios de la HoNOS del grupo ambulatorios 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 de los trastornos mentales. Se requerirá que otras investigaciones en el futuro estudien la validez convergente de la versión al castellano de la HoNOS con otras medidas consolidadas de evaluación rutinaria de resultados de los servicios de salud mental.

En cuanto a la sensibilidad al cambio de la HoNOS, los resultados mostraron que la HoNOS resultó ser una medida capaz de detectar los cambios en las personas con trastornos mentales que reciben tratamiento en los servicios de salud mental. Por otra parte, el cambio observado fue significativamente mayor en los pacientes internados que en los pacientes ambulatorios.

El cuestionario empleado para analizar la viabilidad de la HoNOS como medida para evaluar los resultados en la práctica clínica, encontró que los clínicos valoraron positivamente a la HoNOS, aunque se reflejó un diverso grado de aceptación.

En suma, se encontraron algunos problemas métricos y algunas características deseables de la HoNOS como medida de ROA. Para afrontar los problemas métricos, una de las posibles soluciones encontradas en este estudio, ha sido agrupar las categorías de respuesta de la HoNOS de cinco a cuatro. Por tanto, se sugiere que la investigación futura en Uruguay utilice el sistema de cuatro categorías de respuesta de la HoNOS para lograr una mejor calidad de la medición.

También se destaca que la HoNOS presenta mejores propiedades cuando se utiliza en pacientes con trastornos mentales más severos.

Por último, a pesar de que más estudios se requieren para acumular evidencia sobre la validez, fiabilidad y sensibilidad al cambio de la versión al castellano de la HoNOS; con el análisis de sus características psicométricas en la población clínica uruguaya de adultos, esperamos haber contribuido a avanzar en la futura evaluación rutinaria de los resultados de los servicios de salud mental en Uruguay.

Referencias

- Abad, F. J., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Aitchison, R., Lutchman, R., Mellsop, G., Ruru, R., Savage, A. L., Tait, H., y Thompson, A. (2007). In search of a standardised, comprehensive assessment of functioning. *New Zealand Journal of Occupational Therapy*, 54(1), 33-38.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*. (4a ed.). Washington, DC: Autor.
- Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., y Schulz, H. (2007). A psychometric evaluation of the German version of the 'Health of the Nation Outcome Scales, HoNOS-D': on the Feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology*, 40, 116–125.
- Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., Kawski, S., Schulz, H. (2010). The validity of the German version of the Health of the Nation Outcome Scales (HoNOS-D): A clinician-rating for the differential assessment of the severity of mental disorders. *Journal of Methods in Psychiatric Research*, 19(1), 50–62. doi:10.1002/mpr.305
- Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., y 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:10.1159/000322690

- Andrews, G., y Page, A. C. (2005). Outcome measurement, outcome management and monitoring. *The Australian and New Zealand journal of psychiatry*, *39*, 649–651. doi:10.1111/j.1440-1614.2005.01648.x
- Andrews, G., Peters, L., y Teesson, M. (1994). *The measurement of consumer outcome in mental health: a report to the National Mental Health Information Strategy Committee*. Canberra: Australian Government Publishing Service.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 357–374.
- Andrich, D. (1988). *Rasch Models for measurement*. London: Sage Publications.
- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-Based Practice in Psychology, *American Psychologist*, *61*, 271–278.
- Audin, K., Margison, F. R., Clark, J. M., y Barkham, M. (2001). Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *The British journal of psychiatry: the journal of mental science*, *178*, 561–566. Recuperado de <http://bjp.rcpsych.org/content/178/6/561.full.pdf>
- Barry, A. (2011). *The search for a routine outcome measure for multidisciplinary interventions with young people in New Zealand: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Clinical Psychology at Massey University, Manawatu campus, New Zealand* (Tesis doctoral). Recuperado de http://muir.massey.ac.nz/bitstream/handle/10179/3926/02_whole.pdf?sequence=1
- Bebbington, P., Brugha, T., Hill, T., Marsden, L., y Window, S. (1999). Validation of the Health of the Nation Outcome Scales. *The British journal of psychiatry: the journal of mental science*, *174*, 389–394.

- Bech, P., Bille, J., Waarst, S., Wiese, M., Borberg, L., Treufeldt, P., y Kessing, L. (2006). Validity of HoNOS in identifying frequently hospitalized patients with ICD-10 mental disorders. *Acta psychiatrica Scandinavica*, *113*, 485–491. doi:10.1111/j.1600-0447.2005.00693.x
- Beurs, E. D., Rood, Y. R., Wee, N. J., Giltay, E. J., Noorden, M. S., Lem, R. V., y Fenema, E. V. (2011). Routine Outcome Monitoring in the Netherlands: Practical Experiences with a Web-Based Strategy for the Assessment of Treatment Outcome in Clinical Practice, *18*(1), 1–12. doi:10.1002/cpp.696
- Bond, T. G., y Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2a ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Boswell, J. F., Kraus, D. R., Miller, S. D., y Lambert, M. J. (2013). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*. Publicación anticipada en línea. doi:10.1080/10503307.2013.817696
- Browne, S., Doran, M., y McGauran, S. (2000). Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population. *Irish Journal of Psychological Medicine*, *17*(1), 17–19.
- Burgess, P., Coombs, T., Clarke, A., Dickson, R., y Pirkis, J. (2012). Achievements in mental health outcome measurement in Australia: Reflections on progress made by the Australian Mental Health Outcomes and Classification Network (AMHOCN). *International Journal of Mental Health Systems*, *6*(1), 4. doi:10.1186/1752-4458-6-4

- Burgess, P., Pirkis, J., y Coombs, T. (2006). Do adults in contact with Australia's public sector mental health services get better? *Australia and New Zealand Health Policy*, 3(9), 3–9.
- Burlingame, G., Dunn, T., y Chen, S. (2005). Selection of Outcome Assessment Instruments for Inpatients With Severe and Persistent Mental Illness. *Psychiatric Services*, 56, 444–451.
- Chou, Y., y Wang, W. (2010). Checking Dimensionality in Item Response Models With Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70, 717–731.
- Clark, M. L. (2002). Fichas de MedTrad: El concepto de outcome en el ámbito de la investigación. *Panacea@*, 3(7), 25–31. Recuperado de http://www.medtrad.org/panacea/PanaceaPDFs/Panacea7_Marzo2002.pdf
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2a ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coombs, T., y McKay, R. (2012). *Successful routine outcome assessment requires more than the right measure* [Comentario por pares en el artículo “An evidence-based approach to routine outcome assessment” de M. Slade]. Recuperado de <http://apt.rcpsych.org/content/18/3/180.short/reply>
- Coombs, T., Stapley, K., y Pirkis, J. (2011). The Multiple Uses of Routine Mental Health Outcome Measures in Australia and New Zealand: Experiences From the Field, *Australasia Psychiatry*, 19, 247–253. doi:10.3109/10398562.2011.562507

- Crawford, M. J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., Wykes, T., y Rose, D. (2011). Selecting outcome measures in mental health: the views of service users, *Journal of Mental Health*, 20, 336–346.
- Cyhlarova, E., McCulloch, A., McGuffin, P., y Wykes, T. (2010). *Economic burden of mental illness cannot be tackled without research investment*. London: Mental Health Foundation.
- Deady, M. (2009). *A review of screening, assessment and outcome measures for drug and alcohol settings*. Recuperado de <http://www.thehealthwell.info/node/252189>
- Delaffon, V., Anwar, Z., Noushad, F., Ahmed, A. S., y Brugha, T. S. (2012). Use of Health of the Nation Outcome Scales in psychiatry. *Advances in Psychiatric Treatment*, 18(3), 173–179. doi:10.1192/apt.bp.110.008029
- De la Parra, G. (2013). Psychotherapy research in developing countries: The case of Latin America. *Psychotherapy Research*, 23, 609–623. doi:10.1080/10503307.2013.830794
- De la Parra, G., y Von Bergen, A. (2000). El instrument OQ45.2 que mide resultados terapéuticos como herramienta de investigación psicoanalítica. *Revista Chilena de Psicoanálisis*, 17(2), 143–156.
- Del Castillo, R., Dogmanas, D., y Villar, M. (2011). *Informe de la Evaluación del Sistema de Salud Mental en Uruguay usando el World Health Organization/ Assessment Instrument of Mental Health Systems (WHO-AIMS)*. Documento inédito, Ministerio de Salud Pública, Montevideo, Uruguay.
- Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual-II*. Towson, MD: Clinical Psychometric Research.

