

**ANÁLISIS RASCH DEL CUESTIONARIO DE CALIDAD DE VIDA FORMA CORTA –
WHOQOL-BREF EN POBLACIÓN COLOMBIANA**

DIEGO FERNANDO ROJAS GUALDRÓN

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA DE CIENCIAS SOCIALES

FACULTAD DE PSICOLOGÍA

MAESTRÍA EN PSICOLOGIA

BUCARAMANGA 2014

**ANÁLISIS RASCH DEL CUESTIONARIO DE CALIDAD DE VIDA FORMA CORTA –
WHOQOL-BREF EN POBLACIÓN COLOMBIANA**

DIEGO FERNANDO ROJAS GUALDRÓN

Trabajo de grado presentado para optar al título de: Magister en Psicología

LUIS CARLOS OROZCO VARGAS, MD MSc

Director

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA DE CIENCIAS SOCIALES

FACULTAD DE PSICOLOGÍA

MAESTRÍA EN PSICOLOGIA

BUCARAMANGA 2014

Tabla de contenido

Resumen	9
Abstract	10
Introducción-Justificación.....	11
Pregunta problema	15
Objetivo Principal.....	15
Objetivos Específicos.....	15
Hipótesis de Investigación.	15
Marco teórico.....	16
Conceptos Matemáticos y estadísticos básicos para la comprensión del Modelo de Rasch.	16
Funciones de Probabilidad.	16
Bernoulli	16
Binomial	17
Poisson.....	17
Distribución Normal.....	17
Estimación por Máxima Verosimilitud (Maximum Likelihood)	18
Modelos de Escalamiento de Variables Latentes.....	21
Modelo de Rasch.....	28
Principios básicos del Modelo de Rasch	29
Estimación de Parámetros y determinación del Ajuste	31
Metodología Wolfe-Smith para validación de mediciones con los Modelos Rasch siguiendo la propuesta de Messick.....	32
Contenido.....	33
Sustantiva	34
Estructural.....	35
Generalizabilidad	36
Externa	37
Consecuencial.....	37
Interpretabilidad	37

El concepto de calidad de vida.....	38
Perspectivas sociales de conceptualización de la calidad de vida.....	39
Perspectivas de conceptualización de calidad de vida relacionada con la salud	40
El concepto de calidad de vida subyacente a los instrumentos WHOQOL de la Organización Mundial de la Salud	41
Antecedentes	44
Método	47
Tipo y diseño de la investigación	47
Población	47
Sujetos	47
Muestreo	47
Tamaño de la muestra	47
Variables e Instrumentos.....	48
Análisis de datos.	49
Adecuación de las categorías de valoración de los ítems.	49
Ajuste de los ítems	49
Unidimensionalidad	50
Ajuste de las personas	50
Funcionamiento Diferencia de los Ítems (DIF).....	50
Confiabilidad de los ítems y las personas	51
Jerarquía de los ítems y Targeting	51
Asociación entre valoración global de la salud y la medida del WHOQOL-BREF	51
Comparación con las dificultades reportadas por los antecedentes.....	51
Resultados.....	53
Descripción de la muestra.....	53
Adecuación de las categorías de valoración.	53
Cinco categorías de respuesta.....	53
Tres categorías de respuesta	55
Ajuste de los ítems	57
Unidimensionalidad de la Escala Total	58
Ajuste de las personas	60
Funcionamiento Diferencial de los Ítems.....	60

Funcionamiento Diferencial de los Ítems por sexo	60
Funcionamiento Diferencial de los Ítems por percepción subjetiva de enfermedad....	61
Funcionamiento Diferencial de los Ítems por grupos de edad	62
Funcionamiento Diferencial de los Ítems por nivel de calidad de vida	63
Gráficas de decisión para DIF	64
Comparación de la medida de calidad de vida con y sin ítems DIF.....	66
Confiabilidad de los Ítems y de las Personas	68
Jerarquía de los ítems y Targeting	68
Relación de la medición Rasch del WHOQOL-BREF con autovaloración de la calidad de vida	71
Relación de la medición Rasch del WHOQOL-BREF con autovaloración de la satisfacción con la salud	72
Comparación de los resultados obtenidos con los antecedentes.....	74
Discusión	76
En relación al objetivo “Analizar las características de las categorías de respuesta a los ítems del WHOQOL-BREF.”	76
En relación al objetivo “Determinar el ajuste individual de los ítems del WHOQOL-BREF y las personas al Modelo de Rasch.”	76
En relación al objetivo “Determinar la unidimensionalidad, confiabilidad y separabilidad de la escala total del WHOQOL-BREF.”	77
En relación al objetivo “Evaluar el Funcionamiento Diferencial de los Ítems del WHOQOL-BREF por sexo, reporte subjetivo de enfermedad, edad y nivel de calidad de vida.”	78
En relación al objetivo “Analizar las relaciones entre la medición de calidad de vida del WHOQOL-BREF y autovaloración de variables de salud.”	80
En relación al objetivo “Comparar las dificultades obtenidas en los ítems del WHOQOL-BREF con las reportadas por los antecedentes”	81
En relación al objetivo general “Validar el Constructo de Calidad de vida en el WHOQOL-BREF desde el Modelo de Rasch y siguiendo los lineamientos de validación de constructo de Messick estructurados en la metodología de Wolfe y Smith”	82
Contenido.	82
Sustantiva.	82
Estructural.....	84
Generalizabilidad	84
Validación externa.....	85

Conclusiones	86
Referencias	88

Lista de tablas

Tabla 1 Valores de dificultad para los ítems del WHOQOL-BEF reportados en los antecedentes.....	45
Tabla 2 Caracterización de la muestra de análisis y comparación con los casos excluidos	53
Tabla 3 Estructura de categorías escala total 5 puntos	54
Tabla 4 Estructura de categorías escala total 3 puntos	56
Tabla 5 Polaridad de los ítems Escala total 3 puntos.....	58
Tabla 6 Análisis de componentes principales escala total 3 puntos	59
Tabla 7 Relaciones aproximadas entre las medidas de las personas contraste 1 escala total 3 puntos.....	60
Tabla 8 Funcionamiento diferencial de los ítems por sexo.....	61
Tabla 9 Funcionamiento diferencial de los ítems por reporte de enfermedad.....	62
Tabla 10 Funcionamiento diferencial de los ítems por grupos de edad	63
Tabla 11 Funcionamiento diferencial de los ítems por nivel de calidad de vida	64
Tabla 12 Resumen de cálculos Funcionamiento Diferencial de los Ítems.....	64
Tabla 13 Comparación ajuste de los ítems a partir de 24 y 20 ítems del WHOQOL-BREF	67
Tabla 14 Confiabilidad de las personas.....	68
Tabla 15 Confiabilidad de los Ítems	68
Tabla 16 Estadísticos descriptivos de la medida Rasch por categorías de autovaloración de calidad de vida.....	72
Tabla 17 Análisis de Bonferroni medida Rasch por autovaloración de calidad de vida.....	72
Tabla 18 Estadísticos descriptivos de la medida Rasch por categorías de autovaloración de satisfacción con la salud	73
Tabla 19 Análisis de Bonferroni medida Rasch por autovaloración de satisfacción con la salud	73
Tabla 20 Ítems dentro de los intervalos de acuerdo comparación Rojas 2013 con antecedentes.....	74

Lista de gráficas

Gráfica 1 Evidencia de validez relevante para aspectos de validez (Adaptado de Wolfe y Smith, 2007b).....	33
Gráfica 2 Modelo sistémico de la calidad de vida, tomado de Shye (1989).....	41
Gráfica 3 Dimensiones y facetas del WHOQOL-100 y equivalencia con WHOQOL-BREF	43
Gráfica 4 Medidas empíricas ítem-categoría escala total 5 puntos	54
Gráfica 5 Curva de probabilidad de 5 categorías Ítem 3.....	55
Gráfica 6 Medidas empíricas ítem-categoría escala total 3 puntos	56
Gráfica 7 Curva de probabilidad de 3 categorías Ítem 3.....	57
Gráfica 8 Residuales estandarizados contraste 1 escala total 3 puntos.....	59
Gráfica 9 Gráfica de Bland y Altman Funcionamiento Diferencial por variables de clasificación	65
Gráfica 10 Gráfica de Bland y Altman para diferencias de medida entre formas de 24 y 20 ítems	66
Gráfica 11 Gráfica de Bland y Altman para dificultades de 24 y 20 ítems.....	67
Gráfica 12 Mapa Ítem-Persona WHOQOL-BREF	69
Gráfica 13 Valores en logitos para cada categoría de respuesta por ítem del WHOQOL-BREF	70
Gráfica 14 Comparación de los mapas ítem-persona 24 y 20 ítems.....	71
Gráfica 15 Gráficas de Bland y Altman para la comparación de las dificultades obtenidas con las reportadas en los antecedentes	75
Gráfica 16 Clasificación de los ítems del WHOQOL-BREF según el marco sistémico de Shye (1989).....	83

RESUMEN GENERAL DE TRABAJO DE GRADO

TITULO: Validez y Análisis psicométrico desde el Modelo de Rasch al Cuestionario de Calidad de Vida forma corta – WHOQOL-BREF en sujetos colombianos

AUTOR(ES): Diego Fernando Rojas Gualdrón

FACULTAD: Facultad de Psicología

DIRECTOR(A): Luis Carlos Orozco Vargas

RESUMEN

Se han realizado diversos estudios sobre las propiedades psicométricas del WHOQOL-BREF, sin embargo los trabajos desde el modelo de Rasch son limitados. Objetivo: Validar el Constructo de Calidad de vida en el WHOQOL-BREF desde el Modelo de Rasch y siguiendo los lineamientos de validación de constructo de Messick estructurados en la metodología de Wolfe y Smith. Método: Se aplicó el WHOQOL-BREF a 944 participantes santandereanos mayores de 16 años. Se obtuvieron evidencias de validez las cuales fueron analizadas según los lineamientos de Wolfe y Smith. Resultados: El 84% de los casos se incluyeron en los análisis. La forma original de calificación de 5 categorías (12345) debió ser colapsada a 3 categorías (11123). Los ítems mostraron un ajuste medio OUTFIT de 1.02(.24); las personas mostraron un ajuste medio OUTFIT de 1.02(.48). Se encuentra una estructura unidimensionalidad que explica el 38.8% de la varianza. Cuatro ítems mostraron DIF, sin embargo, al comparar la medida total calculada con y sin estos ítems las diferencias no son significativas. Los ítems muestran una confiabilidad de .99 y un índice de separación de 16.6; se reportan los mapas de ítems e ítems-personas. Se obtienen asociaciones significativas con las variables de autovaloración de salud. La comparación de las dificultades obtenidas con las reportadas en la literatura muestra resultados poco favorables. Discusión: Se analizan los resultados desde las diversas facetas de la validez. La evidencia se considera suficiente para considerar válida la medida de calidad de vida del WHOQOL BREF. Conclusiones: Los resultados plantean un modelo unidimensional y jerárquico de las facetas del WHOQOL-BREF que permitirá fortalecer los análisis conceptuales sobre el constructo de calidad de vida. Se recomienda revisar con mayor detalle la generalizabilidad de la medida del WHOQOL-BREF a partir de análisis Rasch.

PALABRAS CLAVES:

Calidad de vida, WHOQOL-BREF, Validez, Análisis Rasch.

V° B° DIRECTOR DE TRABAJO DE GRADO



GENERAL SUMMARY OF WORK OF GRADE

TITLE: Rasch Analysis of the Quality of Life Questionnaire Short Form – WHOQOL-BREF in Colombia

AUTHOR(S): Diego Fernando Rojas Gualdrón

FACULTY: Facultad de Psicología

DIRECTOR: Luis Carlos Orozco Vargas

ABSTRACT

Many psychometric analysis of the WHOQOL-BREF have been done, however studies from the Rasch perspective are limited. The purpose of the study was to validate the WHOQOL-BREF quality of life construct from the Rasch Model perspective and following the construct validity guidelines by Wolfe and Smith. Methods: 944 participants filled the WHOQOL-BREF. Validity evidences were obtained and analyzed according to Wolfe and Smith guidelines. Results: 84% of cases were included in analyses. The original 5 points response pattern (12345) oughted be collapsed to 3 points (11123) in order to improve fit. Item fit OUTFIT 1.02(.24); Person fit OUTFIT 1.02(.48). An unidimensional structure was adjusted explaining 38.8% of variance. Four items presented DIF, however, item DIF didn't affected test functioning. Item reliability .99, separability 16.6 strates. Item maps are presented. The obtained difficulties on colombian sample are not consistent with international reports. Discussion: Results are analyzed from the different facets of validity. Evidence is enough to consider as valid the WHOQOL-BREF measure of quality of life. Conclusions: The findings supports a unidimensional and hierarchical model of the WHOQOL-BREF that allows to improve the conceptual analysis about the quality of life construct. Further analysis on generalizability from a Rasch perspective are required.

KEYWORDS:

Quality of life, WHOQOL-BREF, Validity, Rasch Analysis

V° B° DIRECTOR OF GRADUATE WORK



Validez y Análisis psicométrico desde el Modelo de Rasch al Cuestionario de Calidad de Vida forma corta – WHOQOL-BREF en sujetos colombianos¹

Introducción-Justificación

Mateus & Rojas, (2009) describen como durante la década de los 90's, la Organización Mundial de la Salud (OMS) reconoce y respalda la tendencia que se viene dando de ampliar las mediciones en salud más allá de indicadores como el de mortalidad y morbilidad (World Health Organization, 1991), para incluir medidas de impacto de las enfermedades y discapacidades en las actividades diarias y el comportamiento, medidas de salud percibida y medidas de funcionalidad. Posteriormente, la OMS reconoce que aun teniendo en cuenta estas medidas del impacto de la enfermedad, se está dejando a un lado la calidad de vida. Bajo este orden de ideas, se comienza a incluir el elemento “humano” en el cuidado de la salud, reconociendo que el objetivo central de ésta es el bienestar del paciente (World Health Organization, 1998). Todos estos postulados acerca de la necesidad de abarcar el aspecto positivo de la salud, entendida como un “Estado de completo bienestar físico, mental y social y no solamente la ausencia de enfermedad.” (World Health Organization, 1946) se ven plasmados en nuevos programas que generan productos como el World Health Organization Quality of Life (WHOQOL y WHOQOL-BREF), una herramienta para la medición de calidad de vida planteada para su uso en la práctica, investigación, auditoría, legislación en salud y medición de la efectividad de tratamientos (World Health Organization, 1998).

El WHOQOL como otros instrumentos de medición, para ámbitos clínicos y de la salud, ha sido pensado tanto para medir estados y condiciones, como resultados de procesos de intervención individual y colectiva e incluso para medir el impacto de políticas. Usualmente, estos instrumentos han sido contruidos de acuerdo con las nociones de Confiabilidad y Validez de la teoría psicométrica tradicional (Nunnally, Bernstein, & Arellano, 1995), la cual ha mostrado dificultades para su justificación estadística. De forma reciente, aproximaciones modernas a la psicometría se han propuesto para solventar el problema de

¹ Investigación financiada por la Fundación Universitaria de San Gil, UNISANGIL.

cómo medir adecuadamente. Entre estas aproximaciones recientes se cuenta con el Modelo de Medición de Rasch, en el cual los datos recogidos a través de estos instrumentos se ponen a prueba en función de los supuestos que plantea el modelo. Estos supuestos que plantea el Modelo de Rasch, han sido considerados como una “plantilla” que operacionaliza los axiomas formales que subyacen a la medición (Karabatsos, 2001). Esto permite un cambio paradigmático importante. El modelo de Rasch abre las puertas a la consideración de mediciones basadas en cuestionarios y técnicas afines, como mediciones específicamente objetivas, permitiendo establecer características métricas robustas y suficientes.

El lograr estas características para instrumentos como el WHOQOL y similares, podría permitir evaluar de mejor forma políticas de salud, pero también políticas de desarrollo y promoción social en términos de su impacto sobre la salud y la calidad de vida de las personas. En particular, Colombia dadas sus condiciones económicas, sociales y políticas necesita que los científicos y profesionales de la salud, asuman una visión transdisciplinar y contribuyan de forma directa y significativa a los problemas sociales de la Nación. El aporte que busca dar este proyecto, está en contribuir a la dotación de herramientas válidas para la toma de decisiones en salud. Esto, con miras a futuro de desarrollar mediciones (en el sentido de Rasch) de salud y de sus determinantes sociales. A continuación se bosquejan unos postulados que justifican proyectos con esta perspectiva y propósito tanto para la medición de la salud, como para la medición de sus determinantes:

En términos de salud.

Los gobiernos a nivel local, departamental y nacional cuentan con un mecanismo de atención a la salud derivado del Sistema de Seguridad Social en Salud (o equivalente), el cual a su vez recoge los aportes realizados por la Organización Mundial de la salud OMS y la Organización Panamericana de la Salud OPS entre otras instituciones internacionales. El sistema colombiano se basa en el concepto de seguridad social integral la cual es definida en la ley 100 de 1993 como:

*“La Seguridad Social Integral es el conjunto de instituciones, normas y procedimientos, de que disponen la persona y la comunidad para **gozar de una calidad de vida**, mediante el cumplimiento progresivo de los planes y programas que el Estado y la sociedad desarrollen para proporcionar la cobertura integral de*

*las contingencias, especialmente las que menoscaban la salud y la capacidad económica, de los habitantes del territorio nacional, con el fin de lograr el bienestar individual y la integración de la comunidad.*² (Colombia, Republica de, 1993)

Esta definición direcciona todas las políticas y programas relacionados con la ley 100, cuyo cumplimiento es monitoreado en informes epidemiológicos tales como los “Indicadores Básicos 2010 Situación de salud en Colombia” (Protección Social, Ministerio de la, 2011) En estos informes se encuentran indicadores de morbilidad, mortalidad, inversión en salud, cobertura en salud, red hospitalaria, talento humano en salud, ingreso económico, acceso a servicios públicos, desigualdad económica, entre otros los cuales se relacionan a nivel de salud pública con la salud, la salud mental y la calidad de vida. Sin embargo no se plantean indicadores que permitan un seguimiento a estos estados propiamente.

Estas medidas de morbilidad, mortalidad, inequidad social e indicadores de servicios son indispensables, aunque solo permiten un análisis indirecto de la salud, salud mental y calidad de vida. Las conclusiones a las que se puede llegar mediante esta información se restringen a la salud como producto del cumplimiento de los objetivos de las políticas. Sin embargo, para conocer el estado de salud, en su concepción amplia, se requieren indicadores directos que la conciben como el proceso vivenciado por las personas, indicadores validos con capacidad para discriminar el estado de salud, salud mental y calidad de vida y el impacto que tienen las políticas y programas sobre los mismos.

El presente proyecto, busca validar el Modelo de medición del WHOQOL-BREF en población general colombiana para su uso en procesos de medición sobre esta población.

En términos de Determinantes Sociales de la salud.

Una de las metas del plan Nacional de desarrollo 2010-2014 en el apartado “Igualdad de oportunidades para la prosperidad social”, es el actuar en pro de una Política integral de desarrollo y protección social; de la promoción social y en políticas diferenciadas para la inclusión social. Esto bajo una visión de sociedad equitativa:

² Negrita añadida.

“La visión de sociedad que queremos es la de una sociedad con igualdad de oportunidades y con movilidad social, en (la) que nacer en condiciones desventajosas no signifique perpetuar dichas condiciones a lo largo de la vida, sino en que el Estado acude eficazmente para garantizar que cada colombiano tenga acceso a las herramientas fundamentales que le permitirán labrar su propio destino, independientemente de su género, etnia, posición social, orientación sexual o lugar de origen.” (Departamento Nacional de Planeación, 2011)

Para lograrlo se reconocen como medio el crecimiento económico sostenible y el desarrollo social integral, y se establece una relación dialógica entre estas. La razón última de ambas estrategias es la búsqueda de una población más productiva y con mayor Calidad de Vida.

Al centrar estas consideraciones con las recomendaciones dadas en los planteamientos del PND 2010-2014 sobre la salud, estos son consecuentes con la Resolución WHA62/R14 de la Asamblea Mundial de la Salud, (2009): “Reducir las inequidades sanitarias actuando sobre los determinantes sociales de la salud” y el Documento A62/9 (Asamblea Mundial de la Salud, 2009a): Informe de la Comisión sobre Determinantes Sociales de la Salud. Los cuales plantean como ejes: (1) mejorar las condiciones de vida; (2) luchar contra la distribución no equitativa del poder, el dinero y los recursos, y (3) medir la magnitud del problema, analizarlo y evaluar los efectos de las intervenciones. Las acciones a ejecutar durante este cuatrienio, según el PND, son consecuentes con los dos primeros puntos.

El presente proyecto, busca realizar un aporte al tercero, validando un Instrumento útil como desenlace en la medición de la magnitud de las desigualdades en términos de capital social y capital económico y su impacto sobre la salud y la calidad de vida.

Así las cosas, dadas las obligaciones y responsabilidades que los actores de la salud, tanto académicos, científicos, practicantes y gobierno tienen para con el colectivo, es indispensable contribuir al aumento de la claridad sobre los fenómenos que se abarcan. Es por esto que difícilmente se pueden aportar soluciones eficaces a problemas conceptuales escasamente elaborados y metodológicamente poco sustentados. En general, se espera que los resultados logrados aporten tanto a nivel conceptual como metodológico en relación a las problemáticas planteadas y justificadas.

Pregunta problema

En consideración a lo expuesto, la investigación busca determinar la calidad métrica del WHOQOL-BREF asumiendo como pregunta principal ¿Se ajusta el WHOQOL-BREF a los supuestos del Modelo de Rasch de tal forma que pueda considerarse un instrumento de Medición específicamente objetivo? Y de no ser así ¿Qué modificaciones deben realizarse sobre este para lograr el ajuste?

Objetivo Principal.

Validar el Constructo de Calidad de vida en el WHOQOL-BREF desde el Modelo de Rasch y siguiendo los lineamientos de validación de constructo de Messick estructurados en la metodología de Wolfe y Smith.

Objetivos Específicos.

Analizar las características de las categorías de respuesta a los ítems del WHOQOL-BREF.

Determinar el ajuste individual de los ítems del WHOQOL-BREF y las personas al Modelo de Rasch.

Determinar la unidimensionalidad, confiabilidad y separabilidad de la escala total del WHOQOL-BREF.

Evaluar el Funcionamiento Diferencial de los Ítems del WHOQOL-BREF por sexo, reporte subjetivo de enfermedad, edad y nivel de calidad de vida.

Analizar las relaciones entre la medición de calidad de vida del WHOQOL-BREF y autovaloración de variables de salud

Comparar las dificultades obtenidas en los ítems del WHOQOL-BREF con las reportadas por los antecedentes

Hipótesis de Investigación.

El WHOQOL-BREF es una medida unidimensional con nivel de intervalo de la calidad de vida.

Marco teórico

Conceptos Matemáticos y estadísticos básicos para la comprensión del Modelo de Rasch.

Este primer apartado no pretende de ninguna manera constituirse en un compendio de estadística teórica. Por el contrario su intención es centrar y traer a escena algunos conceptos y procedimientos matemáticos y estadísticos que son necesarios para la comprensión del modelo de Rasch.

Funciones de Probabilidad.

Las funciones de densidad de probabilidad (FDP) son fórmulas matemáticas que proveen información acerca de cómo están distribuidos los valores de variables aleatorias. Para variables aleatorias discretas (que toman valores particulares dentro de un intervalo), la fórmula da la probabilidad para cada valor de la variable. En el caso de variables aleatorias continuas (que pueden tomar cualquier valor dentro de un intervalo), la fórmula especifica la densidad de probabilidad de que las variables tomen valores dentro de un rango en particular. Esto se hace a partir del área bajo la curva definida por la FDP. En este apartado no se retoman todas las funciones de probabilidad, sino solo aquellas necesarias para la comprensión del modelo de Rasch especialmente algunas funciones para variables aleatorias discretas y por parte de las continuas la distribución normal. El tema puede ser consultado con más detalle en diversos libros de Probabilidad y Estadística (Balakrishnan & Nevzorov, 2004; Evans, Hastings, & Peacock, 2000).

Bernoulli

Función de Probabilidad para una variable aleatoria X que puede tomar solo dos valores (ej. 1 y 0). Lanzar una moneda es el ejemplo más típico. La densidad se define como:

$$P(X = x_1) = p \quad P(X = x_0) = 1 - p$$

Donde x_1 y x_0 son los dos valores posibles para la variable aleatoria (comúnmente: éxito y fracaso) y P denota la probabilidad. El único parámetro para la función de probabilidad de Bernoulli es la probabilidad de "éxito", p . El valor esperado para X es p y su varianza es $p(1 - p)$.

Binomial

Función de Probabilidad para una variable aleatoria X la cual es el número de “éxitos” en una serie de n variables independientes de Bernoulli. El número de “caras” en n lanzamientos de una moneda es el ejemplo más típico. La densidad se define como:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad |(2)$$

$$x = 0, 1, 2, \dots, n.$$

La probabilidad de una variable aleatoria de tomar un valor dentro de un rango se encuentra simplemente sumando las FDP sobre el rango requerido. El valor esperado de X es np y su varianza es $np(1-p)$. La Binomial es importante para asignar probabilidades a eventos simples como la probabilidad de obtener cuatro o más veces un número entre diez lanzamientos de dado; también lo es para la significación estadística de ciertas pruebas de hipótesis (ej. Prueba de los signos) y como la distribución de error asumida en la Regresión Logística.

Poisson

Función de densidad de probabilidad que “surge” naturalmente en muchas situaciones, particularmente como modelo de probabilidad de ocurrencia de eventos raros. Adicionalmente, la función de densidad de Poisson es la distribución límite de la Binomial cuando p es pequeño y n es grande. La función de densidad de Poisson para una variable X que toma valores enteros entre cero e infinito se define como:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad |(3)$$

$$x = 0, 1, 2, \dots, \infty.$$

El único parámetro para la función de densidad de Poisson, λ , es tanto el valor esperado como la varianza, es decir media y varianza de una variable aleatoria Poisson son iguales.

Distribución Normal

La FDP $f(x)$ para una variable aleatoria continua X se define como:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right] \quad |(4)$$

$$-\infty < x < \infty$$

Donde μ y σ^2 son media y varianza respectivamente. Cuando estos valores son cero y uno, la densidad resultante se denomina normal estándar. La función de densidad normal muestra forma de campana. En el caso de las variables aleatorias continuas, la probabilidad de que esta tome un valor particular es estrictamente cero, las probabilidades no-cero solo pueden ser asignadas a algún rango de valores de la variable. La FDP normal es la base de una gran cantidad de métodos estadísticos que asumen densidad normal en los datos observados o en los términos de error de los modelos. Esto en parte se justifica acudiendo al Teorema del Limite Central (Balakrishnan & Nevzorov, 2004).

Estimación por Máxima Verosimilitud (Maximum Likelihood)

La estimación por máxima verosimilitud (MV) es utilizada en varios modelos estadísticos avanzados como el Modelamiento de Ecuaciones Estructurales. También se denomina de esta forma al método empleado para estimar los parámetros de los ítems en el modelo de Rasch, condicionando la función de verosimilitud sobre el número de respuestas correctas (Georg Rasch, 1980). En caso de que haya que estimar los parámetros de ítems y sujetos simultáneamente se emplean los métodos de máxima verosimilitud conjunta (Birnbaum, 1968; Frederic M. Lord, 1974) o de máxima verosimilitud marginal (Bock & Aitkin, 1981; Darrell Bock & Lieberman, 1970). El método de Máxima Verosimilitud Conjunta es el empleado por el Software WINSTEPS al igual que el método PROX el cual permite hacer las estimaciones iniciales a partir de las cuales se realizan las iteraciones (J. Linacre, 2012a).

Las estimaciones obtenidas con MV tienen las ventajas de ser consistentes, es decir, el parámetro se acerca al parámetro poblacional a medida que el tamaño de la muestra aumenta; y además son eficientes, es decir, tienen la menor varianza o fluctuación muestral posible. En el marco de estas ventajas, se debe reconocer que estas características son asintóticas, por lo que se cumplen en muestras de gran tamaño.

El objetivo de la estimación por MV es identificar los parámetros poblacionales (media, coeficientes de regresión, etc.) que con mayor probabilidad hayan producido los datos de la muestra. Esto es posible calculando el valor conocido como Verosimilitud (Likelihood), el cual resume el ajuste de los datos a una estimación de parámetro en particular. La verosimilitud es conceptualmente similar a la Probabilidad pero no son estrictamente lo mismo. Según Clayton y Hillas, 1993 la verosimilitud es una medida del

soporte que provee un cuerpo de datos para un valor en particular del parámetro de un modelo de probabilidad. Para determinar qué tan “verosímiles” son los datos de la muestra se debe en primer lugar hacer una suposición sobre la distribución de las puntuaciones poblacionales, para esto se recurre frecuentemente a la distribución normal.

La Función de Distribución de Probabilidad Normal juega un papel clave en el cálculo del valor de verosimilitud para una muestra, para ejemplificar, supóngase un conjunto de datos del cual se conocen los valores poblacionales media igual a 20 y desviación estándar igual a 5. La verosimilitud asociada a una puntuación de 21 se calcularía remplazando en (5) de la siguiente forma:

$$L_{(21)} = \frac{1}{\sqrt{2(3.14)(5)^2}} 2.71^{-\frac{[21-20]^2}{5^2}} = .07821 \quad | (5)$$

El valor resultante .07821 puede interpretarse como la probabilidad relativa de $x_i=21$ en una población normalmente distribuida con media de 20 y desviación estándar de 5. Gráficamente, ,07821 representaría la altura de la distribución normal al punto del valor 21. Este concepto es extensible para cada caso dentro de un conjunto de datos, estos pueden calcularse de forma similar. Por ejemplo, manteniendo los mismos parámetros poblacionales el valor de verosimilitud para 5 y 33 serían ,001 y ,003 respectivamente. Esto evidencia que la probabilidad relativa de $x_i = 5$ es de aproximadamente ,001 lo cual es mucho menor que la probabilidad relativa de $x_i = 21$, permitiendo establecer que es más probable que 21 haya ocurrido en una población normalmente distribuida con $\mu = 20$ y $\sigma^2 = 5$. Se puede deducir que la verosimilitud asociada a un valor de x_i cambiara a medida que los parámetros poblacionales lo hagan (Clayton, 1993).

Una vez se establecen las verosimilitudes para los valores individuales de x_i , la verosimilitud para la población se obtiene multiplicando. Desde la teoría de la probabilidad, la probabilidad conjunta para un conjunto de eventos independientes se obtiene multiplicando las probabilidades individuales (ej. La probabilidad de observar dos caras en lanzamientos independientes de monedas es $,5^* ,5 = ,25$). En términos estrictos las verosimilitudes no son probabilidades, la probabilidad asociada con cualquier puntuación individual de una FDP continua es cero. Sin embargo, el valor de Verosimilitud para la muestra completa está definido como el producto de las verosimilitudes individuales de la siguiente forma:

$$L = \prod_{i=1}^N \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[x_i - \mu]^2}{2\sigma^2}} \right\} \quad | (6)$$

Llevar a cabo este cálculo utilizando los valores individuales de L_i en una muestra, lleva generalmente a valores muy pequeños. Ya que L se torna cada vez más pequeño conforme la muestra aumenta, el logaritmo de (6) puede utilizarse para solucionar este inconveniente. Cabe recordar que una de las reglas de la logaritmicación dice que $\log(ab) = \log(a) + \log(b)$. Esta aplicación del logaritmo convierte a (6) en la Log-Verosimilitud, la cual además tiene la ventaja de ser un modelo aditivo y no multiplicativo (Clayton, 1993). Esta quedaría definida de la siguiente forma:

$$\ln L = \sum_{i=1}^N \ln \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[x_i - \mu]^2}{2\sigma^2}} \right\} \quad | (7)$$

Una vez calculados los valores individuales $\text{Log } L_i$ el valor de Log-Verosimilitud es igual a la suma de los valores individuales. Este valor resume la Verosimilitud que una muestra tiene respecto a una población normalmente distribuida con parámetros μ y σ . Valores de L_i mayores (ceranos a cero) indican mayor probabilidad relativa. El valor de Log-Verosimilitud es utilizado para estimar el ajuste de una muestra a unos parámetros poblacionales definidos. Si estos valores poblacionales no son conocidos, lo que suele ser el escenario más común, entonces los parámetros deben ser estimados a partir de la información que se tiene. El proceso de identificar dichos parámetros requiere intentar obtener los valores de Log-Verosimilitud a partir de distintos valores y determinar los parámetros que logren el mayor valor. A estos parámetros se los conoce como Estimaciones de Máxima Verosimilitud.