- Dickens, G., Sugarman, P., y Walker, L. (2007). HoNOS-secure: a reliable outcome measure for users of secure and forensic mental health services. *Journal of Forensic Psychiatry and Psychology, 18*, 507–514.
- Digennaro Reed, F. D., y Reed, D. D. (2008). Towards an Understanding of Evidence-Based Practice. *Journal of Early and Intensive Behavior Intervention, 5*(2), 20–29.
Recuperado de <http://files.eric.ed.gov/fulltext/EJ829087.pdf>
- Donabedian, A. (1966). Evaluating the quality of medical care. *Millbank Memorial Fund Quarterly, 44*, 166–206.
- Donabedian, A. (1980). *Explorations in Quality Assessment and Monitoring. The Definition of Quality and Approaches to its Assessment*. Michigan: Health Administration.
- Duncan, E. A., y Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC health services research, 12*(1), 96. doi:10.1186/1472-6963-12-96
- Eisen, S. V., Dill, D. L., y Grob, M. C. (1994). Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. *Hospital and Community Psychiatry, 45*, 242–247.
- Elliott, R., Fox, C. M., Beltyukova, S. A., Stone, G. E., y Gunderson, J. (2006). Deconstructing Therapy Outcome Measurement With Rasch Analysis of a Measure of General Clinical Distress: The Symptom Checklist-90-Revised. *Psychological Assessment, 18*(4), 359–372. doi:10.1037/1040-3590.18.4.359
- Erbes, C., Polusny, M. A., Billig, J., Mylan, M., McGuire, K., Isenhardt, C., y Olson, D. (2004). Developing and applying a systematic process for evaluation of clinical outcome assessment instruments. *Psychological Services, 1*, 31–39.

- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., y Mellor-Clark, J. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry*, 180, 51–60.
doi:10.1192/bjp.180.1.51
- Fernández-Álvarez, H., Hirsh, H., Maristany, M., y Torrente, F. (2005). *Propiedades psicométricas del OQ-45.2 en la Argentina: un estudio piloto*. Póster presentado en el 4º Congreso Mundial de Psicoterapia, Buenos Aires.
- Fernández Castrillo, B., Gerpe, C., Montado, G., y Palermo, A. (2010). Investigación en psicoterapia hospitalaria. En B. Fernández Castrillo, C. Gerpe y L. Villalba (Comps.), *El Programa de Psicoterapia del Hospital de Clínicas*. Montevideo: UdelaR, CSIC, Facultad de Medicina.
- Ferrando, J., y Anguiano-Carrasco, C. (2010). Análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31, 18–33.
- Fitzpatrick, R., Davey, C., Buxton, M. J., y Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*, 2(14), 1–86. Recuperado de http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0003/64830/FullReport-ha2140.pdf
- Fleminger, S., Leigh, E., Eames, P., Langrell, L., Nagraj, R., y Logsdail, S. (2005). HoNOS-ABI: a reliable outcome measure of neuropsychiatric sequelae to brain injury? *Psychiatric Bulletin*, 29(2), 53–55.

-
- Flynn, K., Adams, K., y Alligood, E. (2002). *Outcome measurement in VHA Mental Health Services: An Overview and Series of Diagnosis Specific Short Reports. Part I* (Final Report). US: Technology Assessment Program.
- Fonagy, P. (2011). *Evidence from Adult MH – How Outcomes Have and Have Not Transformed Practice, Cambridge MHNN Meeting Talk*. Cambridge: Mental Health Research Network.
- Fonagy, P., Matthews, R., Pilling, S., y Glover, G. (2005). *Outcomes Measures Implementation Best Practice Guidance*. London: National Institute for Mental Health in England.
- Froyd, J. E., Lambert, M. J., y Froyd, J. D. (1996). A review of practices of psychotherapy outcome measurement. *Journal of Mental Health, 5*, 11–15.
- Gilbody, S. M., House, A. O., y Sheldon, T. A. (2002a). Outcome research in mental health. *British Journal of Psychiatry, 181*, 8–16.
- Gilbody, S. M., House, A. O., y Sheldon, T. A. (2002b). Psychiatrists in the UK do not use outcome measures: National Survey. *Psychiatric Bulletin, 180*, 101–103.
- Gilbody, S. M., House, A. O., y Sheldon, T. A. (2003). *Outcomes measurement in psychiatry: A critical review of outcomes measurement in psychiatric research and practice* (Report No. 24). Recuperado del sitio de internet de University of York, Centre for Reviews and Dissemination: <http://www.york.ac.uk/inst/crd/CRD-Reports/crdreport24.pdf>
- Hampson, M., Killaspy, H., Mynors-Wallis, L., y Meier, R. (2011). *Outcome Measures Recommended for Use in Adult Psychiatry* (Occasional Paper No. 78). London: Royal College of Psychiatrists.

-
- Hardiman, E., Theriot, M., y Hodges, J. (2005). Evidence-Based Practice in Mental Health: Implications and Challenges for Consumer-Run Programs. *Best Practices in Mental Health, 1*(1).
- Hatfield, D. R., y Ogles, B. M. (2006). The influence of outcome measures in assessing client change and treatment decisions. *Journal of Clinical Psychology, 62*, 325–338.
- Heinemann, A., Fisher, W., y Gershon, R. (2006). Improving Health Care Quality With Outcomes Management. *Proceedings, 6*, 46–50.
- Horowitz, L. M., Strupp, H., Lambert, M. J., y Elkin, I. (1997). Overview and summary of the Core Battery Conference. In H. H. Strupp, L. M. Horowitz, y M. J. Lambert (Eds.), *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery* (pp. 491–502). Washington, DC: American Psychological Association Press.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., y Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059–1064.
- Iglesias, C., Illescas, J., Lucio, R., Santamarina, S., y Alonso, M. (2005). Evaluación de las "necesidades asistenciales" en pacientes esquizofrénicos atendidos en un hospital de día. *Psiquiatria.com, 05*(4). Recuperado de <http://hdl.handle.net/10401/796>
- Jacobs, R. (2007). *Investigating patient outcome measures in mental health*. London: Office of Health Economics.
- Jacobs, R. (2009). *Investigating patient outcome measures in mental health*. (CHE Research Paper No. 48). York: Centre for Health Economics, University of York.