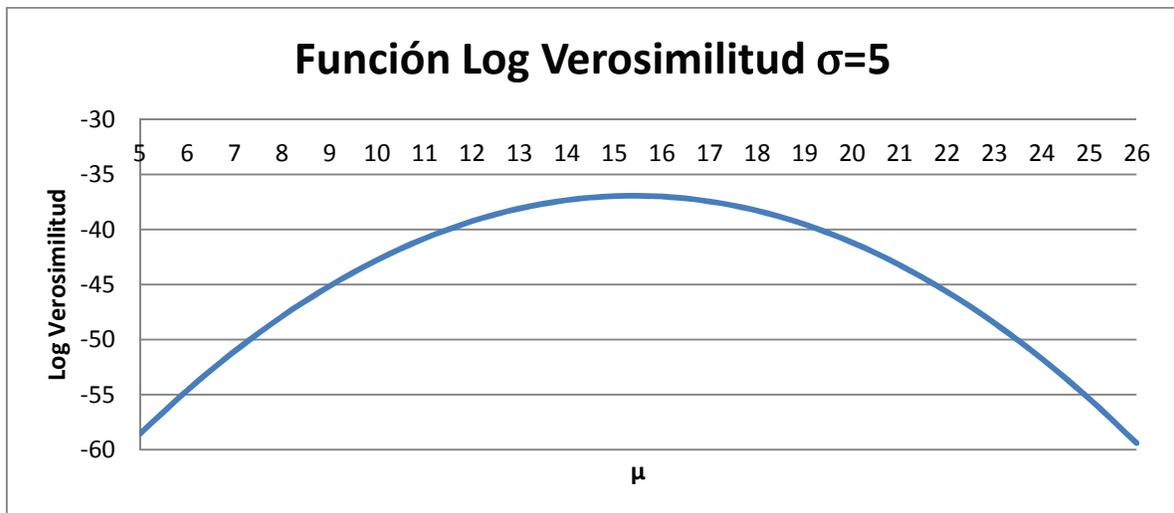


Gráfico 1 Función de Log-Verosimilitud

Al intentar estimar el parámetro μ suponiendo un valor para σ , se realiza un proceso iterativo asumiendo distintos valores x . Lógicamente, dado que μ es un estadístico de centralidad, al intentar el valor x más cercano a este se obtendrá la Log-Verosimilitud mayor y al intentar para μ el valor $x+1$ su Log-Verosimilitud será menor que la de x . Por conclusión, el valor x de mayor Log-Verosimilitud será el de mayor probabilidad relativa con respecto a μ . Esto permite deducir que existe una relación entre los parámetros poblacionales y el valor de Log-Verosimilitud, la cual se conoce como función de Log-Verosimilitud con la forma del gráfico 1.

Modelos de Escalamiento de Variables Latentes

Muchos constructos que son de interés para los científicos sociales no pueden ser observados directamente. Entre los ejemplos se encuentran las preferencias, actitudes, intenciones, los rasgos de personalidad y la calidad de vida. Estos constructos solo pueden ser medidos indirectamente a través de indicadores observables como ítems diseñados para generar respuestas relacionadas con algún rasgo particular. Diferentes tipos de técnicas de escalamiento han sido diseñadas para derivar información sobre constructos no observables a partir de los indicadores. Uno de estos grupos de métodos de escalamiento son los modelos de variable latente.

Un modelo de variable latente es un modelo, posiblemente no lineal, gráfico o de "Path Analysis". En adición a las variables manifiestas, el modelo incluye una o más variables latentes que representan los constructos de interés. Estas se relacionan a partir

de dos supuestos que definen los mecanismos causales que subyacen a las respuestas. Primero, se asume que las respuestas a los indicadores son resultado de la posición de un individuo sobre la variable latente; Segundo, se supone que la variables manifiestas no tienen otra relación además de controlar la(s) variable(s) latente(s), esto se conoce como el axioma de la independencia local (Bartholomew & Knott, 1999a; Birnbaum, 1968; Heinen, 1996).

De igual forma, dos supuestos hacen referencia a las distribuciones de las variables latentes y manifiestas. Dependiendo de estos supuestos, se obtienen diferentes clases de modelos de variable latente. Siguiendo a Bartholomew y Knott (1999) las cuatro clases principales son³: Análisis Factorial (FA), Análisis de Rasgos Latentes (LTA), Análisis de Perfil Latente (LPA) y Análisis de Clase Latente (LCA). Estos modelos son resultado de las posibles combinaciones de supuestos sobre las distribuciones de variables latentes.

En los análisis factoriales y de rasgo las variables latentes se asumen continuas y normalmente distribuidas. En los análisis de perfil y de clase, las variables latentes se asumen discretas y distribuidas bajo la forma Multinomial. Por otra parte, las variables manifiestas en los análisis factoriales y de rasgo son continuas y con distribución normal; en los análisis de perfil y de clase, los indicadores son dicotómicos, ordinales o nominales y sus distribuciones binomiales o multinomiales.

Evidentemente, la distinción fundamental de la tipología de Bartholomew radica en la clasificación de las variables latentes y manifiestas como continuas o categóricas. Es el investigador quien debe decidir si es más apropiado tratar la variable latente como continua o discreta. Sin embargo, como lo ha demostrado Heinen (1996) la distribución de un modelo de variable latente continua puede ser aproximada por un modelo de distribución discreta. Esto evidencia que la distinción entre variable latente continua o categórica es menos fundamental de lo que aparenta. De igual forma, la distinción entre modelos continuos o discretos para las variables manifiestas resulta no ser fundamental. La especificación de la distribución de las variables manifiestas es condicional a su nivel de escala. Los desarrollos recientes en modelamiento de variables latentes permiten diferentes formas de distribución para cada indicador. Estas pueden ser normal, lognormal, Gamma o exponencial para

³ Siglas en inglés.

variables continuas; o binomial (variables dicotómicas), multinomial (variables nominales y ordinales), Poisson o binomial negativa (Recuentos) en el caso de las variables discretas.

El modelo más ampliamente utilizado por la psicología en general y por la psicometría en particular es el de Análisis Factorial. En 1904, Charles Spearman publicó dos artículos que tuvieron una importante influencia en la psicología y la psicometría. En el primero (Spearman, 1904), bosquejó el modelo clásico de confiabilidad basado en la puntuación verdadera al dividir el desempeño en la prueba, y por ende la varianza en las puntuaciones al test, en dos partes: una porción que se debe al nivel “real” del individuo en el rasgo medido; y otra porción que se debe al error aleatorio de medición. Esto se representa mediante la ecuación:

$$x_{ij} = v_i + e_{ij} \quad | (8)$$

La puntuación observada de un individuo i en una variable X en el momento j se compone (es igual a) de la puntuación verdadera del sujeto i en X (V_i) más el error en la medición de la persona i en el momento j (E_{ij}). Si la puntuación verdadera es constante, entonces las variaciones en el desempeño de una persona en sucesivas mediciones se debe a errores que fluctúan de forma aleatoria. Esta distinción entre los componentes de la puntuación observada sentó las bases de la Teoría Clásica de la Medición y es hasta la actualidad altamente influyente.

Al aplicar la ecuación (8) a un conjunto de puntuaciones de una muestra con N individuos y obtener la varianza, se puede dividir la varianza de las puntuaciones observadas en dos componentes: varianza de las puntuaciones verdaderas y varianza de los errores de medición:

$$\sigma_x^2 = \sigma_v^2 + \sigma_e^2 \quad | (9)$$

La razón varianza de puntuaciones verdaderas sobre varianza total conforma el coeficiente de confiabilidad y la raíz cuadrada de la varianza de error es el error de medición, el cual puede ser utilizado para determinar el intervalo de confianza para una puntuación estimada.

En el segundo artículo de 1904, Spearman (2010), enunció una de las teorías de mayor influencia sobre las habilidades cognitivas humanas del siglo XX, la teoría de la inteligencia general, sentando con ello las bases del método de análisis de datos conocido

como Análisis de Factor. El autor dividió las puntuaciones de una persona a una variable observada, en una porción que representa lo que la variable tiene en común con otras variables que conforman el análisis, lo que llamó factor “g” o inteligencia general; y una porción que era exclusiva a la variable en cuestión, lo que llamó factor “s” o desempeño específico:

$$x_{ij} = g_i + s_{ij} \quad |(10)$$

La puntuación de un sujeto i a la variable j (X_{ij}) se compone de la puntuación del sujeto i en la variable de habilidad general (G_i) más la porción de puntuación en la variable habilidad específica (S_{ij}). Según la lógica usada en (9) para un conjunto de datos, en (10) la varianza de una variable puede descomponerse en una porción que se debe a un factor común y otra porción que se debe a un factor específico:

$$\sigma_x^2 = \sigma_g^2 + \sigma_s^2 \quad |(11)$$

Ya que hay dos fuentes, o factores, que contribuyen a la varianza de cada variable, esta teoría se conoce como la teoría de dos factores.

Un poco antes, Karl Pearson (Pearson, 1901) había derivado lo que él llamó Componente Principal de un conjunto de variables para acoger la mayor cantidad de varianza que podía ser explicada por una sola dimensión del conjunto. Posteriormente (Hotelling, 1933) lo generaliza al conjunto completo de componentes principales. Ya que este procedimiento no estaba asociado a ninguna teoría psicológica y era muy demandante para su cómputo, no fue utilizado por la psicología en ese momento (Nunnally et al., 1995).

Hay una importante diferencia ontológica entre el análisis de componentes como lo conciben Pearson y Hotelling y el análisis de factores como lo concibe Spearman. El primero es un procedimiento de reducción de datos. El análisis resulta en una representación espacial ortogonal (no correlacionada) de las variables, pero no implica nada acerca de constructos que subyazcan a las variables. El análisis de factores, por otra parte, es un método para descubrir constructos significativos subyacentes a las variables manifiestas entre las cuales hay una relación causal que explica las correlaciones entre las variables. Algunos autores como Velicer & Jackson (1990) consideran esta distinción innecesaria. En parte esto es cierto, ya que ambos procedimientos derivan en descripciones similares de los datos. Sin embargo, el análisis de factores generalmente arroja mejores resultados en

términos de calidad de la solución(Bartholomew & Knott, 1999b; Birnbaum, 1968; Heinen, 1996).

El planteamiento de Spearman sobre un único factor de habilidad cognitiva llevó a un debate con Thorndike y colaboradores (1909), Thorndike argumentaba que habían varios factores de inteligencia. En respuesta a esta crítica, Spearman se vio forzado a desarrollar un método analítico que soportara su planteamiento de que un único factor era suficiente para explicar las correlaciones en un conjunto de tests. Hart y Spearman (1912) logran demostrar que la condición suficiente para la existencia de un único factor era que una ecuación de la forma:

$$r_{ab} \cdot r_{cd} - r_{ac} \cdot r_{bd} = 0 \quad (12)$$

fuera cierta para todos los conjuntos posibles de cuatro tests. Este criterio conocido como la Ecuación de las diferencias tetraédricas, no podría ser cumplido con exactitud, para todos los posibles conjuntos de cuatro tests, a partir de datos reales, pero podría ser aproximado.

El debate acerca de la naturaleza de la inteligencia (unidimensional o no) continuó ya que los partidarios mostraban conjuntos de datos que satisfacían el criterio y los opositores mostraban datos que no lo cumplían. En 1917, Burt publicó un método para extraer factores de una matriz de correlaciones, el cual se aproximaba al método de componentes principales de Pearson, pero que era más fácil de calcular (Ballard, 1918). Ya que este método ubicaba el primer factor a través del promedio o centro geométrico del conjunto de variables, se conoce como el método centroide para extracción de factores, determinar la ubicación inicial de un factor se conoce como extracción del factor (Clayton, 1993).

Thurstone utiliza este método y lo convierte en un estándar hasta casi 1960, época en la que los computadores comienzan a ser más accesibles(Embretson, 2000). Por su parte Spearman en 1927 continúa con el criterio tetraédrico mostrando evidencia sobre una única dimensión de la inteligencia (Spearman, 2008). Por su parte Kelley (1928), utilizando el método de las correlaciones parciales para remover “g” de la matriz de correlaciones de un conjunto de variables, muestra cómo se pueden encontrar factores significativos en la matriz de correlaciones residuales. Kelley argumentaba que la distribución de las correlaciones residuales, posteriores a la extracción de “g” podrían ser usadas para probar y rechazar la hipótesis de un único factor de inteligencia general y que por ende, era un

importante objetivo de la medición psicológica la construcción de tests que sean medidas puras (específicas) de los múltiples factores que el encontró. Un poco antes, Thomson, (1920) había propuesto una aproximación muestral para la conceptualización de factores que resultaban en factores lógicos y jerárquicos dependiendo de la amplitud de la muestra. El concepto de jerarquía fue desarrollado explícitamente por Vernon (1950) y Humphreys (1962) en teorías acerca de la organización de las habilidades humanas.

Thurstone, (1947), reconoció que la satisfacción del criterio tetraédrico (ecuación 12) para cualquier conjunto de cuatro variables era equivalente a plantear que el rango de una matriz de correlaciones 4×4 es igual a 1. En este postulado, Thurstone argumenta que el rango de una matriz es el equivalente al número de factores que esta requiere para explicar las correlaciones. A menos que el rango de la matriz fuera 1, se requerirían más de un factor para reproducir las correlaciones. Además, el autor demostró como el método centroide podía ser usado para extraer factores sucesivos de forma más simple que con el método de correlaciones parciales de Kelley.

En un trabajo conocido como el Estudio de las Habilidades Mentales Primarias, Thurstone (1947, 1936, 2011) identifica doce factores de los cuales siete cumplen los criterios para ser considerados constructos hipotéticos. A partir de este trabajo Thurstone desarrolla una interpretación geométrica de los factores como ejes de un espacio multidimensional definido por las variables. Esta nueva interpretación lo llevó a reconocer que la ubicación de un factor es arbitraria, una vez se define el espacio multidimensional (según el rango de la matriz de correlaciones). Los factores centroides o componentes principales son utilizados para definir los ejes no-cero del espacio satisfaciendo ciertas condiciones, pero estos factores iniciales rara vez son significativos. Thurstone argumentó que se debían mover los ejes a nuevas posiciones que tuvieran mayor significado psicológico, este proceso actualmente se conoce como Rotación de Factores (Clayton, 1993).

En este trabajó Thurstone roto los factores mediante un criterio rígido que mantuvo los factores ortogonales. En 1938, realiza una rotación oblicua lo cual implica, geoméricamente, que los factores pueden asumir posiciones diferentes a 90 grados unos respecto a los otros. Vernon (1950) y Humphreys (1962) aplican posteriormente el análisis de factores a las matrices de correlaciones de los factores de primer orden para obtener sus modelos jerárquicos.

Los descubrimientos de Thurstone crearon tres preguntas importantes. Primero, el rango de cualquier matriz de correlaciones siempre sería igual al número de variables ya que las entradas diagonales de las matrices no incluían únicamente la varianza común (como la varianza de “g” en el modelo bifactorial de Spearman), sino que también incluyen la varianza específica. Ante esto, el autor sugirió que la matriz de correlaciones a ser explicada por los factores, no debería ser la matriz original sino una que contenga la estimación de la varianza común. Segundo, el problema de cómo estimar esta varianza común, hoy conocida como comunalidades, está relacionado con la pregunta de cuantos factores se requerían para explicar las correlaciones. Thurstone propone asumir como comunalidad el valor más alto de correlación que una variable tuviera con cualquier otra variable. Por su parte, Roff (1936) planteó que el cuadrado de la correlación múltiple de cada variable con las demás variables de la matriz proveía la mejor estimación de la comunalidad, esta es la forma actual de estimar las comunalidades iniciales. El tercer problema se relacionaba con el procedimiento de rotación, ya que una vez que se rotaba, había múltiples soluciones igualmente válidas. Thurstone intentó resolver este problema introduciendo el concepto de Estructura Simple.

Sin entrar en detalle, el principio de estructura simple plantea que cada variable observada debería componerse de la mínima cantidad de factores posible, idealmente uno. Thurstone (2011) plantea cinco criterios que un patrón de cargas factoriales debería cumplir para satisfacer el principio de estructura simple, sin embargo, el foco de atención se ha dirigido a encontrar rotaciones que produzcan una pequeña cantidad de cargas no-cero para cualquier variable.

Lo que sucede de este momento en adelante es ampliamente conocido y difundido (Nunnally et al., 1995). Lo importante de retomar de los inicios de la medición psicométrica, que se dan intentando responder a la necesidad de medir la inteligencia, es la búsqueda de ciertas características métricas en lo psicométrico: Unidimensionalidad, Independencia local, Parsimonia y Equiparación. Sin embargo, los métodos creados por los pioneros (como el análisis factorial) no respondían siempre con estas características a los datos que eran analizados, en gran parte porque los análisis propuestos lo que buscaban era construir un modelo que explicase los datos.

El siguiente apartado, desarrolla brevemente la propuesta de análisis del modelo de Rasch, una forma de ver la medición desarrollada a partir de los años 50 (unos 15 años después de los trabajos seminales de la psicometría clásica) que deja lo procedimental y

técnico a un lado para esclarecer las preguntas fundamentales. Unidimensionalidad, Independencia local, Parsimonia y Equiparación mediante el establecimiento de una “plantilla” de medición a la cual los datos deben ajustarse.

Modelo de Rasch

En 1952, el matemático Danés Georg Rasch (1901-1980) siendo consultor de proyectos en el ministerio de Asuntos Sociales, introdujo el Modelo Multiplicativo de Poisson para el análisis de errores de lectura en niños escolares. El consideró que el número de errores en los que incurría un examinado S_v en el texto I_i como una variable Poisson con parámetro λ_{vi} ⁴, midiendo la propensión del examinado al error cuando lee cierto texto. Posteriormente dividió λ_{vi} en un factor perteneciente al examinado, la habilidad de S_v para leer denominada Θ_v ; y un factor perteneciente al texto, la dificultad del texto I_i denominada δ_i . En consecuencia con el teorema de las variables Poisson, llego a la conclusión de que si un examinado lee dos textos I_i e I_j , la probabilidad de observar los errores K_{vi} y K_{vj} en estos textos, condicionada al total de la suma de los errores $K_{vi} + K_{vj}$, debía seguir una distribución Binomial caracterizada por K_v y el parámetro $\pi = 1/(1 + \frac{\delta_j}{\delta_i})$.

Esto llevo a Rasch a una nueva aproximación a la medición en las ciencias del comportamiento. Ya que el parámetro π , no dependía del parámetro Θ_v del examinado, los datos de todos los examinados que habían leído los dos mismos textos podían ser agrupados para obtener una estimación (máxima verosimilitud)⁵ de π . A partir de esta estimación, en retorno, se podía obtener una estimación del cociente de las dificultades de los textos $\frac{\delta_j}{\delta_i}$. Esto permitió una medición de las dificultades relativas de los textos, independientemente de la muestra de examinados, y esto era posible, a partir de un modelo extensible a más de dos textos. Rasch consideró tales comparaciones de las dificultades de los textos como “objetivas” a razón de la generalizabilidad de los resultados por sobre una muestra particular examinada, independientemente de la que pudiese ser la habilidad individual para leer de quienes la componen. Rasch (1977) plantea funciones que hacen de tales comparaciones “comparadores” factibles, y a las comparaciones realizadas de esta manera “específicamente objetivas”

⁴ Ver apartado de Conceptos estadísticos y matemáticos básicos. Poisson.

⁵ Ver Estimación por Máxima Verosimilitud.

En 1953, cuando Rasch analizó datos de mediciones de inteligencia del ejército danés, decidió aplicar el mismo principio en el área de análisis de tests. En primer lugar, planteó una Función de Respuesta al Ítem (FRI) adecuada $P(+|S_v, I_i) = f(\xi_v, \epsilon_i)$, donde + representa una respuesta correcta, ξ_v la habilidad del examinado y ϵ_i la facilidad del ítem del test I_i . Como una función algebraica particular para los valores reales positivos en el intervalo semiabierto (0,1) propuso: $f = x/(1 + x)$ y planteó a x como el producto de la habilidad del examinado ξ_v y la facilidad del ítem ϵ_i ; $x = \xi_v \epsilon_i$, con valores para ambos parámetros ≥ 0 . Este modelo general se conoce como “Modelo de Rasch”, pero comúnmente es reparametrizado tomando los logaritmos de la siguiente forma: $\theta_v = Ln(\xi_v)$ como la habilidad del examinado, y $\beta_i = Ln(\epsilon_i)$ como la dificultad del ítem.

Principios básicos del Modelo de Rasch

El Modelo de Rasch (MR) para respuestas dicotómicas (+,-) como resultado de la función de probabilidad cuando se asume la logística se define como:

$$P(X_{vi} = 1 | \theta_v, \beta_i) = \frac{e^{(\theta_v - \beta_i)}}{1 + e^{(\theta_v - \beta_i)}} \quad (13)$$

Donde X_{vi} es la respuesta a la variable con resultado $x_{vi} = 1$, si el examinado S_v da una respuesta “+” a I_i , y $x_{vi} = 0$ si S_v responde “-” a I_i ; $-\infty < \theta_v < \infty$ es la habilidad latente del examinado; y $-\infty < \beta_i < \infty$ es la dificultad del ítem. Todas las respuestas a los ítems se asumen localmente independientes, esto es, que la probabilidad de patrón de respuesta en un test de longitud k , es el producto de k probabilidades (13) o complementos del mismo. Los parámetros θ_v y β_i son definidos únicamente hasta una constante de normalización aditiva arbitraria c , esta última usualmente se especifica ya sea configurando un parámetro a cero (ej: $\beta_1 = 0$) o configurando $\sum \beta_i = 0$. La ecuación (13) implica que todas las funciones de respuesta al ítem (FRI) son curvas paralelas. Esto significa que si el Modelo de Rasch ajusta a un conjunto de datos, entonces todos los ítems deben tener (aproximadamente) igual discriminación (Bond & Fox, 2001).

La probabilidad de una matriz ($n \times k$) completa de puntuación al ítem X como una función de parámetros desconocidos (Función de Verosimilitud) se puede mostrar dependiente únicamente de las sumas marginales de X , es decir, la puntuación directa $r_v = \sum_i x_{vi}$; y de los marginales de los ítems $x_i = \sum_v x_{vi}$. Esto implica que r_v y x_i contienen toda la información relevante en los datos con respecto a los parámetros (conjuntamente son

estadísticos suficientes⁶), mientras que los patrones de respuesta individual no dan información adicional acerca de los parámetros.

Otra característica importante del MR es que la probabilidad condicional (Verosimilitud) de una matriz de puntuaciones a los ítems X , dadas unas puntuaciones directas, es una función de los parámetros de los ítems exclusivamente. La probabilidad condicional puede, de esta forma, servir como comparador para los parámetros de los ítems, permitiendo comparaciones específicamente objetivas entre las dificultades de los ítems. Las comparaciones entre personas también serían posibles pero realizarlas directamente no es práctico. La forma de comparar personas supone primero estimar todos los parámetros de los ítems y luego estimar los parámetros de las personas considerando constantes los parámetros de los ítems (Bond & Fox, 2001; Karabatsos, 2001).

Estas dos características de suficiencia de las puntuaciones directas y de los marginales de los ítems; y comparabilidad específicamente objetiva entre ítems y entre examinados, son exclusivas del Modelo de Rasch. Esto, asumiendo independencia local de los ítems con FRI continuas y monótonas estrictas con límites inferiores a cero y a superiores uno. Dentro del marco de estas condiciones, surge el catálogo de Modelos de Rasch a partir de las siguientes suposiciones: Suficiencia de las puntuaciones directas para los parámetros de personas; Existencia de una Función de Verosimilitud que pueda servir como comparador específicamente objetivo para los ítems y las personas; Existencia de estadísticos suficientes para los parámetros de personas independientes de los parámetros de ítems; y, Ordenamiento estocástico consistente de los ítems (Karabatsos, 2001; Georg Rasch, 1980). El catálogo de Modelos Rasch se conforma de todos los modelos de la forma (13) en los cuales se remplacen los parámetros Θ_v y β_i por $a\theta v + c$ y $a\beta i + c$ respectivamente. En estas, c es la constante de normalización mencionada con anterioridad; y $a > 0$ es el parámetro inespecífico de discriminación, es decir, la pendiente máxima (paralela) de las FRI. Si $a = 1$ se produce el modelo de Rasch, no se debe olvidar que la especificación es arbitraria (Karabatsos, 2001).

Esto tiene importantes connotaciones respecto a lo que son las propiedades de un test. Si para un conjunto de datos X se ajusta el MR y se estiman adecuadamente los parámetros Θ y β ; y al ser estos susceptibles de transformaciones con parámetros arbitrarios $a > 0$ y c , se concluye que dichas escalas son escalas de intervalo, con una unidad

⁶ Ver estimación.

de medida común. Sin embargo, se debe tener en cuenta como la evidencia que lleva a estas conclusiones se basa en el supuesto de que existe una población de examinados y un denso universo de ítems, se asume que tanto los parámetros de ítem como de persona pueden variar de forma continua. Suponer esto para las personas no es problema, el aumento de la habilidad suele concebirse como continuo. Por su parte los parámetros de ítem se asumen ser k número de valores discretos específicos, lo que debilita la conclusión respecto a las propiedades de los MR, ya que esto lleva a que las habilidades solo sean medibles a nivel de escala métrica ordenada⁷. Esto es, una escala ordinal con algunas propiedades de intervalo a ciertos puntos del rango. Esta situación que pareciera ser una desventaja, le da una fortaleza particular al MR, obliga a pensar en un orden conceptual jerárquico para los ítems que permita saturar los diferentes niveles (puntos) de rasgo. La construcción de escalas desde el Modelo de Rasch debe ser un proceso de base conceptual (Bond & Fox, 2001; Wilson, 2005).

Estimación de Parámetros y determinación del Ajuste

Se han propuesto varios métodos para la estimación de parámetros β , algunos por Rasch (1960) y otros por autores posteriores (Bock & Aitkin, 1981; M. Linacre, 1994). Sin embargo, para el modelo que asume los estimadores como únicos, estadísticamente suficientes y con propiedades asintóticas, predominan dos aproximaciones. El método de Máxima Verosimilitud Condicional MVC, el cual maximiza la probabilidad condicional de X , dadas unas puntuaciones directas, en términos de parámetros de los ítems. Esta función de Verosimilitud es independiente de los parámetros de personas, y por ende, es una función que permite establecer comparaciones específicamente objetivas entre los ítems. En muestras finitas, las estimaciones MVC normalizadas son únicas si una gráfica específica está asociada con la matriz de X (Fischer & Molenaar, 1995); probar esta relación en la práctica se realiza a partir de herramientas estándar de la representación gráfica estadística. Si esta condición, necesaria y suficiente establecida por Pfanzagal (1994) se cumple para la distribución de los parámetros de persona en la población de examinados, los estimadores son consistentes y distribuidos asintóticamente y normalmente alrededor de los parámetros reales de β_i para un test de longitud k y $n = \infty$.

Los errores estándar de los estimados de β pueden ser extraídos de la matriz de información. El estimador no es eficiente, aunque la pérdida de información estadística que acarrea el condicionar sobre los datos brutos es muy pequeña. Eggen (2000) demuestra, a

⁷ Taxonomía de Torgerson.

partir del concepto de F-información, que la eficiencia en el uso de información del MVC, del MVC y del MVM (Máxima Verosimilitud Marginal) es superior al 93% y que cuando el test supera los 20 ítems es superior al 99%.

Otros de los procedimientos de estimación, además del MVC, son el MVM, utilizado por el software BILOG. El procedimiento de MVM tiene una versión no paramétrica. Sin embargo, al usar este método dentro del modelo Rasch, autores han demostrado que es la única forma de lograr identificación de los parámetros de los ítems asumiendo una distribución latente de θ desconocida (de Leeuw & Verhelst, 1986; Pfanzagl, 1994). Es decir, MVC y MVM no paramétrica son asintóticamente equivalentes, lo cual se convierte en un argumento más a favor del modelo Rasch.

Metodología Wolfe-Smith para validación de mediciones con los Modelos Rasch siguiendo la propuesta de Messick

Wolfe y Smith a partir de una revisión de metodologías y trabajos empíricos sobre validación desde los modelos de Rasch consolidan una serie de recomendaciones para el desarrollo de instrumentos (Wolfe & Smith, 2007a) y para la validación de las mediciones (Wolfe & Smith, 2007b). El WHOQOL-BREF es un instrumento desarrollado por el comité de calidad de vida de la Organización Mundial de la Salud, un grupo de expertos que le dan soporte a las actividades de validación a partir del desarrollo del instrumento sugeridas por Wolfe & Smith (2007a) y que no se retomarán en este estudio. El objetivo del presente proyecto se direcciona a las tareas de la validación de mediciones. A continuación se plantean los aspectos más relevantes del documento de Wolfe y Smith (2007b).

El propósito de los lineamientos propuestos por los autores es “identificar varios tipos de evidencia que resultan de ejecutar las actividades propuestas (para validación) y cómo esta evidencia puede ser utilizada para soportar argumentos de validez que respaldan las inferencias y decisiones realizadas a partir de las medidas que son generadas a partir de un instrumento.” (p. 244)⁸ Para esto los autores toman como referencia el trabajo de Messick (1995) “Validity of psychological Assessment: Validation of inferences from persons responses and performances as scientific inquiry into score meaning” quien propone como aspectos de la validez el contenido, lo sustantivo, lo estructural, la generalizabilidad, los criterios externos y lo consecuencial. De forma complementaria los autores toman como referencia los “Instrument Review Criteria” del Medical Outcomes Trust Scientific Advisory

⁸ Las especificaciones de página, que se hacen a lo largo del apartado, se refieren a (Wolfe & Smith, 2007b)

Committee (1995) los cuales sugieren considerar como criterios de validez la sensibilidad y la interpretabilidad.

La gráfica 1 es una adaptación al español de la original publicada en Wolfe y Smith (2007b), de los aspectos no subrayados, el comité y los centros colaboradores que desarrollaron los instrumentos han cubierto gran parte del trabajo relacionado con el contenido. El subrayado indica las evidencias de la validez susceptibles de ser verificadas en el presente estudio. Sobre estas evidencias para cada aspecto de la validez se hace una descripción a partir del trabajo de los autores.

<p style="text-align: center;">Contenido</p> <p>Propósito del instrumento Especificaciones del test Desarrollo de ítems Revisiones de expertos <u>Calidad técnica de los ítems</u></p> <p style="text-align: center;">Generalizabilidad</p> <p><u>Invarianza de la calibración de los ítems</u> Predicción diferencial Generalización de la validez <u>Confiabilidad</u></p> <p style="text-align: center;">Estructural</p> <p>Correlaciones entre subescalas <u>Análisis de dimensionalidad</u></p> <p style="text-align: center;">Interpretabilidad</p> <p><u>Mapas Ítems-Personas</u> <u>Desarrollo de la métrica</u></p>	<p style="text-align: center;">Sustantiva</p> <p><u>Modelamiento cognitivo</u> Observación comportamental <u>Distractores/Funcionamiento de la escala de valoración</u> <u>Ajuste de las personas</u> <u>Jerarquía de la dificultad de los ítems</u></p> <p style="text-align: center;">Externa</p> <p>Matrices multirasgo-multimetodo <u>Comparación de grupos</u> Cambios individuales Calibración de ítems entre grupos</p> <p style="text-align: center;">Consecuencial</p> <p>Metodo de la configuración estandar Calificación de evaluadores Calidad de las calificaciones Estudios de Clasificación Selección de criterios</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Gráfica 1 Evidencia de validez relevante para aspectos de validez (Adaptado de Wolfe y Smith, 2007b)

Contenido

El aspecto de la validez de contenido se refiere a “la relevancia y representatividad de los contenidos sobre los cuales los ítems se basan” (p.245); también hace referencia a la calidad técnica de los mismos.

El aspecto de contenido, depende mucho de la racionalidad que sustenta el instrumento y los criterios para la inclusión de ítems. Las tareas que Wolfe y Smith denominan Propósito del instrumento, Especificaciones del test, Desarrollo de ítems y Revisiones de expertos, fueron realizadas por el WHOQOL Group a partir de equipos internacionales de expertos y centros colaboradores (WHOQOL1995; WHO, 1998).

En relación a la calidad técnica de los ítems, desde los modelos Rasch se pueden establecer dos índices para su valoración: la correlación ítem-medida y los índices de medias cuadráticas para ajuste. La correlación ítem-medida es “la correlación de Pearson entre el vector de puntuaciones en un ítem particular y las medidas de los participantes que recibieron esas puntuaciones”(p. 246). Este índice indica la consistencia entre la puntuación al ítem y la puntuación a la escala total. Para ítems politómicos, como es el caso del WHOQOL-BREF, se esperan correlaciones positivas y mayores a .40 cuando se utilizan en calificación normativa.

Por su parte los índices de medias cuadráticas para ajuste expresan “discrepancias entre los valores esperados por el modelo y los valores observados en las personas”(p.247). Los inconvenientes en estos índices pueden reflejar problemas de multidimensionalidad o pobre calidad del ítem. Estos índices suelen reexpresarse en forma estandarizada para evitar problemas relacionados con tamaño de la muestra y del test, en su forma estandarizada los índices pueden tomar valores centrados en cero en el rango $-\infty$ a ∞ . Un valor mayor a cero “sugiere la presencia de grandes residuos, o acumulación de varios residuos moderados”(p.248), un valor menor a cero “indica menor variabilidad que la esperada por el modelo Rasch”(p.248).

Sustantiva

El aspecto sustantivo de la validez se refiere “al grado con el cual la justificación teórica, relacionada tanto con el contenido de los ítems como con los modelos de procesamiento cognitivo, explica la consistencia observada entre las respuestas a los ítems.”(p. 248).

Para este propósito los autores plantean evidencias de validez a partir del análisis del funcionamiento de la escala de valoración, el ajuste de las personas y la confirmación de la jerarquía de los ítems.