-
- Jacobs, R., y Moran, V. (2010). Uptake of mandatory outcome measures in mental health services. *The Psychiatrist*, *34*(8), 338–343.
- Jacobson, N. S., y Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. doi:10.1037/0022-006X.59.1.12
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, *63*, 146–159. doi:10.1037/0003-066X.63.3.146
- Kisely, S., Campbell, L. A., Cartwright, J., Cox, M., y Campbell, J. (2010). Do the Health of the Nation Outcome Scales measure outcome? *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *55*(7), 431–439.
- Kisely, S., Campbell, L. A., Crossman, D., Gleich, S., y Campbell, J. (2007). Are the Health of the Nation Outcome Scales a valid and practical instrument to measure outcomes in North America? A three-site evaluation across Nova Scotia. *Community mental health journal*, *43*(2), 91–107. doi:10.1007/s10597-006-9067-2
- Knaup, C., Koesters, M., Schoefer, D., Becker, T., y Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta- analysis. *British Journal of Psychiatry*, *195*, 15–22.
- Koch, N., Wasserfallen, J.-B., Camus, D., Gebel, S., y Bonsack, C. (2013). Implementing validated routine outcome assessment in psychiatry: a field study. *Swiss archives of neurology and psychiatry*, *164*(4), 134–144.

- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., Shimokawa, K., . . . Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire 45*. Salt Lake City, UT: OQ Measures L.L.C.
- Lane-Herman, L. (2012). *A Pilot Study Examining the Effect of an Intensive Skills-Based Training in Cognitive-Behavioral Therapy: Impact on Graduate Students' Competence* (Tesis doctoral). Recuperado de http://digitalcommons.pcom.edu/cgi/viewcontent.cgi?article=1206&context=psychology_dissertations
- Lauzon, S., Corbiere, M., Bonin, J. P., Bonsack, C., Lesage, A. D., y Ricard, N. (2001). Validation de la version française du Health of the Nation Outcome Scales (HoNOS-F). *Canadian Journal of Psychiatry*, 46, 841–846.
- Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3, 85–106.
- Linacre, J. M. (2012). *A user's guide to Winstep Ministep: Rasch model computer program*. Recuperado de <http://ifile.hkedcity.net/1/001/950/public/Secondary/EI0020070012/winsteps.pdf>
- Linacre, J. M., y Wright, B. D. (2000). *WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch análisis (Computer software)*. Chicago: MESA Press.
- Londoño, C., y Valencia, S. C. (2005). La investigación en el proceso de soporte empírico de las terapias: Alcances y limitaciones. *Acta Colombiana de Psicología*, 13, 163–181.
- Lora, A., Bezzi, R., y Erlicher, A. (2007). Estimating the prevalence of severe mental illness in mental health services in Lombardy (Italy). *Community mental health journal*, 43(4), 341–357. doi:10.1007/s10597-006-9078-z

- Lorenzo-Seva, U., y Ferrando, P. (2006). *Manual of the program FACTOR*. Tarragona: Universitat Rovira i Virgili. Recuperado de <http://psico.fcep.urv.es/utilitats/factor/>
- Lovaglio, G., y Monzani, E. (2011a). Health of the nation outcome scales evaluation in a community setting population. *Quality of life research*, 21, 1643–1653. doi:10.1007/s11136-011-0071-9
- Lovaglio, G., y Monzani, E. (2011b). Validation aspects of the health of the nation outcome scales. *International journal of mental health systems*, 5(1), 20. doi:10.1186/1752-4458-5-20
- Lutchman, R., Thompson, A., Tait, H., Savage, A., Aitchison, R., Ruru, R., y Mellsop G., (2007). In search of a standardised, comprehensive assessment of functioning. *New Zealand Journal of Occupational Therapy*, 54(1), 33–38.
- Maloney, K., y Chaiken, B. P. (1999). An Overview of Outcomes Research and Measurement. *Journal of Healthcare Quality*, 21(6), 4–9. Recuperado de <http://www.docsnetwork.com/articles/BPC99115.pdf>
- Matthies, M. C., Uriarte, J. J., Martínez, M. J., Rodríguez, A., y López, M. (2012). *Atención a las personas con enfermedad mental grave en los centros de SMB*. Bizkaia: Red de Salud Mental de Bizkaia. Recuperado de http://www.osakidetza.euskadi.net/r85-pkdocc08/eu/contenidos/informacion/premios_osakidetza/eu_osk/adjuntos/04/217.pdf
- Mavranouzouli, I., Brazier, J. E., Young, T. A., y Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of Life Research*, 20, 321–333.
- McClelland, R. (2000). Validation of an outcome scale for use in adult psychiatric practice.

Quality in Health Care, 9(2), 98–105. doi:10.1136/qhc.9.2.98

McPheeters, H. L. (1984). Statewide mental health outcome evaluation: A perspective of two southern states. *Community Mental Health Journal*, 20, 44–55.

Meehan, T., McCombes, S., Hatzipetrou, L., y Catchpoole, R. (2006). Introduction of routine outcome measures: staff reactions and issues for consideration. *Journal of Psychiatric and Mental Health Nursing*, 13, 581–587.

Mellsop, G., y Wilson, J. (2006). Outcome measures in mental health services: Humpty Dumpty is alive and well. *Australasian Psychiatry* 14(2), 137–140.
doi:10.1111/j.1440-1665.2006.02270.x

Milne, D., Reichelt, K., y Wood, E. (2001). Implementing HoNOS: An eight stage approach. *Clinical Psychology y Psychotherapy*, 8(2), 106–116. doi:10.1002/cpp.252

Ministerio de Salud Pública. (2005). Salud Mental en la emergencia social y en el nuevo modelo asistencial. Documento programático de la Comisión Asesora Técnica Permanente del Programa Nacional de Salud Mental. *Revista de Psiquiatría del Uruguay*, 69(2), 155–160.

Ministerio de Salud Pública. (2011). *Plan de implementación de prestaciones en salud mental en el Sistema Nacional Integrado de Salud*. Recuperado de <http://www.msp.gub.uy/sites/default/files/Plan%20de%20Prestaciones%20en%20Salud%20Mental.pdf>

Ministerio de Salud Pública. (2013). *Avances del Plan de implementación de prestaciones en salud mental en el Sistema Nacional Integrado de Salud*. Conferencia por el Día Nacional de Salud Mental, Montevideo.

- Mulder, C. L., Staring, A. B. P., Loos, J., Buwalda, V. J. A., Kuijpers, D., Sytema, S., y Wierdsma, A. I. (2004). De Health of the Nations Outcome Scales (HoNOS) als instrument voor 'routine outcome assessment'. *Tijdschrift Voor Psychiatrie*, 46, 273–285.
- Montado, G., Palermo, A., Fernández, B., y Gerpe, C. (2011). Investigación empírica: Puente entre psicoterapia y psiquiatría. *Revista de Psiquiatría del Uruguay*, 75(1), 48–55.
- Najson, S. (2008). *Validación del Symptom Checklist-90-R para la población uruguaya*. Documento inédito, Universidad de la República, Montevideo, Uruguay.
- National Institute of Mental Health. (2004). *Organising and Delivering Psychological Therapies, Department of Health UK*. Recuperado de http://webarchive.nationalarchives.gov.uk/20040823180324/dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4086097.pdf
- National Institute for Mental Health. (2008). *Mental health Outcomes Compendium*. Bimingham: Autor.
- Newnham, E. A., y Page, A. C. (2010). Bridging the gap between best evidence and best practice in mental health. *Clinical Psychology Review*, 30, 127–142.
- Organización Mundial de la Salud. (1992). *CIE-10: Trastornos mentales y del comportamiento*. Madrid: Meditor.
- Organización Mundial de la Salud. (2013). *Plan Mundial en Salud Mental 2013–2020*. Ginebra; Autor. Recuperado de http://apps.who.int/iris/bitstream/10665/97488/1/9789243506029_spa.pdf