El funcionamiento de la escala de valoración puede dar algunas luces en relación al grado con el cual las personas responden al test siguiendo la intención de sus desarrolladores. El aumento progresivo de la media a lo largo de la clasificación jerárquica de las alternativas de respuesta genera información de mayor especificidad que la correlación ítem-medida descrita en la validez de contenido. El que se de este aumento progresivo, sugiere que la intención de los desarrolladores del test, se refleja adecuadamente en la forma en que las personas responden. Linacre (2004) propuso

algunos lineamientos para valorar a profundidad el funcionamiento de la escala de valoración, este se describe en detalle en la metodología.

El ajuste de las personas se da en relación a lo esperado por el modelo de Rasch y lo expresado por quienes responden a los ítems. El ajuste de las personas se calcula mediante índices de medias cuadráticas estandarizadas de forma similar a lo descrito en el apartado de contenido. Los problemas en el ajuste pueden sugerir “características no modeladas entre quienes responden”(p.255) y por ende no concebidas por los desarrolladores durante el proceso de construcción. Es posible que también se deba a sesgos de respuesta, sesgos en los ítems o no interés de parte de los participantes en contestar con seriedad.

La confirmación de la jerarquía de los ítems, es dentro de las evidencias de validez sustantiva la que tiene una mayor connotación teórica. Si el ordenamiento supuesto para los ítems, es decir el nivel de dificultad supuesto para los ítems, coincide con las dificultades calculadas para los ítems en el proceso de análisis, entonces esto puede usarse como argumento para sustentar la teoría que respalda el constructo subyacente a la medida que se intenta validar.

Estructural

El aspecto estructural de la validez “confronta la fidelidad de la estructura de puntuación con la estructura del dominio de constructo.”(p.256). Para este aspecto se pueden generar evidencias de validez a partir del análisis de correlaciones entre subescalas y del análisis de dimensionalidad.

El análisis de unidimensionalidad consiste en tres pasos los cuales son similares al análisis de componentes principales. En primer lugar se obtiene una escala que ajusta a los criterios de los modelos Rasch, lo cual es “equivalente al primer componente extraído de los datos con la restricción que todos los ítems cargan igualmente al componente”(p.257). El segundo paso consiste en obtener componentes que expliquen varianza a partir de los residuales estandarizados de la escala Rasch y el tercer paso consiste en valorar si estos componentes de los residuales tiene una capacidad explicativa que haga sospechar de la unidimensionalidad de la medida. Para esto se toma como criterio los autovalores los cuales son índices que expresan si un componente tiene mayor capacidad explicativa que un ítem. De forma genérica se asumen valores de autovalores mayores a 1 - el criterio de Kaiser o se grafican los autovalores para determinar si se

incumple el supuesto de comportamiento asintótico de los residuales (p.258). La metodología describe el criterio usado para este estudio.

Generalizabilidad

El aspecto de generalizabilidad de la validez se refiere al “grado con el cual las medidas mantienen su significado a través de contextos de medición. Las evidencias para este aspecto de la validez toman la forma de correlaciones entre medidas generadas en distintos contextos, sin embargo la generalizabilidad como aspecto de la validez trasciende las concepciones de confiabilidad.”(p.260).

La Invarianza de la calibración de los ítems evalúa el grado con el cual “los ítems mantienen su significado e interpretabilidad a través de varios contextos.”(p.261). La técnica para analizar esto se conoce como Funcionamiento diferencial de los Ítems y determina el grado en que un ítem falla en mantener su nivel de dificultad entre grupos o contextos. Es decir, en el caso de que se igualen las mediciones al constructo de dos grupos y un ítem no muestre igual probabilidad de arrojar respuestas particulares entre estos, entonces el ítem falla en el supuesto de invarianza y su interpretabilidad entre contextos no puede ser garantizada. Existen varias formas de hacer el análisis de funcionamiento diferencial. En el apartado de metodología se describe la forma como se realiza este análisis en este estudio.

La confiabilidad hace referencia a “comparaciones entre mediciones a través de variaciones menos extremas de contextos de medición.”(p.263). Existen diversas formas de calcular la confiabilidad a las cuales se asocian inferencias posibles de realizar. A todas ellas es común el considerarse como límites a las inferencias sobre generalizabilidad que se pretenden realizar. La consistencia interna, la forma a estimar en este estudio, “toma en consideración el error de medición que se ha introducido por variaciones en el contenido o calidad de los ítems debidas al muestreo de los ítems, pero lo hace sin la necesidad de crear múltiples formas de un instrumento.”(p.264).

Para el modelo de Rasch, “el coeficiente de consistencia se designa como la varianza de las medidas latentes de las personas y la varianza de las estimaciones de las medidas de las personas, $(V(\text{latente})/V(\text{observada}))$, referido como confiabilidad de la separación de las personas. Dado que la varianza de las medidas latentes de las personas no puede ser observada, este índice se estima como $1 - (V(\text{error})/V(\text{observada}))$, el cual es estadísticamente equivalente a $(V(\text{latente})/V(\text{observada}))$ y donde la varianza de error es

igual al promedio de los errores cuadrados asociados con la estimación de las medidas de las personas.”(p.264)

Externa

Este aspecto de validez es considerado el de mayor importancia y es el más cercano a lo que comúnmente se ha llamado “Validez de criterio”. La validez externa se refiere al “grado con el cual las medidas están relacionadas con medidas externas del mismo constructo, constructos similares u otros constructos.”(p.268).

Las comparaciones entre grupos contribuyen a argumentar la validez externa al generar evidencia sobre las experiencias de individuos pertenecientes a diferentes grupos en los cuales se espera por razones teóricas o prácticas que las puntuaciones difieran significativamente. Caso contrario, si dos grupos de los cuales se tiene una medición externa relacionada con el constructo bajo evaluación, y se conoce que esta medición externa es diferente entre ellos, no generan diferencias en la medida en proceso de validación, los argumentos de validez externa no son posibles de sustentar.

Consecuencial

Este aspecto de la validez “se enfoca en las implicaciones de valor de la interpretación de puntuaciones como fuente para la acción.”(p.273). Esta evidencia permite inferir las implicaciones potenciales del uso de las medidas. Este aspecto de la validez no puede ser determinado en el estudio y por ende no se profundiza en él.

Interpretabilidad

El aspecto de interpretabilidad “se enfoca en el grado con el cual el significado de las mediciones es claramente comunicado a aquellos que quieren interpretar las mediciones.”(p.279). La interpretabilidad se puede generar a partir de métodos normativos y referenciales. No hay procedimientos formales para evaluar la interpretabilidad de las medidas. En el contexto del análisis Rasch se puede recurrir a la interpretación basada en la interacción entre la escala de las personas y la escala de los ítems, sacando provecho a la característica de expresión de estas dos escalas en la misma unidad (logitos) y su definición probabilística conjunta. Esta forma permite una aproximación a la interpretación basada en criterio al permitir describir los ítems que se espera que una persona pueda cumplir (aquellos con dificultades iguales o menores a lo que su nivel de habilidad puede equiparar). Igualmente es posible obtener valores tipificados por variables sociodemográficas para una aproximación normativa a la interpretación.

Tanto la aproximación normativa como la referencial pueden ser analizadas a partir de los mapas de ítems-personas o mapas de Wright. Estos mapas dan información útil para comprender el targeting de un conjunto de ítems a una muestra de personas. Sobre estos mapas se pueden hacer análisis de interpretabilidad relacionando habilidad de los sujetos con dificultad de los ítems.

El concepto de calidad de vida

Calidad de vida es un término común, para el cual no hay una definición universal. La calidad de vida ha sido históricamente estudiada desde las ciencias sociales, donde la literatura reconoce cinco grandes aproximaciones. Sociológica, Económica, Psicológica, Filosófica y Ética (Taylor, Gibson, & Franck, 2008). Sólo hasta después de la segunda guerra mundial, las áreas disciplinares relacionadas con el cuidado de la salud, la seguridad social y la salud pública han cobrado interés en la calidad de vida, desde el reconocimiento de su fuerte relación con el proceso de salud-enfermedad; y sólo hasta 1976 el término se agrega al “Index Medicus”.

La ambigüedad en la definición y en su diferenciación con la salud se refleja en trabajos documentales que intentar cumplir este propósito que datan de los noventa (Kleinpell, 1991) hasta años recientes (Condon, 2013; Plummer & Molzahn, 2009). Aún hoy no se dan consensos gremiales, científico, ni disciplinares. Lo usual es asumir la definición de la Organización Mundial de la Salud, práctica muy discutida en los círculos académicos de forma similar a como sucede con la definición de salud (Haas, 1999).

Los expertos reconocen como primer trabajo publicado sobre calidad de vida, el reporte “Americans View Their Mental Health” liderado por Gurin (1975) y realizado por un equipo de la universidad de Michigan. Esta investigación fue ordenada por una comisión de Salud y Enfermedad Mental del Congreso de los Estados Unidos con el propósito de “Medir la salud mental subjetiva de los adultos americanos ‘normales’ y determinar en detalle como afrontaban los problemas de ajuste” tanto psicológicos como económicos. Si bien esta investigación no contempla la “salud física”, es el primer ejercicio formal de averiguar cómo interactúa la salud (mental) con las condiciones percibidas de vida.

Los autores en general se cuestionan por la dificultad para diferenciar la calidad de vida de la salud, pero las indagaciones sobre la diferencia calidad de vida – salud mental no son comunes. Se esperaría lo contrario, ya que los grandes hitos que marcan la introducción de la calidad de vida en el foco de interés del cuidado de la salud están

asociados a la salud mental: el reporte de Gurin y otros (1957) y el comité de calidad de vida de la Organización Mundial de la Salud, el cual hace parte de la división de Salud Mental de esta Organización. De igual forma, los análisis a las cuatro dimensiones del WHOQOL-BREF muestran una relación de imagen entre la dimensión de salud psicológica y la calificación total de calidad de vida (Rojas Gualdrón, 2012; World Health Organization, 1998).

Estos antecedentes que se cuestionan por la relación entre la enfermedad mental y la percepción de los recursos económicos y psicológicos con los que se cuenta para afrontarla abren la vía para llevar esta preocupación a las condiciones de enfermedad física. Más aún cuando los cambios en los paradigmas del cuidado (Plummer & Molzahn, 2009) posibilitan el reconocimiento de aspectos positivos del bienestar entre las personas enfermas (Taylor et al., 2008); y los debates éticos propenden por el mejoramiento de sus condiciones de satisfacción y bienestar más allá del sometimiento a detrimentos de estas condiciones en el intento de combatir la enfermedad o prolongar la vida (Bodek, 2013; Johns, 2013; Kaasa & Loge, 2003).

Perspectivas sociales de conceptualización de la calidad de vida

En los periodos tempranos de la historia, la perspectiva filosófica de la calidad de vida se enfoca en el reconocimiento de la naturaleza de la existencia humana, para a partir de allí recomendar las pautas para llevar una buena vida (Eiser & Morse, 2001; Mandzuk & McMillan, 2005). Por su parte, la perspectiva ética, hija de la perspectiva filosófica, redirecciona esta visión de la buena vida hacia una búsqueda de la santidad de la vida (Anderson & Burckhardt, 1999). El análisis de lo planteado por el autor da pie a considerar este planteamiento más como una perspectiva moral que ética.

La consolidación de las ciencias sociales y del desarrollo capitalista de la sociedad le dan un cambio paradigmático a la definición de la calidad de vida. Surge la perspectiva económica que asume la calidad de vida como la medición del crecimiento económico, el aumento del producto interno bruto y el mejoramiento de la esperanza de vida ((Anderson & Burckhardt, 1999; Eiser & Morse, 2001)). Paralelamente, se va desarrollando la perspectiva sociológica que conceptualiza la calidad de vida como la relación entre las circunstancias sociales y la cultura (Eiser & Morse, 2001).

Posteriormente, se consolida la perspectiva psicológica de la calidad de vida que la plantea como la valoración personal tanto de la vida, como del cumplimiento de los objetivos

vitales((Condon, 2013; Eiser & Morse, 2001; Plummer & Molzahn, 2009; Taylor et al., 2008)). Esta perspectiva psicológica, le da sustento a la forma como más tarde la Organización Mundial de la Salud conceptualizará la calidad de vida.

Perspectivas de conceptualización de calidad de vida relacionada con la salud

En una revisión sistemática de la literatura, Taillefer y otros (2003) encuentran 68 modelos de calidad de vida relacionada con la salud publicados entre 1965 y 2001. Según los criterios de valoración de los autores que incluían aspectos como nivel de formalización, distinción entre calidad de vida y factores que la afectan, y la presencia de instrumentos para medirla, solo el 15% de esos modelos fueron considerados modelos sofisticados; el 75% eran ambiguos y el 25% no daba si quiera una definición de la calidad de vida.

Este trabajo revela la falta de consenso para referirse a la calidad de vida relacionada con la salud. Entre otros resultados, los autores logran clasificar estos 68 modelos en conceptuales, marcos conceptuales y marcos teóricos. Estos autores consideran un modelo conceptual como “un modelo que especifica las dimensiones y propiedades del concepto de calidad de vida”. Un marco conceptual se define como “un modelo que describe, explica y/o predice la naturaleza o direccionalidad entre los elementos o dimensiones de la calidad de vida”. Un marco teórico se define como “un modelo que incluye la estructura de los elementos y sus relaciones bajo una teoría que explica estas relaciones”.

El único modelo considerado como marco teórico en la revisión sistemática fue el de Shye (1989). Este autor en su artículo “The systemic life quality model: A basis for urban renewal evaluation” plantea un marco teórico para sustentar las relaciones entre los elementos supuestos que influyen y constituyen la calidad de vida. Empíricamente el autor prueba su modelo de forma jerárquica mediante escalogramas. La gráfica 1 muestra este marco con ejemplos de ítems propuestos por el autor.

TABLE III
Relative QOL coefficients in IG neighborhood in each of the sixteen systemic life quality components (a heuristic description illustrates the significance of each mode)

Subsystem: Mode	Personality	Physical	Social	Cultural
Expressive	Self actualization 1.00	Physical activity 0.90	Social influence 0.85	Cultural-educational activity 0.90
Adaptive	Personal recreation 0.65	Physical conditions 0.71	Institutional roles 0.84	Cultural compatibility w. environment 0.85
Integrative	Peace of mind 0.90	Physical health 0.96	Intimate friendship 0.68	Integrity of values held 0.91
Conservative	Self confidence 0.90	Physical security 1.14	Social confidence 0.91	Stable structure of beliefs 0.92
Total QOL			0.79	

Gráfica 2 Modelo sistémico de la calidad de vida, tomado de Shye (1989)

Según el autor, los modos de funcionamiento hacen referencia a:

Expresivo: las características del sistema se reflejan en sus actividades. Emergen dentro del sistema y se actualizan fuera del sistema.

Adaptativo: se dan relaciones compensatorias entre el sistema y su entorno. Emergen y se actualizan fuera del sistema.

Integrativo: relaciones de complementación entre los componentes del sistema. Emergen y se actualizan dentro del sistema.

Conservativa: las características del sistema son estables a través del tiempo. Emergen fuera del sistema y se actualizan dentro del sistema.

Este marco teórico avalado por la revisión sistemática se empleará para discutir la estructura del WHOQOL-BREF.

El concepto de calidad de vida subyacente a los instrumentos WHOQOL de la Organización Mundial de la Salud

La Organización Mundial de la Salud define la salud como “el estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades”(World Health Organization, 1946). De esta definición se deduce que la medición de la salud y los efectos de la atención en salud debe trascender los indicadores de frecuencia y severidad de enfermedad, para considerar la medición del bienestar. Una

forma de aproximarse al bienestar es mediante la medición de la calidad de vida. La OMS define la calidad de vida como “la percepción individual de la propia posición en la vida dentro del contexto del sistema cultural y de valores en que se vive y en relación con sus objetivos, esperanzas, normas y preocupaciones”(WHOQOL, 1995).

En el reconocimiento de estas consideraciones la Organización Mundial de la Salud conforma un comité “The WHOQOL Group” dentro de la división de Salud Mental. El objetivo de este comité fue desarrollar un instrumento internacional de medición de la calidad de vida que permitiese a la organización “reafirmar su compromiso con la promoción de un enfoque holístico de la salud y el cuidado de la salud” (WHOQOL, 1995). A este propósito el comité añade “El cuidado de la salud es esencialmente una transacción humanista entre un profesional de la salud y un paciente, donde el bienestar del paciente es el principal fin” (WHOQOL, 1995).

Para este propósito el comité inicia con 15 centros colaborativos a nivel mundial una primera fase de clarificación del concepto. Entre los resultados de esta fase descritos en WHO (1995; 1997, 1998), se destaca el reconocimiento de que la calidad de vida es subjetiva. Siguiendo los desarrollos teóricos vigentes a la época el comité reconoce que la calidad de vida incluye la percepción de aspectos objetivos (recursos materiales) y subjetivos (satisfacción con los recursos) y lleva este planteamiento a consideraciones sobre el cómo medirlo. De esta forma, se llega al consenso de considerar tres tipos de preguntas(WHOQOL, 1995):

1. Información sobre el funcionamiento: ¿Cuántas horas durmió anoche?
2. Evaluación global del funcionamiento: ¿Qué tan bien duerme?
3. Valoraciones personalizadas del funcionamiento: ¿Qué tan satisfecho está con su sueño?

De estos niveles de preguntas, como los llama el comité, prevalece en importancia relativa el conjunto de preguntas que indagan sobre las “evaluaciones globales de las personas sobre sus comportamientos, estados y capacidades, al igual que su satisfacción/insatisfacción con estos.” (WHOQOL, 1995).

WHOQOL-100		WWHOQOL-BREF
Dominio	Facetas	Dominio
	Calidad de vida global y Salud general	
Salud física	Energía y fatiga Dolor y Disconfort Sueño y descanso	Salud física
Nivel de Independencia	Movilidad Actividades de la vida cotidiana Dependencia a medicamentos y soporte médico Capacidad para trabajar	
Salud psicológica	Imagen corporal y apariencia Sentimientos negativos Sentimientos positivos Autoestima Pensamiento, Aprendizaje, Memoria y Concentración	Salud mental
Espiritualidad	Religión/Espiritualidad/Creencias personales	
Relaciones sociales	Relaciones sociales Soporte social Actividad sexual	Relaciones sociales
Entorno	Recursos financieros Libertad, seguridad física y seguridad Servicios sociales y de salud: Accesibilidad y calidad Ambiente del hogar Oportunidades para adquirir nueva información y habilidades Participación y oportunidades para recreación/placer Ambiente físico (Polución,/Ruido/Trafico/Clima) Transporte	Entorno

Gráfica 3 Dimensiones y facetas del WHOQOL-100 y equivalencia con WHOQOL-BREF

El otro gran resultado del consenso, establece el carácter multidimensional de la calidad de vida (WHO 1995; 1997, 1998). De esta forma el comité genera el cuestionario WHOQOL-100 originalmente pensado a partir de seis grandes dimensiones, las cuales a su vez se componen de facetas (gráfica 3). A partir de esta forma general de 100 ítems el comité genera una forma, corta el WHOQOL-BREF, la cual se compone de dos preguntas de valoración global de la salud y la calidad de vida y de 24 ítems de forma tal que se incluya uno por cada faceta del WHOQOL-100. Para la calificación de la forma corta y en base a análisis realizados sobre el WHOQOL-100 y en las pruebas del BREF se decide agrupar las dimensiones de Salud física y nivel de funcionamiento, al igual que las dimensiones de Salud psicológica y Espiritualidad (WHO, 1997).

Definiciones extensas para cada faceta del WHOQOL-100 se pueden encontrar en el apéndice 1 de WHO (1998 pp. 57-69). Estas facetas en el WHOQOL-BREF se convierten en el contenido de los ítems que cargan las cuatro dimensiones de la estructura del cuestionario. El comité no genera definiciones para estas dimensiones.

Es importante reconocer que la conceptualización de la calidad de vida que respalda tanto al WHOQOL-100 como al WHOQOL-BREF, es el producto del trabajo empírico de comités y centros colaboradores de expertos internacionales. Ninguno de los documentos originales discute el modelo conceptual jerárquico de las facetas a partir de las cuales se mide la calidad de vida. Esta información sería de alta importancia para robustecer la interpretabilidad de las medidas de calidad de vida de la Organización Mundial de la Salud. Posterior a los trabajos originales se han realizado análisis de las propiedades psicométricas de ambos instrumentos en diversas adaptaciones y traducciones de ambos cuestionarios, la mayoría de estos trabajos desde la Teoría Clásica de los Test, en menor medida desde la Teoría de la Respuesta al Ítem y muy pocos desde los modelos de Rasch. Estos últimos trabajos se relacionan a continuación en los antecedentes.

Antecedentes

La búsqueda de antecedentes⁹ se hizo en las bases de datos PubMed, EBSCO Psychology and Behavioral Science Collection, SCOPUS y SpringerLink. Los instrumentos de la Organización Mundial de la Salud para medir la calidad de vida, el WHOQOL-100 y WHOQOL-BREF han sido ampliamente utilizados alrededor del mundo en diferentes poblaciones. Al buscar en SCOPUS con la palabra “WHOQOL” se obtienen 1.568 documentos. Por otro lado, al buscar específicamente por “WHOQOL-BREF” se obtienen 1.097 documentos. Al reducir la búsqueda a estudios psicométricos agregando la palabra “Psychometric” se encuentran 109 documentos. Sin embargo al utilizar como criterio de búsqueda “WHOQOL-BREF RASCH” solo se obtienen nueve documentos, de los cuales seis documentos son pertinentes a análisis psicométricos.

Entre estos seis documentos, el trabajo de Krägeloh y colaboradores (2013) no está disponible para consulta completa dado que será publicado en la siguiente edición. El resumen reporta que los resultados fueron favorables después de solucionar problemas con el ordenamiento de categorías, dependencia local y funcionamiento diferencial de los Ítems.

Se presentan en la tabla 1 los valores de dificultad de los estudios disponibles para cada ítem del WHOQOL-BREF. Las descripciones de muestras son las siguientes:

⁹ A agosto de 2013

Tabla 1 Valores de dificultad para los ítems del WHOQOL-BEF reportados en los antecedentes

	Rocha y otros, 2012 Depresivos	Liang y otros, 2009 Mayor 65 años	Wang y otros, 2006 Adultos 20 - 65
Salud física			
Actividades cotidianas	0.5	0.21	-0.1
Dependencia a medicamentos	-0.61	-0.6	-0.77
Energía y fatiga	0.28	0.48	0.61
Movilidad	-0.85	0.42	---
Dolor y disconfort	-0.43	-1.1	-0.45
Sueño y descanso	0.75	-1.6	0.61
Capacidad para trabajar	0.35	0.42	0.09
Salud psicológica			
Imagen corporal y apariencia	-0.39	-0.52	-0.57
Sentimientos negativos	0.44	-1.19	-0.03
Sentimientos positivos	0.3	1.29	1.32
Autoestima	0.16	-0.61	-0.54
Espiritualidad	0.3	0.37	-0.21
Cognición	-0.21	0.67	0.03
Relaciones sociales			
Relaciones personales	0.06	-0.71	-0.22
Soporte social	-0.33	0.31	-0.39
Actividad sexual	0.27	1.17	---
Entorno			
Recursos financieros	0.59	0.94	---
Seguridad física	0.17	-0.63	0.43
Seguridad social y en salud	-0.44	0.01	-0.27
Ambiente del hogar	-0.05	-0.56	0.08
Acceso a información	-0.54	0.29	---
Actividades de ocio y placer	0.51	0.96	---
Ambiente físico	-0.01	-0.63	---
Transporte	-0.23	-0.03	-0.25

Rocha y otros, 2012: 1193 pacientes confirmados para depresión (71% mujeres, edad media 40.3 SD 14.5) miembros del estudio LIDO (Longitudinal Investigation of Depression Outcomes) provenientes de Australia, Brasil, Israel, Rusia, España y Estados Unidos.

Rocha y Almeida, 2009: 208 pacientes confirmados para depresión miembros del estudio LIDO de nacionalidad brasilera. Estos pacientes son incluidos en el estudio posterior de Rocha y otros (2012) razón por la cual no serán nuevamente reportados en la tabla.

Liang y otros, 2009: 1200 adultos mayores de 65 años taiwaneses 59% hombres.

Wang y otros, 2006: 13.083 adultos taiwaneses entre los 20 y 65 años.

Noerholm y otros, 2004: 1101 adultos (52% mujeres) entre los 20 y 60 años. Reportan no obtener ajuste al modelo de Rasch y presentan resultados con TRI de dos parámetros (no se muestran en la tabla).

Esta revisión demuestra que ha sido poco lo que se ha estudiado del WHOQOL-BREF mediante análisis Rasch.

Método

Tipo y diseño de la investigación

Investigación cuantitativa. Diseño explicativo no experimental de corte transversal. Se busca mediante el Análisis Rasch explicar los trazos latentes a partir de las respuestas observadas en los ítems.

Población

Jóvenes adultos y adultos medios santandereanos con condiciones de salud favorables establecidas a partir de su juicio subjetivo. El foco de atención en esta investigación es el constructo, no la población. Por esta razón no hay más criterios definitorios.

Sujetos

Los participantes fueron obtenidos durante una actividad universitaria de promoción y prevención realizada en el 2012 en la ciudad de San Gil. Se les tomó consentimiento informado y contestaron un protocolo con datos sociodemográficos y cuestionarios sobre salud, incluido el WHOQOL-BREF.

Muestreo

No probabilístico por conveniencia. El análisis de Rasch, como lo explica el capítulo 3, consiste en la estimación por máxima verosimilitud de los parámetros de los ítems para asumirlos como invariantes y poder hacer estimaciones sobre las personas. Estas estimaciones se hacen no para las variables observadas sino para los hipotetizados trazos latentes. Esto supone determinar poblaciones (dificultad ítems o habilidad personas) como una tarea imposible e inútil dado el rango posible $-\infty; \infty$, de forma que se excluye el muestreo por tamaño de la población. La otra opción a la que se puede recurrir es el muestro por tamaño del efecto, en este caso el error tipo I (rechazar erróneamente un ítem que se muestra sin ajuste) se exagera para los indicadores basados en t siendo más estables los indicadores basados en medias cuadráticas (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008).

Tamaño de la muestra

No hay consenso ni procedimientos establecidos para el cálculo del tamaño de la muestra en estudios de validación de constructo. Sin embargo, en función de las consideraciones expuestas en el muestreo se puede retomar el trabajo de Linacre, (2002) donde recomienda al menos 10 respuestas por categoría dentro de las opciones de

respuesta politómicas, estas consideraciones son extrapolables al Modelo de Andrich, pero según el autor (1994) no al Modelo de Crédito Parcial de Masters, en función de los cambios en los grados de libertad. Sin embargo Smith, 2008 y Linacre 2002, 1994 coinciden en considerar que para las observaciones politómicas, el tamaño de la muestra real podría ser menor que en variables dicotómicas dada la mayor riqueza de información de las primeras.

Así las cosas al seguir la guía de Linacre, 1994 para lograr estabilidad contundente en las calibraciones de los ítems (dicotómicos) se requeriría de una muestra entre 250 y 20x número de ítems. El trabajo de Smith, 2008 muestra que en tamaños cercanos a este valor no se da tanta discrepancia entre los indicadores basados en t y los indicadores basados en medias cuadráticas.

Los subgrupos de interés para el proyecto son género y cohortes por lo cual el tamaño final de la muestra debe ser de aproximadamente 750 a 800 datos.

Variables e Instrumentos

WHOQOL-BREF. Consiste en 26 preguntas que constituyen cuatro dominios: Salud física, Salud psicológica, Relaciones sociales y Entorno, y dos preguntas no calificadas que indagan sobre la calidad de vida y satisfacción con la salud. El instrumento tiene la siguiente estructura:

Salud física	Actividades cotidianas Dependencia medicamentos y sustancias medicas Energía y fatiga Movilidad Dolor e incomodidad Sueño y descanso Capacidad para trabajar
Psicológica	Imagen corporal Sentimientos negativos Sentimientos positivos Autoestima Espiritualidad Pensamiento, memoria, aprendizaje y concentración.
Relaciones sociales	Relaciones interpersonales

Entorno	Apoyo social
	Actividad sexual
	Recursos económicos
	Libertad y seguridad
	Acceso y calidad del servicio social y en salud
	Entorno familiar
	Acceso a nueva información y habilidades
	Recreación
	Ambiente físico
	Transporte

Análisis de datos.

Los cálculos y tablas se realizaron mediante el software WINSTEPS 3.75.1 (J. Linacre, 2012b). Para el análisis de funcionamiento diferencial se utilizaron procedimientos escritos para STATA por el profesor Luis Carlos Orozco Vargas. Estos fueron implementados en STATA versión 12. Se ejecutaron los siguientes análisis:

Adecuación de las categorías de valoración de los ítems.

La valoración de las escalas de valoración de cinco puntos de los ítems del WHOQOL-BREF se determinó mediante el cumplimiento de los tres criterios de optimización de categorías de Linacre (2002). 1) al menos diez observaciones por categoría; 2) las medidas promedio de las categorías aumentan a lo largo de la escala; y 3) el valor de Outfit Mean Square MNSQ debe oscilar entre el rango .5-2.0.

El orden y calibración de las categorías (threshold) se analizó para establecer el aumento monótonico. Si el aumento monótonico no se cumple, esto indica que algunas de las categorías no son la respuesta más probable a ciertos puntos de rasgo latente (Bond & Fox, 2001). En estos casos las categorías deben ser colapsadas.

Ajuste de los ítems

Dado el cumplimiento de las categorías dentro de la escala, se estableció la posibilidad de usar los ítems para construir una medida unidimensional. Los estadísticos de ajuste del ítem y el análisis de componentes principales sobre los residuos se emplearon para determinar la unidimensionalidad.

Los estadísticos de ajuste analizan la compatibilidad de los datos brutos con los requerimientos del modelo Rasch. En general se reportan el Outfit mencionado

anteriormente y el Infit. Los Infit son sensibles a comportamientos inesperados que afectan las respuestas a los ítems cercanos al nivel de “habilidad” exhibido por los individuos; por su parte los Outfit son sensibles a comportamientos inesperados que afectan las respuestas a los ítems lejanos al nivel de “habilidad” exhibido por los individuos (Bond & Fox, 2001). Para ambos se mantiene el rango .5-2 en MNSQ. A diferencia del Outfit, un ítem con valores no esperados de Infit muestra sobre ajuste al modelo y se considera redundante.

Unidimensionalidad

Una vez se garantizaron los estadísticos de ajuste de los ítems, se ejecutaron análisis de componentes principales sobre los residuos (puntuaciones observadas – puntuaciones esperadas). Al cumplirse los criterios de ajuste de ítems se espera que los residuos se distribuyan aleatoriamente (M. Linacre, 1994; Smith et al., 2008). Para esto se estableció como criterio que las medidas explicarán la mayor proporción de varianza, y que el primer contraste explicase menos de 3 autovalores-

Ajuste de las personas

El ajuste de personas se evaluó mediante sus estadísticos Infit y Outfit con los mismos criterios que el ajuste de ítems. Con el propósito de obtener estimaciones más precisas de la calibración de los ítems, se eliminaron los sujetos que no ajustaban antes del análisis de ajuste de ítems.

Funcionamiento Diferencia de los Ítems (DIF)

El DIF se evaluó para establecer si las calibraciones de los ítems eran estables entre géneros, presencia de enfermedad, rangos de edad y nivel de calidad de vida. Las estimaciones de dificultad de ítem por estos grupos se analizaron gráfica y matemáticamente. Los ítems que no caen dentro de las líneas de control se consideran DIF (Bond & Fox, 2001), en contraparte, los que si ajustan se consideran invariantes en relación con las variables analizadas. Los cálculos de tamaño del DIF y significación por Welch y Mantel Haenszel fueron realizados en winsteps y posteriormente exportados al software Stata versión 12. En este software, el DIF fue analizado mediante procedimientos de cálculo de valores críticos para significación por el procedimiento de Benjamini-Hochberg con el propósito de controlar los falsos positivos producto de las comparaciones múltiples (Thissen, Steinberg, & Kuang, 2002). De forma complementaria mediante el Stata se generaron los gráficos de límites de acuerdo (95% de confianza) de Bland y Altman, los cuales tienen como propósito realizar comparaciones sencillas de medidas clínicas en términos de acuerdo entre estas (Bland & Altman, 1986).

Confiabilidad de los ítems y las personas

La confiabilidad de los ítems indica la replicabilidad del ordenamiento de estos a lo largo de la escala de rasgo latente si dichos ítems se aplicarán a otra muestra de sujetos. De forma similar la confiabilidad de las personas indica la replicabilidad de las personas en función de su habilidad, si a las mismas personas le fuesen aplicados otros ítems que midan el mismo constructo. Los valores de confiabilidad de ítems y personas se muestran en un rango 0 a 1 y pueden ser interpretados de forma similar al Alfa de Cronbach y por lo tanto su interpretación se asume como consistencia interna.

Además se calculó el índice de separación (G_p) el cual indica que es posible separar la habilidad (nivel de calidad de vida) de las personas en diferentes estratos, entre mayor es el valor de G_p , mayor es el número de estratos. Para calcular el número de estratos se utilizó la fórmula $(4G_p+1)/3$ (Wright & Masters, 1982). Esta clasificación se utilizó para establecer si las personas podían ser separadas en estratos diferentes.