- Organización Panamericana de la Salud. (2009). *Estrategia y plan de acción sobre salud mental*. Washington, DC: Autor. Recuperado de http://new.paho.org/hq/dmdocuments/2009/SALUD_MENTAL_final_web.pdf
- Organización Panamericana de la Salud. (2011). *Marco de Referencia para la Implementación de la Estrategia Regional en Salud Mental*. Washington, DC: Autor. Recuperado de <http://www.paho.org/hq/dmdocuments/MARCO%20DE%20REFERENCIA%20Salud%20Mental%20final.pdf?ua=1>
- Parabiaghi, A., Barbato, A., D'Avanzo, B., Erlicher, A., y Lora, A. (2005). Assessing reliable and clinically significant change on Health of the Nation Outcome Scales: method for displaying longitudinal data. *The Australian and New Zealand journal of psychiatry*, 39, 719–725. doi:10.1111/j.1440-1614.2005.01656.x
- Pastor, D. y Beretvas, S. (2006). Longitudinal Rasch Modeling in the Context of Psychotherapy Outcomes Assessment. *Applied Psychological Measurement*, 30(2), 100–120. doi:10.1177/0146621605279761
- Phuaphanprasert, B., Srisurapanont, M., y Silpakit, C. (2007). Reliability and validity of the Thai version of the Health of the Nation Outcome Scales (HoNOS). *Journal of the Medical Association of Thailand*, 90, 2487–2493.
- Pirkis, J., Burgess, P., Coombs, T., Clarke, A., David Jones, E., y Dickson, R. (2005). Routine measurement of outcomes in Australian public sector mental health services. *Australia and New Zealand Health Policy*, 2, 8.
- Pirkis, J. E., Burgess, P. M., Kirk, P. K., Dodson, S., Coombs, T. J., y 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:10.1186/1477-7525-3-76
- Prasad-Ildes, R., y Wright, B. (2004). One Size Does Not Fit All: the case of standardised Mental Health Outcome Measures in multicultural Australia. *Synergy - Multicultural Mental Health Australia*, 1, 6–7.
- Prieto, G., y Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94–100.
- Prieto, G., y Velasco, A. (2003). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en Psicología*, 19(106), 5–23.
- Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., y Phillips, M. R. (2007). No health without mental health. *Lancet*, 370, 859–877. doi:10.1016/S0140-6736(07)61238-0
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*.
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, 207–230.
- Rosen, A., Hadzi-Pavlovic, D., y Parker, G. (1989). The Life Skills Profile: A measure assessing function and disability in schizophrenia. *Schizophrenia Bulletin*, 15, 325–337.
- Roth, A., y Fonagy, P. (2005). *What Works for Whom?: A Critical Review of Psychotherapy Research* (2a ed.). New York: Guilford.
- Roy, A., Matthews, H., Clifford, P., Fowler, V., y Martin, D. (2002). Health of the Nation Outcome Scales for People with Learning Disabilities (HoNOS-LD). *British Journal of Psychiatry*, 180, 61–66.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., y Haynes, R. B. (2000).

Evidence-based medicine: How to practice and teach EBM (2a ed.). New York: Churchill Livingstone.

Saldivia, S., Grandón, P., y Vicente, B. (2012). Investigación en servicios de salud mental en Chile: avances y desafíos. *Medwave*, 12(10), 5549.

doi:10.5867/medwave.2012.10.5549

Schmidt, L. J., Garratt, A. M., y Fitzpatrick, R. (2000). *Instruments for Mental Health: a Review* (Report to the Department of Health, September). Oxford: National Centre for Health Outcomes Development. Recuperado de

http://phi.uhce.ox.ac.uk/pdf/phig_mental_health_report.pdf

Seligman, M. E. (1995). The effectiveness of psychotherapy: The Consumer Reports Study. *American Psychologist*, 50, 965–974.

Slade, M. (2002a). What outcomes to measure in routine mental health services and how to assess them: a systematic review. *Australian and New Zealand Journal of Psychiatry*, 36, 743–753.

Slade, M. (2002b). Routine outcome assessment in mental health services. *Psychological Medicine*, 32, 1339–1343. doi:10.1017/S0033291701004974

Slade, M. (2004). *The development and evaluation of an evidence-based approach to implementing outcome measures in routine mental health services* (Tesis doctoral). Institute of Psychiatry, King's College London.

Slade, M. (2006). Commissioning Outcome-focused Mental Health Services. *Mental Health Review Journal*, 11(3), 31–36.

-
- Slade, M. (2012). An evidence-based approach to routine outcome assessment. *Advances in Psychiatric Treatment*, 18, 180–182.
- Slade, M., McCrone, P., Kupiers, E., Leese, M., Cahill, S., Parabiaghi, A., Priebe, S., y Thornicroft, G. (2006). Use of standardised outcome measures in adult mental health services: Randomised controlled trial. *British Journal of Psychiatry*, 189, 330–336.
- Slade, M., Powell, R., y Rosen, A. (2000). Threshold Assessment Grid (TAG): the development of a valid and brief scale to assess the severity of mental illness. *Social Psychiatry and Psychiatric Epidemiology*, 35, 78–85.
- Slade, M., Thornicroft, G., y Glover, G. S. (1999). The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology*, 34, 243–249.
- Smith, G. R., Fischer, E. P., Nordquist, C., Mosley, C., y Ledbetter, N. S. (1997). Implementing outcomes management systems in mental health settings. *Psychiatric Services*, 48, 364–368.
- Smith, R. M., y Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. En M. Wilson (Ed.), *Objective Measurement: Theory into Practice. Volume 2*. Greenwich: Ablex.
- Speak, B. L., y Hay, P. (2012). *The utility of the Mental health Clustering Tool (MHCT) as a generic outcome measure for the payment by results quality and outcomes framework in working age and older adult mental health services*. (Final Report). Recuperado de <http://www.cppconsortium.nhs.uk/admin/files/The%20utility%20of%20the%20MHC%20as%20a%20generic%20outcome%20measure%20for%20the%20PbR%20qualit>

y%20and%20outcomes%20framework%20for%20working%20age%20and%20older%20adult%20mental%20health%20V2.pdf

Sperlinger, D. (2002). *Outcome Assessment in Routine Clinical Practice in psychosocial services, British Psychological Society's Centre for Outcomes, Research and Effectiveness*. London: The British Psychological Society. Recuperado de <http://www.psige.org/public/files/BPS%20Measuring%20Outcomes%20in%20Routine%20Clinical%20Practice%20Paper%201.pdf>

Stedman, T., Yellowlees, P., Mellsoy, G., Clarke, R., y Drake, S. (1997). *Measuring Consumer Outcomes In Mental Health: Field Testing of Selected Measures of Consumer Outcome in Mental Health*. St Lucia, Queensland: University of Queensland. Recuperado de [http://www.health.gov.au/internet/main/publishing.nsf/Content/55CDF7AE496207BECA257BF0001F97FD/\\$File/consfield.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/55CDF7AE496207BECA257BF0001F97FD/$File/consfield.pdf)

Stewart, A. L., Hays, R. D., y Ware, J. E. (1988). The MOS Short-Form General Health Survey: reliability and validity in a patient population. *Medical Care*, 26, 724–732.

Thomas, M. L., y Thomas, M. L. (2011). The Value of Item Response Theory in Clinical Assessment: A Review. *Assessment*, 18, 291–307. doi:10.1177/1073191110374797

Thornicroft, G., y Slade, M. (2000). Are routine outcome measures feasible in mental health? *Quality in Health Care*, 9, 84. doi:10.1136/qhc.9.2.84

Trauer, T. (2010a). Introduction. En T. Trauer (Ed.). *Outcome measurement in mental health: theory and practice* (pp. 1–11). Cambridge: Cambridge University Press.