Un índice de separación también se calculó para los ítems (G_i) y se realizó similar procedimiento de estratificación.

Jerarquía de los ítems y Targeting

A partir de los ítems que cumplieron con los requisitos hasta el momento expuesto se construyeron escalas de intervalo. De forma gráfica se analizó la jerarquía de los ítems utilizando el método propuesto por Bond y Fox (2001). Finalmente se muestra la relación entre dificultades de los ítems y habilidades de las personas mediante mapas ítem-persona para analizar el targeting del rango de habilidades de la muestra en relación al WHOQOL-BREF.

Asociación entre valoración global de la salud y la medida del WHOQOL-BREF

Para los dos ítems no calificables del WHOQOL-BREF: satisfacción con la salud y percepción de calidad de vida, se realizaron análisis de varianza y comparaciones múltiples mediante el método de Bonferroni. En estos análisis se esperaron diferencias significativas en el promedio de la medida rasch entre las categorías semánticas de respuesta a los dos ítems que aumentasen de forma lógica según la jerarquía de respuesta.

Comparación con las dificultades reportadas por los antecedentes

Mediante las dificultades y errores estándar obtenidos en la muestra estudiada y los reportados por los antecedentes descritos en la tabla 1, se construyeron gráficas de Bland

y Altman (95% de confianza) para establecer diferencias entre contextos de evaluación con el WHOQOL-BREF en términos de acuerdo.

Resultados

Descripción de la muestra

La muestra estuvo conformada por un total de 944 personas. Para los análisis se optó por incluir los casos que hubiesen contestado a las 24 preguntas del WHOQOL-BREF, para un total de 792 casos analizados (84% valido según lista). La tabla 1 muestra las características de los casos incluidos y los compara, mediante prueba t y chi-cuadrado con los casos excluidos, no se encuentran diferencias significativas que sugieran un sesgo en que explique la no respuesta.

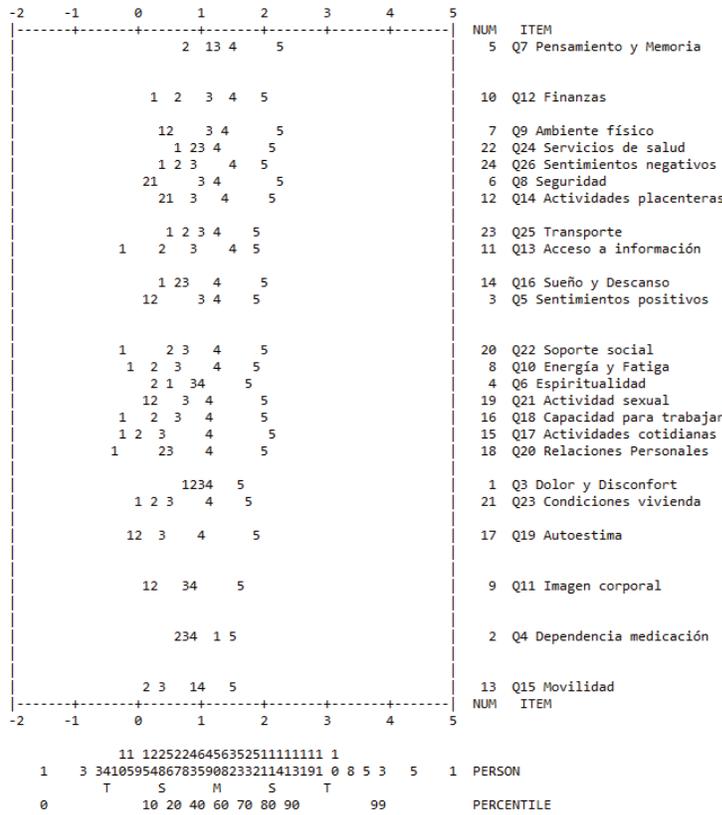
Tabla 2 Caracterización de la muestra de análisis y comparación con los casos excluidos

	Incluido		Excluido		p.
	Media	SD	Media	SD	
Edad	21.71	6.382	20.97	6.014	0.15
Nivel de calidad de vida	4.14	.709	4.05	.832	0.20
Satisfacción con la salud	3.87	1.066	3.70	1.149	0.08
Sexo	%		%		
Femenino	61.0%		58.7%		0.55
Masculino	39.0%		41.3%		
Enfermo?	%		%		
Si	12.7%		15.9%		0.49
No	70.0%		66.5%		
No esta seguro	17.3%		17.6%		

Adecuación de las categorías de valoración.

Cinco categorías de respuesta

Al analizar las medidas empíricas ítem-categoría se observa que la escala de 5 puntos para los ítems 7. Pensamiento y Memoria, 8. Seguridad, 14. Actividades placenteras, 6. Espiritualidad, 4. Dependencia a medicación y 15. Movilidad están desordenadas y no se logra en estos ítems el supuesto del instrumento (tabla 2). De igual forma en la mayoría de los ítems la distancia en logitos entre las categorías no es relevante (gráfica 4), con algunas excepciones como 12. Finanzas, 13. Acceso a la información, 10. Energía y fatiga y 18. Capacidad para trabajar.



Gráfica 4 Medidas empíricas ítem-categoría escala total 5 puntos

En relación al análisis de estructura de categorías (tabla 3) se identifica que no se cumple el criterio de mínimo 10 respuestas por categoría para todos los ítems¹⁰. De igual forma los valores de coherencia de las categorías 1, y 3 no alcanzan para M->C y C->M un valor de al menos 40%. La categoría 1 muestra un valor MNSQ OUTFIT mayor a 2.

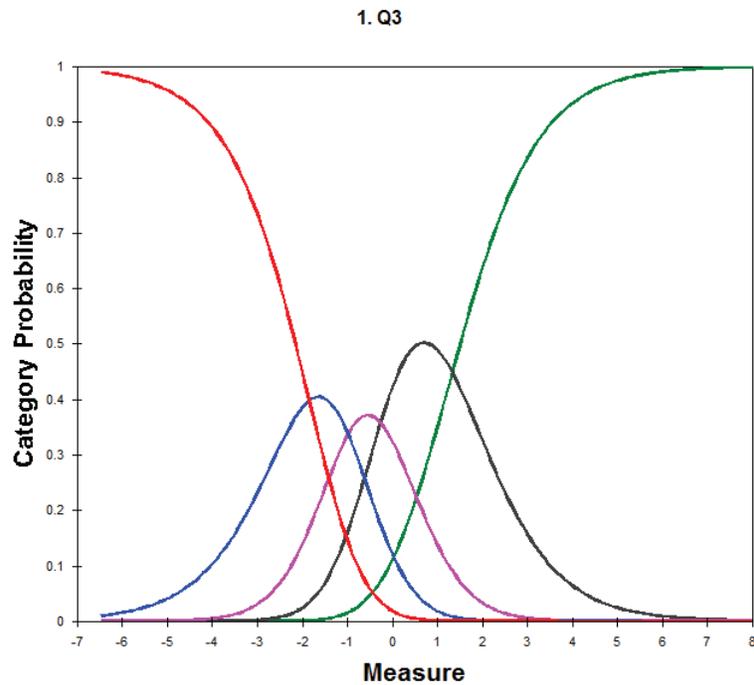
Tabla 3 Estructura de categorías escala total 5 puntos

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INFINIT AVRGE	OUTFIT EXPECT	ANDRICH MNSQ	CATEGORY THRESHOLD	COHERENCE MEASURE	M->C	C->M	
1	1	339	2	.20	-.28	1.46	2.18	NONE	(-2.75)	33%	1%
2	2	1327	7	.25	.19	1.10	1.25	-1.41	-1.19	41%	12%
3	3	3496	18	.62	.67	.90	.86	-.54	-.10	37%	37%
4	4	7524	40	1.15	1.24	.96	.83	.18	1.15	49%	75%
5	5	6322	33	2.08	2.00	.93	.95	1.77	(3.01)	79%	44%

En consideración con los análisis expuestos sobre la escala total a partir de 5 categorías se concluye que estas deben ser colapsadas para mejorar el funcionamiento de la escala, los análisis de categoría ítem y estructura de categorías sugieren colapsar las

¹⁰ Por extensión, no se reportan las tablas por cada ítem.

categorías 1, 2 y 3 en una única categoría. Se realiza este proceso y se reinicia el análisis pasando la escala del formato 1 2 3 4 5 al formato 1 1 1 2 3.



Gráfica 5 Curva de probabilidad de 5 categorías Ítem 3

Tres categorías de respuesta

Con la nueva categorización de la escala total se observa ordenamiento entre estas y distancia relevante en logitos (gráfica 6), con excepción de los ítems 3. Dolor y Disconfort y 4. Dependencia medicación.

OBSERVED AVERAGE MEASURES FOR PERSON (unscored) (BY OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	ITEM
				1	2	3			5	Q7 Pensamiento y Memoria
					1	2	3		10	Q12 Finanzas
				1	2	3			7	Q9 Ambiente físico
				1	2	3			24	Q26 Sentimientos negativos
				1	2	3			6	Q8 Seguridad
				1	2	3			12	Q14 Actividades placenteras
				1	2	3			22	Q24 Servicios de salud
				1	2	3			11	Q13 Acceso a información
				1	2	3			23	Q25 Transporte
				1	2	3			3	Q5 Sentimientos positivos
				1	2	3			14	Q16 Sueño y Descanso
				1	2	3			20	Q22 Soporte social
				1	2	3			8	Q10 Energía y Fatiga
				1	2	3			19	Q21 Actividad sexual
				1	2	3			4	Q6 Espiritualidad
				1	2	3			15	Q17 Actividades cotidianas
				1	2	3			16	Q18 Capacidad para trabajar
				1	2	3			18	Q20 Relaciones Personales
				1	2	3			21	Q23 Condiciones vivienda
				1	2	3			17	Q19 Autoestima
				12	3				1	Q3 Dolor y Discomfort
				1	2	3			9	Q11 Imagen corporal
				12	3				2	Q4 Dependencia medicación
				1	2	3			13	Q15 Movilidad

1112141353625272251111111111

1 5 2 6 3274011155730365998378760171832 7 6 5 7 1 PERSON

0 T S M S T PERCENTILE

10 20 30 50 70 80 90 99

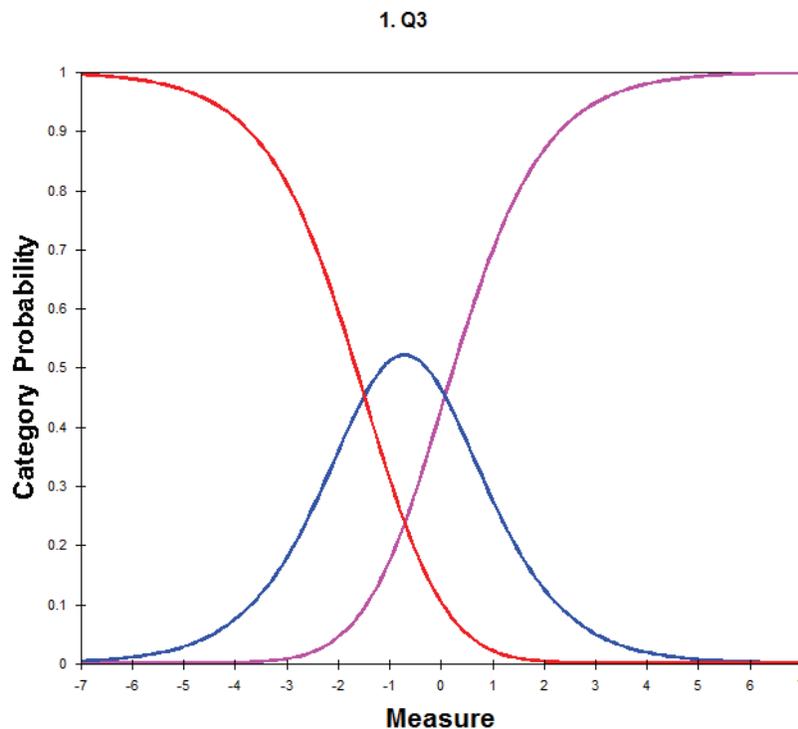
Gráfica 6 Medidas empíricas ítem-categoría escala total 3 puntos

En términos de la estructura de categorías (tabla 4) se logra el mínimo de observaciones por categoría en los ítems, se mejora el ajuste de las categorías con valores MNSQ OUTFIT e INFIT cercanos a 1, se obtiene aumento mono tónico de la medida entre categorías y la coherencia M->C y C->M es superior al 40%.

Tabla 4 Estructura de categorías escala total 3 puntos

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	AVRGE	EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	COHERENCE M->C	C->M
1	1	5162	27	-.85	-.90	1.08	1.17	NONE	(-2.02)	68%	37%
2	2	7524	40	.00	.08	.92	.86	-.78	.00	49%	79%
3	3	6322	33	1.23	1.18	.95	.98	.78	(2.02)	77%	51%

En conclusión el uso de tres categorías en vez de cinco mejora el ajuste de los ítems y de las categorías beneficiando la escala total



Gráfica 7 Curva de probabilidad de 3 categorías Ítem 3

Ajuste de los ítems

Al analizar la escala total con las categorías colapsadas (tabla 5), se obtiene una leve mejora en las correlaciones de los ítems 3. Dolor y Disconfort, 7. Pensamiento y memoria y 4. Dependencia medicación. El ajuste de los ítems mejora en el INFIT y el OUTFIT los cuales pasan un SD de .34 y .35 al usar cinco categorías a un SD de .22 y .24 al colapsar a tres categorías. No se observan razones para eliminar ítems de la escala.

Tabla 5 Polaridad de los ítems Escala total 3 puntos

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		ITEM
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
1	1852	792	-.71	.06	1.34	6.8	1.50	7.8	.34	.52	50.9	57.4	Q3 Dolor y Disconfort	
5	1241	792	1.34	.06	1.09	1.9	1.21	3.3	.38	.53	56.0	60.6	Q7 Pensamiento y Memoria	
2	2027	792	-1.39	.07	1.32	5.6	1.45	5.4	.39	.47	65.7	65.8	Q4 Dependencia medicación	
13	2067	792	-1.58	.07	.99	-.1	1.04	.6	.43	.46	71.0	68.7	Q15 Movilidad	
22	1428	792	.67	.06	1.10	2.2	1.11	2.1	.47	.55	54.2	55.5	Q24 Servicios de salud	
23	1538	792	.32	.06	1.19	4.3	1.19	3.8	.47	.55	52.3	54.3	Q25 Transporte	
10	1326	792	1.02	.06	1.16	3.3	1.15	2.7	.47	.54	56.6	57.6	Q12 Finanzas	
24	1390	792	.80	.06	.94	-1.5	1.00	.0	.48	.54	58.5	55.9	Q26 Sentimientos negativos	
7	1372	792	.86	.06	.97	-.7	1.03	.6	.50	.54	55.9	56.7	Q9 Ambiente físico	
11	1464	792	.55	.06	1.17	3.7	1.13	2.6	.52	.55	50.3	54.3	Q13 Acceso a información	
3	1548	792	.28	.06	1.15	3.5	1.14	2.8	.52	.55	48.3	54.3	Q5 Sentimientos positivos	
9	1963	792	-1.13	.06	1.17	3.4	1.15	2.2	.53	.49	63.8	62.1	Q11 Imagen corporal	
4	1706	792	-.22	.06	1.43	9.0	1.42	7.5	.53	.54	37.9	54.2	Q6 Espiritualidad	
20	1629	792	.03	.06	.89	-2.8	.91	-1.9	.56	.54	58.4	54.1	Q22 Soporte social	
6	1399	792	.77	.06	1.02	.6	1.00	.1	.56	.54	52.5	55.9	Q8 Seguridad	
14	1610	792	.09	.06	.98	-.5	.96	-.9	.56	.54	55.2	54.0	Q16 Sueño y Descanso	
19	1701	792	-.20	.06	.90	-2.4	.89	-2.4	.59	.54	61.4	54.2	Q21 Actividad sexual	
21	1813	792	-.58	.06	.82	-4.3	.81	-3.7	.59	.52	63.0	56.5	Q23 Condiciones vivienda	
12	1412	792	.73	.06	.99	-.3	.94	-1.2	.60	.55	56.3	55.2	Q14 Actividades placenteras	
8	1678	792	-.13	.06	.90	-2.5	.89	-2.3	.62	.54	56.4	54.2	Q10 Energía y Fatiga	
18	1734	792	-.31	.06	.72	-7.2	.72	-6.3	.64	.53	64.5	54.7	Q20 Relaciones Personales	
16	1716	792	-.25	.06	.67	-8.6	.66	-7.9	.65	.54	65.7	54.7	Q18 Capacidad para trabajar	
17	1850	792	-.70	.06	.68	-8.1	.65	-7.2	.68	.52	70.0	57.3	Q19 Autoestima	
15	1712	792	-.24	.06	.54	-9.9	.54	-9.9	.69	.54	72.8	54.9	Q17 Actividades cotidianas	
MEAN	1632.3	792.0	.00	.06	1.00	-.2	1.02	-.1			58.2	56.8		
S.D.	221.9	.0	.75	.00	.22	4.9	.24	4.5			7.8	3.8		

Unidimensionalidad de la Escala Total

El análisis de los residuales sobre los resultados del modelo Rasch de la escala total (tabla 6) muestran que se explica a partir de este el 38.8% de la varianza de las observaciones brutas. Al realizar este análisis sobre una base de datos simulada se observa que en este ideal escenario se explica el 39.9% de la varianza¹¹. Del porcentaje explicado en la base de datos real (tabla 6), el 16.6% se explica por la varianza de las personas y 22.2% por la varianza de los ítems. El primer contraste explica el 6.6% de la varianza (10.7% modelada) lo cual corresponde a 2.6 autovalores, esto levanta dudas respecto a la unidimensionalidad de la escala total, pero es posible aceptarla al no ser el porcentaje de varianza explicado por este contraste de alta relevancia; y al ser su autovalor menor a 3, lo cual sugiere baja capacidad para formar una segunda dimensión.