Trauer, T. (2010b). Outcome measurement in chronic mental illness. *International Review of Psychiatry*, 22(2), 99–113. doi:10.3109/09540261003667525

- Trauer, T., y Buckingham, B. (2006). *National Outcomes and Casemix Collection Discussion Paper - The Health of the Nation Outcome Scales (HoNOS): Towards an Agenda for Future Development*. Canberra: Adult Mental Health Outcomes Expert Group.
- Trauer, T., Callaly, T., Hantz, P., Little, J., Shields, R., y Smith, J. (1999). Health of the Nation Outcome Scales. Results of the Victorian field trial. *British Journal of Psychiatry*, *174*, 380–388.
- 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:10.1080/13607860410001725063
- Uriarte, J. J., Beramendi, V., Medrano, J., Wing, J. K., Beevor, A. S., y Curtis, R. H. (1999). Presentación de la traducción al castellano de la escala HoNOS. *Psiquiatría Pública*, *11*, 93–101.
- Veit, C. T., y Ware, J. E., Jr. (1983). The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology*, *51*, 730–742. doi:10.1037/0022-006X.51.5.730
- Ware, J. E., y Sherbourne, C. D. (1992). The MOS 36-item short form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*, 473–481. doi:10.1097/00005650-199206000-00002
- 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.

-
- Wing, J. K., Curtis, R. H., y Beevor, A. S. (1996). *HoNOS: Health of the Nation Outcome Scales* (Report on Research and Development, July 1993 - December 1995). London: Royal College of Psychiatrists.
- Wolfe, E. W., y Smith, E. V. (2007a). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part I-Instrument Development Tools. *Journal of Applied Measurement*, 8, 97–123.
- Wolfe, E. W., y Smith, E. V. (2007b). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part II-Validation Activities. *Journal of Applied Measurement*, 8, 204–234.
- Wright, B. D., y Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Zickar, M. J., y Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. En C. E. Lance, y R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 37–61). New York: Routledge.
- Zytner, R. (2011). *Informe de lo actuado de enero a diciembre en el Servicio de Atención Psicológica Preventivo-Asistencial de la Facultad de Psicología*. Documento inédito, Universidad de la República, Montevideo, Uruguay.

APÉNDICES

ESCALAS DE EVOLUCIÓN HoNOS: VERSIÓN EN LENGUA CASTELLANA

**Escalas de Evolución
HoNOS**

Uriarte JJ, Beramendi V*, Medrano J**
Wing JK***, Beevor AS***, Curtis R****

Versión en Lengua Castellana de:**HEALTH OF THE NATION OUTCOME SCALES
(Wing, Beevor, Curtis)**

*Hospital Psiquiátrico de Zamudio. Sº Rehabilitación. Osakidetza-Servicio Vasco de Salud

**Hospital Psiquiátrico Las Nieves. Osakidetza-Servicio Vasco de Salud

***Royal College of Psychiatrists. College Research Unit

Resumen de las Instrucciones de Puntuación

1. **Puntuar** cada una de las escalas siguiendo el orden de la 1 a la 12
2. **No** incluir información ya puntuada en un ítem previo excepto en el ítem 10 que es una evaluación global
3. **Puntuar** el problema MÁS SEVERO ocurrido durante el periodo de tiempo evaluado
4. **Todas las escalas** siguen el formato:

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

ESCALAS DE EVOLUCIÓN HoNOS: VERSIÓN EN LENGUA CASTELLANA

Glosario de la Hoja de Puntuación HoNOS**1. Conducta hiperactiva, agresiva, disruptiva o agitada**

Incluye tal comportamiento debido a cualquier causa (por ejemplo drogas, alcohol, demencia, psicosis, depresión, etc.)

No incluye la conducta bizarra, que se puntúa en la escala 6

0. Sin problema de este tipo durante el periodo evaluado
1. Irritabilidad, enfrentamientos, inquietud etc. que no requiere intervención
2. Incluye actitudes agresivas; empujar o importunar a otras personas; amenazas o agresiones verbales; daños menores a bienes y enseres (por ejemplo romper vasos, cristales); hiperactividad marcada o agitación
3. Agresividad física hacia otras personas o animales (insuficiente para puntuar 4); actitud amenazante; hiperactividad más grave o destrucción de bienes o enseres
4. Al menos un ataque físico serio a otras personas o animales; destrucción de propiedades (por ejemplo provocando incendios); conducta intimidatoria u obscena

2. Autolesiones no accidentales

*No incluye las autolesiones accidentales (debidas, por ejemplo, a demencia o a incapacidad intelectual severa); **los problemas cognitivos han de ser puntuados en la Escala 4 y el daño resultante en la Escala 5***

No incluye las enfermedades o lesiones producidas como consecuencia directa del consumo de alcohol o drogas, que se puntúan en la Escala 3; (por ejemplo la cirrosis hepática o las lesiones derivadas de conducir en estado ebrio se puntúan en la escala 5).

0. Ningún problema de este tipo durante el periodo evaluado
1. Pensamientos pasajeros de acabar con todo pero riesgo pequeño durante el periodo evaluado; no autolesiones
2. Riesgo leve durante el periodo evaluado; incluye autolesiones no peligrosas (por ejemplo arañarse las muñecas)
3. Riesgo moderado o grave de autoagresión deliberada durante el periodo evaluado; incluye actos preparatorios (por ejemplo acumular medicación)
4. Tentativa seria de suicidio y/o autolesión grave deliberada durante el periodo evaluado

3. Consumo Problemático de Alcohol o Drogas

No incluye la conducta agresiva/destructiva secundaria al consumo de alcohol o drogas, puntuado en la Escala 1

No incluye a las enfermedades o discapacidades derivadas del consumo de alcohol o drogas, puntuadas en la Escala 5

ESCALAS DE EVOLUCIÓN HONOS: VERSIÓN EN LENGUA CASTELLANA

0. Ningún problema de este tipo durante el periodo evaluado
1. Consumo excesivo pero dentro de las normas sociales
2. Pérdida de control en el consumo de alcohol o drogas pero sin dependencia seria
3. Marcada apatencia o dependencia del alcohol o drogas con frecuente pérdida del control; conductas de riesgo durante los periodos de intoxicación
4. Incapacitado por los problemas de alcohol/drogas

4. Problemas Cognitivos

Incluye problemas de memoria, orientación y comprensión asociadas a cualquier trastorno: incapacidad intelectual, demencia, esquizofrenia, etc.

No incluir trastornos temporales (por ejemplo resacas) secundarios al consumo de alcohol o drogas, puntuados en la Escala 3

0. Ningún problema de este tipo durante el periodo evaluado
1. Problemas menores de memoria o comprensión (por ejemplo olvido ocasional de nombres)
2. Problemas leves pero claros (por ejemplo se ha perdido en un lugar familiar o no ha reconocido a una persona conocida); confusión ocasional en relación con decisiones simples
3. Marcada desorientación en el tiempo, el espacio o las personas; desconcierto ante acontecimientos cotidianos; lenguaje algunas veces incoherente; enlentecimiento mental
4. Desorientación severa (por ejemplo incapacidad para reconocer a familiares); riesgo de accidentes; habla incomprensible; obnubilación o estupor

5. Problemas por enfermedad física o discapacidad

Incluye enfermedades o discapacidades de cualquier causa que limiten o impidan el movimiento, o deterioren la capacidad visual o auditiva, o que de otra manera interfieran con la autonomía personal

Incluye a los efectos secundarios de la medicación; los efectos del consumo de drogas/alcohol; las discapacidades físicas derivadas de accidentes o de autolesiones asociadas a problemas cognitivos, conducción de vehículos bajo los efectos del alcohol, etc.

No incluye los problemas mentales o conductuales puntuados en la Escala 4

0. Ningún problema de salud física durante el periodo evaluado
1. Problema menor de salud durante dicho periodo (por ejemplo resfriados, caídas de poca importancia, etc.)
2. Problema de salud física que provoca limitación leve de la movilidad y actividad
3. Grado moderado de limitación de la actividad debido a un problema de salud física
4. Incapacidad severa o completa secundaria a un problema de salud física

ESCALAS DE EVOLUCIÓN HONOS: VERSIÓN EN LENGUA CASTELLANA

6. Problemas asociados a la presencia de ideas delirantes y alucinaciones

Incluye alucinaciones e ideas delirantes independientemente del diagnóstico

Incluye la conducta extraña o bizarra asociada a la presencia de alucinaciones o ideas delirantes

No incluye la conducta agresiva, destructiva o hiperactiva atribuible a la presencia de alucinaciones o ideas delirantes, que se puntúa en la Escala 1

0. No hay evidencia de alucinaciones, ideas delirantes durante el periodo evaluado
1. Creencias excéntricas o en alguna medida extrañas no congruentes con el ambiente cultural del paciente
2. Ideas delirantes o alucinaciones (por ejemplo voces o visiones) presentes, pero originan escaso malestar o conducta bizarra en el paciente, es decir clínicamente presente pero de forma leve
3. Preocupación marcada en relación con las ideas delirantes o las alucinaciones, que originan mucho malestar y/o se manifiestan en una conducta bizarra obvia, es decir problema de moderada severidad clínica
4. El estado mental y la conducta del paciente están seria y negativamente afectados como consecuencia de las ideas delirantes o de las alucinaciones, con una repercusión severa sobre el paciente

7. Problemas en relación con el humor depresivo

No incluye la hiperactividad o la agitación, que se puntúa en la Escala 1

No incluye la ideación o tentativas de suicidio, que se puntúan en la Escala 2

No incluye la presencia de ideas delirantes o alucinaciones, que se puntúan en la Escala 6

0. Ningún problema asociado con humor depresivo durante el periodo evaluado
1. Tristeza; o cambios menores en el estado de ánimo
2. Depresión y malestar leve pero claro (por ejemplo sentimientos de culpa; pérdida de la autoestima)
3. Depresión con autoinculpación injustificada; preocupación acerca de sentimientos de culpa
4. Depresión severa o muy severa, con ideas de culpa o autoacusación

8. Otros problemas mentales o conductuales

Puntuar únicamente el problema clínico más severo que no haya sido considerado en los ítems 6 y 7 como sigue. Especificar el tipo de problema consignando la letra apropiada: A fóbico; B ansiedad; C obsesivo-compulsivo; D sobrecarga mental y tensión; E disociativo; F somatoforme; G alimentación; H sueño; I sexual; J otros, especificar

0. Sin evidencia de cualquiera de estos problemas a lo largo del periodo evaluado
1. Problemas menores solamente

ESCALAS DE EVOLUCIÓN HONOS: VERSIÓN EN LENGUA CASTELLANA

2. Problema clínicamente presente con intensidad leve (por ejemplo el paciente tiene un cierto grado de control)
3. Crisis o malestar severo de forma ocasional, con pérdida de control (por ejemplo se ve obligado a evitar por completo situaciones que generan ansiedad, pedir ayuda a un vecino, etc.) es decir nivel de problema moderadamente severo
4. Problema grave que domina la mayoría de las actividades

9. Problemas con las relaciones

Puntuar el problema más severo del paciente asociado a la carencia activa o pasiva de relaciones sociales, y/o a relaciones carentes de apoyo, destructivas o autolesivas

0. Ningún problema significativo durante el periodo evaluado
1. Problema menor sin entidad clínica
2. Problema claro para crear o mantener relaciones de apoyo: el paciente se queja de ello y/o tales problemas son evidentes para los demás
3. Problema severo y persistente debido a carencia pasiva o activa de relaciones sociales y/o que las relaciones existentes aportan un escaso o nulo apoyo o consuelo
4. Aislamiento social grave y penoso debido a incapacidad para comunicarse socialmente y/o a la pérdida de las relaciones sociales

10. Problemas con las actividades de la vida cotidiana

Puntuar el nivel medio de funcionamiento en las actividades de la vida diaria (AVD): por ejemplo problemas con actividades básicas de cuidados personales como la alimentación, la limpieza personal, la vestimenta, el uso del lavabo; también habilidades complejas como la administración del dinero, la organización de donde vivir, la ocupación y el tiempo libre, la movilidad y el uso del transporte, las compras, el desarrollo personal, etc.

Incluye: cualquier falta de motivación para manejarse o valerse por sí mismo, puesto que esto contribuye a un menor nivel de funcionamiento

No incluye la falta de oportunidades para ejercer las habilidades y capacidades intactas, que se registra en las Escalas 11 y 12

0. Sin problemas durante el periodo evaluado; buena capacidad de funcionamiento en todas las áreas
1. Problemas menores solamente (por ejemplo desordenado, desorganizado)
2. Cuidados personales adecuados pero fallo importante en la realización de una o más habilidades complejas (ver arriba)
3. Problemas severos en una o más áreas de los cuidados personales (alimentación, limpieza personal, vestimenta, uso del lavabo) así como discapacidad severa para ejecutar varias habilidades complejas
4. Discapacidad grave o incapacidad en todas o casi todas las áreas de cuidados personales y habilidades complejas

ESCALAS DE EVOLUCIÓN HoNOS: VERSIÓN EN LENGUA CASTELLANA

11. Problemas con las condiciones de vida

Puntuar la severidad media de los problemas en relación con la calidad de las condiciones de vida y con la rutina doméstica cotidiana

¿Están cubiertas las necesidades básicas (calefacción, luz, higiene)? Si es así, ¿se dispone de ayuda para afrontar las discapacidades y de variedad de oportunidades para utilizar habilidades y desarrollar otras nuevas?

No puntuar el nivel de discapacidad funcional en sí mismo, puntuado en la Escala 10

Nota : puntuar el alojamiento habitual del paciente. Si la evaluación se lleva a cabo en una unidad de internamiento de corta estancia, puntuar su alojamiento habitual.

0. El alojamiento y las condiciones de vida son aceptables; útil para mantener cualquier discapacidad registrada en la Escala 10 en el más bajo nivel posible, y para ayudar al paciente a valerse por sí mismo
1. El alojamiento es razonablemente aceptable aunque existen problemas menores o transitorios (por ejemplo la situación no es la ideal, no es la opción preferida, la comida no es de su agrado, etc.)
2. Problema significativo con uno o más aspectos del alojamiento y/o régimen (por ejemplo elección restringida; el personal o las personas que viven con el paciente tienen conocimientos insuficientes acerca de como limitar la discapacidad o de como ayudar a utilizar o a desarrollar habilidades nuevas o intactas)
3. Múltiples problemas preocupantes en relación con el alojamiento (por ejemplo carencia de algunas necesidades básicas); los recursos del entorno doméstico para mejorar la independencia del paciente son mínimos o inexistentes
4. El alojamiento es inaceptable (por ejemplo carece de las necesidades básicas, el paciente está en riesgo de desahucio o de quedarse en la calle, o las condiciones de vida son por lo demás intolerables), agravando los problemas del paciente

12. Problemas en relación con la ocupación y las actividades

Puntuar el nivel global de problemas en relación con la calidad del entorno cotidiano.

¿Existe ayuda disponible para afrontar las discapacidades, y oportunidades para mantener o mejorar capacidades ocupacionales y recreativas y actividades? Considerar factores tales como el estigma, falta de personal cualificado, acceso a recursos de apoyo (por ejemplo personal y equipamiento de centros de día, talleres, clubes sociales, etc.)

No puntuar el nivel de incapacidad funcional en sí mismo, puntuado en la Escala 10

Nota : Puntuar la situación habitual del paciente. Si el paciente está internado en una unidad de corta estancia, puntuar las actividades durante el periodo previo al ingreso.

0. El entorno cotidiano del paciente es aceptable: útil para mantener cualquier discapacidad puntuada en la Escala 10 en el menor nivel posible, y para ayudar al paciente a valerse por sí mismo.

ESCALAS DE EVOLUCIÓN HONOS: VERSIÓN EN LENGUA CASTELLANA

1. Problema menor o transitorio (por ejemplo retraso en cobrar el paro): recursos razonables disponibles pero no siempre en el momento deseado, etc.
2. Surtido limitado de actividades; carencia de tolerancia razonable (por ejemplo denegar injustificadamente la entrada en bibliotecas u otros lugares públicos, etc.); en desventaja por la falta de domicilio estable; apoyo insuficiente por parte de profesionales o cuidadores; ayuda durante el día disponible pero durante tiempo muy limitado
3. Deficiencia marcada de servicios cualificados disponibles que ayuden a minimizar el nivel de discapacidad existente; ausencia de oportunidades para utilizar habilidades intactas o para adquirir nuevas; atención no cualificada de difícil acceso
4. La falta de oportunidades para realizar actividades durante el día contribuye a empeorar los problemas del paciente

ESCALAS DE EVOLUCIÓN HoNOS: VERSIÓN EN LENGUA CASTELLANA

HoNOS
Hoja De Puntuación

Nombre :

Fecha :

- | | | |
|---|-----------|--------------------------|
| 1. Conducta hiperactiva, agresiva, disruptiva o agitada | 0 1 2 3 4 | <input type="checkbox"/> |
| 2. Autolesiones no accidentales | 0 1 2 3 4 | <input type="checkbox"/> |
| 3. Consumo Problemático de Alcohol o Drogas | 0 1 2 3 4 | <input type="checkbox"/> |
| 4. Problemas cognitivos | 0 1 2 3 4 | <input type="checkbox"/> |
| 5. Problemas por enfermedad física o discapacidad | 0 1 2 3 4 | <input type="checkbox"/> |
| 6. Problemas asociados a la presencia de ideas delirantes y alucinaciones | 0 1 2 3 4 | <input type="checkbox"/> |
| 7. Problemas en relación con el humor depresivo | 0 1 2 3 4 | <input type="checkbox"/> |
| 8. Otros problemas mentales o conductuales | 0 1 2 3 4 | <input type="checkbox"/> |
| Especificar el tipo de trastorno: A, B, C, D, E, F, G, H, I, J | | <input type="checkbox"/> |
| 9. Problemas con las relaciones | 0 1 2 3 4 | <input type="checkbox"/> |
| 10. Problemas en relación con las actividades de la vida cotidiana | 0 1 2 3 4 | <input type="checkbox"/> |
| 11. Problemas con las condiciones de vida | 0 1 2 3 4 | <input type="checkbox"/> |
| 12. Problemas en relación con la ocupación y las actividades | 0 1 2 3 4 | <input type="checkbox"/> |

Puntuación Total :

0 - 48

FICHA DE INFORMACIÓN COMPLEMENTARIA A LA ESCALA HoNOS

Fecha de hoy: ___/___/_____**Fecha de la consulta:** ___/___/_____**DATOS DEL EVALUADOR****Identificación** _____**Profesión del evaluador:** psicólogo (1) _____ **psiquiatra** (2) _____**Experiencia profesional en el área:** Años (1) _____ Meses (2) _____**Ámbito de aplicación de la Escala:**

Policlínica (1) _____ Centro Diurno (2) _____ Internación H. psiquiátrico (3) _____

Experiencia previa en utilización de escalas de evaluación:

Ninguna (1) _____ Alguna (2) _____ Mucha (3) _____

Primera consulta del profesional con este paciente Si (1) _____ No (2) _____

En caso de responder No (2), seleccione y especifique el tiempo que lleva asistiendo al paciente.

días ___ ___ meses ___ ___ años ___ ___

DATOS DEL PACIENTE**Identificación** _____**Sexo (F/M)** _____**Edad** _____**Ciudad de residencia** _____**Nivel educativo alcanzado:** Primaria (1) _____ Secundaria (2) _____ UTU (3) _____

Universitario (4) _____

En caso de Universitario: Completo (1) _____ Incompleto (2) _____

Trabaja actualmente Si (1) _____ No (2) _____**Estado civil:** soltero/a (1) _____ casado/a (2) _____ separado/divorciado/viudo (3) _____

Unión libre (4) _____

FICHA DE INFORMACIÓN COMPLEMENTARIA A LA ESCALA HoNOS

Diagnóstico primario según CIE-10, indique con una cruz el grupo

1. Trastornos mentales y de conducta debidos a uso de sustancias psicoactivas (F10-F19) ____
2. Esquizofrenia, trastornos esquizotípicos y delirantes (F20-F29) ____
3. Trastornos del estado del ánimo [afectivos] (F30-F39) ____
4. Trastornos neuróticos, relacionados con el estrés y somatoformos (F40-F48) ____
5. Trastornos de personalidad y conducta de adultos (F60-F69) ____
6. Otros (ej. epilepsia, trastornos mentales orgánicos, retardo mental, trastornos conductuales y emocionales cuya primera aparición generalmente ocurre en la infancia o adolescencia, trastornos de evolución psicológica) ____ Especifique _____

Inicio del tratamiento actual (seleccione y especifique)

días ____ meses ____ años _____

Antecedentes: fecha del primer tratamiento recibido (indique la fecha o en caso de no poseer la información, indique el año):

Fecha _____ Año _____

Comentarios: _____

CUESTIONARIO PARA LA ADAPTACIÓN DE LA HONOS EN URUGUAY

Nombre del profesional _____

1) Valore las instrucciones sobre cómo completar la Escala en cuanto a:

Claridad en su redacción : muy poca (1)____ poca(2)____ regular(3)____
bastante(4)____ total(5)____

¿Realizaría alguna modificación a las instrucciones? Si____ No____
En caso afirmativo especifique

2) En cuanto al contenido de los ítems de la escala

2.a) Valore la relevancia de cada ítem para los pacientes con trastornos mentales severos de nuestra población

Ítem 1: muy poco (1)____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

Ítem 2: muy poco (1) ____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

Ítem 3:muy poco (1) ____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

Ítem 4: muy poco (1) ____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

Ítem 5: muy poco (1) ____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

Ítem 6: muy poco (1) ____ poco(2)____ regular(3)____ bastante(4)____ total(5)____

Modificaciones _____

CUESTIONARIO PARA LA ADAPTACIÓN DE LA HONOS EN URUGUAY

Ítem 7: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 8: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 9 : muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 10: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 11 : muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 12 : muy poco (1)___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

2.b) Valore los ítems con respecto a la claridad en su forma de redacción

Ítem 1: muy poco (1)___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 2: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 3:muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

CUESTIONARIO PARA LA ADAPTACIÓN DE LA HONOS EN URUGUAY

Ítem 4: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 5: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 6: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 7: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 8: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 9 : muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 10: muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 11 : muy poco (1) ___ poco(2)___ regular(3)___ bastante(4)___ total(5)___
 Modificaciones_____

Ítem 12 : muy poco (1)___ poco(2)___ regular(3)___ bastante(4)___
 total(5)_____
 Modificaciones_____

CUESTIONARIO PARA LA ADAPTACIÓN DE LA HONOS EN URUGUAY

3) Valore el Formato de la Escala en general considerando los objetivos de la misma (utilidad, relevancia, brevedad, sencillez en la aplicación y puntuación, etc.)

3.a) Extensión: inadecuada (1)____ poco adecuada(2)____ adecuada(3)____ muy adecuada (4)____ totalmente adecuada (5)____

Modificaciones: _____

3.b) Orden: inadecuado (1) ____ poco adecuado (2) ____ adecuado (3)____ muy adecuado (4)____ totalmente adecuado (5)____

Modificaciones: _____

3.c) Cantidad de opciones de respuesta: inadecuada (1)____ poco adecuada(2)____ adecuada(3)____ muy adecuada (4)____ totalmente adecuada (5)____

Modificaciones: _____

**CUESTIONARIO SOBRE LA FICHA COMPLEMENTARIA A LA ESCALA
HoNOS URUGUAY**

Nombre del profesional _____

Considerando los fines de la Escala HoNOS y de su Ficha complementaria

1) ¿Considera necesario agregar alguna pregunta/dato? Si ____ No ____

En caso afirmativo, especifique

2) ¿Considera necesario quitar alguna pregunta? Si ____ No ____

En caso afirmativo, especifique

3) ¿Considera necesario modificar alguna pregunta? Si ____ No ____

En caso afirmativo, especifique

3) Valore el Formato de la Hoja Complementaria en general

3.a) Extensión: inadecuada (1)____ poco adecuada(2)____ adecuada(3)____ muy adecuada (4)____ totalmente adecuada (5)____

Modificaciones: _____

3.b) Orden: inadecuado (1) ____ poco adecuado (2) ____ adecuado(3)____muy adecuado (4)____ totalmente adecuado (5)____

Modificaciones: _____

3.c) Cantidad de opciones de respuesta: inadecuada (1)____ poco adecuada(2)____ adecuada(3)____ muy adecuada (4)____ totalmente adecuada (5)____

Modificaciones: _____

Otros: _____

HOJA DE REGISTRO DE HoNOS

HOJA DE REGISTRO DE HoNOS**INFORMACIÓN BÁSICA****Nombre del Centro** _____**Identificación del usuario** _____**Fecha de nacimiento del usuario** ____/____/____**Sexo paciente** 1. Mujer 2. Hombre

HOJA DE REGISTRO DE HONOS

HOJA DE EVALUACIÓN ESCALA HONOS

Identificación del paciente _____

Fecha de hoy (dd/mm/aa)	/ /	/ /	/ /	/ /
Identificación del evaluador				
Profesión del evaluador	1. Psiquiatra	2. Psicólogo	3. Otro (especifique)	
Período evaluado	1. 1° consulta/ingreso	2. Egreso/alta	3. Consulta	4. Otro (especifique)
Ámbito de aplicación de la Escala:	1. Hospital psiquiátrico de Agudos	2. Hospital psiquiátrico de Larga estadía	3. Hospital General	4. Centro Diurno
	5. Policlínica ambulatoria	6. Otro (especificar)		
Diagnóstico primario:	0. demencia	1. Drogas/alcohol/ Abuso de sustancias	2. Esquizofrenia/ trastornos delirantes	
	3. T. E. de ánimo	4. Neurosis/ansiedad	5. trastornos sueño, alimentación, etc	
	6. Trastorno de personalidad	7. Otro (especificar)		
1. Conducta hiperactiva, agresiva, disruptiva o agitada	0 1 2 3 4			
2. Autolesiones no accidentales	0 1 2 3 4			
3. Consumo Problemático de Alcohol o Drogas	0 1 2 3 4			
4. Problemas cognitivos	0 1 2 3 4			
5. Problemas por enfermedad física o discapacidad	0 1 2 3 4			
6. Problemas asociados a la presencia de ideas delirantes y alucinaciones	0 1 2 3 4			
7. Problemas en relación con el humor depresivo	0 1 2 3 4			
8. Otros problemas mentales o conductuales	0 1 2 3 4			
Especificar el tipo de trastorno:	A, B, C, D, E, F, G, H, I, J			
9. Problemas con las relaciones	0 1 2 3 4			
10. Problemas en relación con las actividades de la vida cotidiana	0 1 2 3 4			
11. Problemas con las condiciones de vida	0 1 2 3 4			
12. Problemas en relación con la ocupación y las actividades	0 1 2 3 4			
Puntuación Total :	0 - 48			

HOJA DE REGISTRO DE HoNOS

HOJA DE REGISTRO DE HoNOS- INFORMACIÓN COMPLEMENTARIA *

Nombre del Centro _____

Identificación del usuario _____

Identificación del evaluador _____

DATOS DEL USUARIO

Nivel educativo alcanzado: 1. Primaria 2. Secundaria 3. UTU 4. Universitario incompleto 5. Universitario completo	
Trabaja actualmente : 1.Si 2.No	
Ciudad de residencia: 1. Montevideo 2. Interior(especifique)	
Ciudad natal: 1.Montevideo 2. Interior (especifique)	
Estado civil: 1.soltero/a 2.casado/a 3.separado/divorciado/viudo 4.Unión libre	
Vive con: 1.padres o tutores 2. Pareja 3.solo 4.Otro(especifique)	
Es la primera vez que recibe atención en un servicio de salud mental 1.Si 2.No (especifique el año de inicio usando aaaa)	

DATOS DEL TRATAMIENTO

Fecha de hoy (dd/mm/aa)	/ /	/ /	/ /
Tratamiento indicado Farmacológico/biológico: 1.No 2.Si (especifique)			
Psicoterapéutico 1.No 2.Si (especifique)			
Social 1.No 2.Si (especifique)			
Otro (especifique)			
Eventos relevantes desde la última evaluación			

* Los datos de esta hoja amplían la información de la Hoja de Evaluación HoNOS básica y permiten un análisis más completo para evaluar los resultados del tratamiento, pero su uso no es obligatorio.

CUESTIONARIO DE VIABILIDAD DE HoNOS

Cuestionario de viabilidad para el profesional que ha aplicado la escala HoNOS**Instrucciones (verbales)**

Luego de haber participado del estudio se solicita su opinión respecto a la Escala y su experiencia en general.

En primer lugar deben completar datos sobre su participación. En segundo lugar, calificar la Escala HoNOS en algunos aspectos. Se plantean 5 opciones donde 1 es “ninguno/nada”, 2 “algo”, 3 “regular”, 4 “bastante”, 5 “máximo/total”.

CUESTIONARIO DE VIABILIDAD DE HoNOS

A partir de su experiencia califique la aplicación de la escala HoNOS con un puntaje del 1 al 5, donde 1 es ninguno y 5 es máximo o total.

Utilidad	1	2	3	4	5
Relevancia del contenido que evalúa	1	2	3	4	5
Facilidad para la aplicación	1	2	3	4	5
El tiempo que insume su aplicación	1	2	3	4	5
Necesidad de recibir una capacitación previa	1	2	3	4	5
¿En qué medida la utilizaría para evaluar resultados en su práctica clínica?	1	2	3	4	5