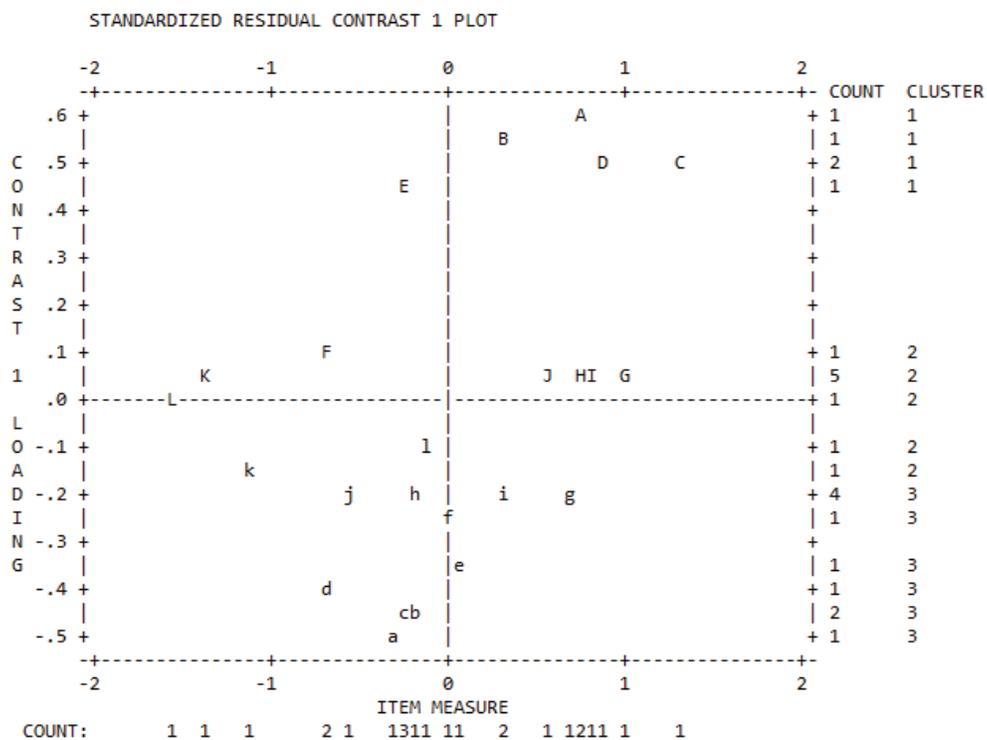
¹¹ El análisis no se incluye en las tablas

Tabla 6 Análisis de componentes principales escala total 3 puntos

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	39.2 100.0%	100.0%
Raw variance explained by measures	=	15.2 38.8%	38.8%
Raw variance explained by persons	=	6.5 16.6%	16.6%
Raw Variance explained by items	=	8.7 22.2%	22.2%
Raw unexplained variance (total)	=	24.0 61.2%	61.2%
Unexplned variance in 1st contrast	=	2.6 6.6%	10.7%
Unexplned variance in 2nd contrast	=	2.2 5.6%	9.1%
Unexplned variance in 3rd contrast	=	1.9 4.8%	7.8%
Unexplned variance in 4th contrast	=	1.6 4.0%	6.6%
Unexplned variance in 5th contrast	=	1.4 3.5%	5.8%

Este primer factor de contraste (gráfica 8) muestra 3 clúster de ítems, de los cuales evidencian independencia el 1 respecto a los clúster 2 y 3. Este primer clúster está constituido por los ítems (A) 8. Seguridad, (B) 5. Sentimientos positivos, (C) 7. Pensamiento y memoria, (D) 9. Ambiente físico y (D) 6. Espiritualidad. Estos ítems no guardan una fuerte relación conceptual entre sí por lo cual en análisis del residual reafirma la posibilidad de asumir la escala total como unidimensional.



Gráfica 8 Residuales estandarizados contraste 1 escala total 3 puntos

Las correlaciones obtenidas entre estos tres clúster en el primer contraste (tabla 7), respaldan la interpretación unidimensional para el total de la escala.

Tabla 7 Relaciones aproximadas entre las medidas de las personas contraste 1 escala total 3 puntos

PCA Contrast	ITEM Clusters	Pearson Correlation	Disattenuated Correlation	Pearson+Extr Correlation	Disattenuated+Extr Correlation
1	1 - 3	0.3962	0.5506	0.4008	0.5566
1	1 - 2	0.3989	0.5640	0.4046	0.5714
1	2 - 3	0.5948	0.7696	0.5988	0.7745

Ajuste de las personas

El análisis de ajuste de personas¹² arroja valores MNSQ INFIT con una media de 1.01 (SD .38) y valores MNSQ OUTFIT con una media de 1.02 (SD .48) lo cual indica que aproximadamente el 95% de las personas tiene valores INFIT entre .26 y 1.75 y OUTFIT entre .07 y 1.98. En ambos casos los valores son aceptables, según los criterios establecidos en la metodología. Respecto a un ajuste deseable entre .5 y 2 se logra en el 91% de la muestra para INFIT y en el 86% de la muestra para OUTFIT.

Funcionamiento Diferencial de los Ítems

Se presentan los análisis de funcionamiento diferencial de los ítems con los 24 ítems originales del WHOQOL-BREF. Se realizan los cálculos por sexo, percepción subjetiva de enfermedad, grupos de edad con corte en 30 años y nivel de calidad de vida alto bajo según puntuación del modelo Rasch para el WHOQOL-BREF. Se consideran como con DIF significativo aquellos ítems cuya diferencia en logitos es de al menos .5 y que su valor de significación sea menor al valor crítico reportado por el procedimiento de Benjamini-Hochberg (BH). Se realizan gráficas de Bland y Altman para corroborar los hallazgos mediante cálculos.

Funcionamiento Diferencial de los Ítems por sexo

La tabla 8 muestra los resultados del análisis de funcionamiento diferencial de los ítems por sexo. Considerando relevante un contraste de DIF mayor o igual a .5 logitos, sólo el ítem Sentimientos negativos logra este criterio. De igual forma se obtiene significación mediante las pruebas de -Welch y Mantel Haenszel los cuales son menor al valor crítico BH.

¹² Dada la extensión del output de este análisis no se reporta en el informe.

Los ítems Energía y Fatiga y Soporte social muestran significación por ambas pruebas menor al valor crítico BH. Sin embargo, el contraste DIF en ambos casos es menor a .5 con valores de .41 y .42 respectivamente.

Tabla 8 Funcionamiento diferencial de los ítems por sexo

Ítem	Mujeres			Hombres			Contraste DIF	E.E. conjunto	Rasch Welch			Mantel		BH crítico
	promedio Obs-Esp	Medida DIF	E.E. DIF	promedio Obs-Esp	Medida DIF	E.E. DIF			t	d.f.	p	Chi-squ	p	
Q3 Dolor y Discomfort	-0.04	-0.59	0.07	0.07	-0.93	0.1	0.34	0.13	2.7	660	0.008	6.4	0.011	0.004
Q4 Dependencia medicación	-0.02	-1.32	0.08	0.04	-1.54	0.12	0.22	0.14	1.6	657	0.115	0.5	0.466	0.017
Q5 Sentimientos positivos	-0.03	0.36	0.07	0.05	0.16	0.09	0.2	0.12	1.7	678	0.094	2.4	0.124	0.011
Q6 Espiritualidad	0	-0.22	0.07	0	-0.22	0.1	0	0.12	0.0	673	1.000	0.3	0.587	0.019
Q7 Pensamiento y Memoria	0.03	1.25	0.08	-0.05	1.48	0.1	-0.23	0.13	1.8	680	0.076	1.1	0.300	0.013
Q8 Seguridad	0.01	0.74	0.07	-0.02	0.81	0.1	-0.07	0.12	-0.6	681	0.549	0.9	0.342	0.015
Q9 Ambiente físico	0.01	0.82	0.07	-0.02	0.92	0.1	-0.11	0.12	-0.9	681	0.382	0.3	0.564	0.018
Q10 Energía y Fatiga	-0.06	0.01	0.07	0.1	-0.39	0.1	0.41	0.12	3.4	669	0.001	10.6	0.001	0.003
Q11 Imagen corporal	-0.04	-1.02	0.08	0.06	-1.33	0.11	0.32	0.14	2.3	655	0.020	3.7	0.054	0.008
Q12 Finanzas	0.01	1	0.08	-0.02	1.07	0.1	-0.08	0.12	-0.6	683	0.533	0.0	0.977	0.025
Q13 Acceso a información	0.01	0.52	0.07	-0.02	0.62	0.09	-0.1	0.12	-0.8	679	0.418	0.2	0.649	0.020
Q14 Actividades placenteras	-0.01	0.75	0.07	0.01	0.69	0.09	0.06	0.12	0.5	683	0.619	0.1	0.748	0.022
Q15 Movilidad	0.02	-1.67	0.09	-0.04	-1.42	0.11	-0.24	0.14	1.7	682	0.091	3.4	0.064	0.009
Q16 Sueño y Descanso	0.04	-0.02	0.07	-0.07	0.28	0.09	-0.3	0.12	2.6	677	0.010	5.8	0.016	0.005
Q17 Actividades cotidianas	0.01	-0.24	0.07	-0.01	-0.22	0.1	-0.03	0.12	-0.2	674	0.825	0.6	0.437	0.016
Q18 Capacidad para trabajar	0.01	-0.28	0.07	-0.01	-0.22	0.1	-0.06	0.12	-0.5	674	0.630	1.1	0.302	0.014
Q19 Autoestima	0	-0.7	0.07	0	-0.7	0.1	0	0.12	0.0	670	1.000	0.0	0.894	0.024
Q20 Relaciones Personales	0	-0.31	0.07	0	-0.31	0.1	0	0.12	0.0	673	1.000	0.0	0.853	0.023
Q21 Actividad sexual	0	-0.2	0.07	0	-0.2	0.1	0	0.12	0.0	674	1.000	0.2	0.693	0.021
Q22 Soporte social	0.06	-0.12	0.07	-0.1	0.3	0.09	-0.42	0.12	3.5	677	0.000	13.7	0.000	0.002
Q23 Condiciones vivienda	0.04	-0.68	0.07	-0.07	-0.37	0.1	-0.31	0.12	2.6	679	0.011	5.6	0.018	0.006
Q24 Servicios de salud	0.04	0.57	0.07	-0.07	0.84	0.1	-0.27	0.12	2.3	677	0.022	2.7	0.098	0.010
Q25 Transporte	-0.03	0.39	0.07	0.05	0.17	0.09	0.22	0.12	1.9	678	0.063	3.8	0.051	0.007
Q26 Sentimientos negativos	-0.07	0.99	0.08	0.12	0.49	0.09	0.5	0.12	4.1	691	0.000	22.6	0.000	0.001

Funcionamiento Diferencial de los Ítems por percepción subjetiva de enfermedad

El ítem de percepción subjetiva de enfermedad del WHOQOL-BREF tiene tres opciones de respuesta: Sí (12.76%), No (70.36%) y No está seguro (16.88%). El análisis de varianza arroja diferencias de medias significativas, en la medida calculada con Rasch, entre estos tres grupos con un valor $F = 36.75$ ($\text{sig.} < .001$). El análisis Post Hoc mediante Bonferroni evidencia que los grupos Si y No está seguro no son significativamente diferentes en sus promedios ($\text{sig.} = .636$). Las diferencias significativas se presentan entre No enfermo y Si enfermo ($\text{sig.} < .001$) y entre No enfermo y No está seguro ($\text{sig.} < .001$). De esta forma se procede a hacer el análisis DIF con la dupla Sí/No está seguro vs No.

Tabla 9 Funcionamiento diferencial de los ítems por reporte de enfermedad

Ítem	Enfermo/No está seguro			No enfermo			Contraste DIF	E.E. conjunto	Rasch Welch		Mantel		BH crítico	
	promedio Obs-Esp	Medida DIF	E.E. DIF	promedio Obs-Esp	Medida DIF	E.E. DIF			t	d.f.	p	Chi-squ		p
Q3 Dolor y Disconfort	-0.08	-0.49	0.1	0.04	-0.8	0.07	0.31	0.13	2.4	554	0.015	11.9	0.001	0.003
Q4 Dependencia medicación	-0.22	-0.82	0.11	0.09	-1.8	0.09	0.98	0.14	7.0	610	0.000	42.9	0.000	0.001
Q5 Sentimientos positivos	0.13	-0.04	0.1	-0.06	0.44	0.07	-0.47	0.12	3.8	529	0.000	10.9	0.001	0.004
Q6 Espiritualidad	0.03	-0.31	0.1	-0.01	-0.19	0.07	-0.12	0.12	-1.0	536	0.342	3.5	0.062	0.011
Q7 Pensamiento y Memoria	0.05	1.18	0.12	-0.02	1.41	0.07	-0.22	0.14	1.6	510	0.117	0.0	0.931	0.024
Q8 Seguridad	0.12	0.46	0.11	-0.05	0.91	0.07	-0.45	0.13	3.5	525	0.001	13.7	0.000	0.002
Q9 Ambiente físico	0.09	0.61	0.11	-0.04	0.95	0.07	-0.34	0.13	2.6	522	0.009	2.5	0.115	0.013
Q10 Energía y Fatiga	-0.14	0.22	0.11	0.06	-0.31	0.07	0.52	0.13	4.1	533	0.000	10.3	0.001	0.005
Q11 Imagen corporal	-0.04	1.02	0.11	0.02	-1.21	0.08	0.19	0.13	1.4	566	0.158	0.0	0.832	0.023
Q12 Finanzas	0.06	0.82	0.11	-0.03	1.1	0.07	-0.27	0.13	2.0	517	0.043	2.3	0.125	0.014
Q13 Acceso a información	0.05	0.43	0.11	-0.02	0.62	0.07	-0.18	0.13	1.4	522	0.150	1.6	0.206	0.016
Q14 Actividades placenteras	0	0.74	0.11	0	0.74	0.07	0	0.13	0.0	514	1.000	1.6	0.203	0.015
Q15 Movilidad	-0.06	1.43	0.11	0.02	-1.71	0.09	0.28	0.15	1.9	583	0.054	5.2	0.022	0.009
Q16 Sueño y Descanso	-0.09	0.31	0.11	0.04	-0.03	0.07	0.34	0.13	2.7	526	0.008	8.6	0.003	0.007
Q17 Actividades cotidianas	-0.06	-0.08	0.1	0.03	-0.31	0.07	0.23	0.13	1.9	538	0.063	5.3	0.022	0.008
Q18 Capacidad para trabajar	-0.05	-0.15	0.1	0.02	-0.31	0.07	0.17	0.13	1.4	538	0.178	1.5	0.216	0.017
Q19 Autoestima	-0.03	-0.63	0.1	0.01	-0.76	0.07	0.13	0.13	1.0	551	0.324	0.2	0.680	0.021
Q20 Relaciones Personales	0.02	-0.36	0.1	-0.01	-0.32	0.07	-0.05	0.13	-0.4	538	0.714	1.4	0.231	0.019
Q21 Actividad sexual	-0.01	-0.18	0.1	0	-0.21	0.07	0.03	0.12	0.2	536	0.831	0.0	0.991	0.025
Q22 Soporte social	0.09	-0.2	0.1	-0.04	0.12	0.07	-0.32	0.12	2.6	531	0.010	5.2	0.023	0.010
Q23 Condiciones vivienda	0.08	-0.77	0.11	-0.03	-0.49	0.07	-0.28	0.13	2.2	539	0.026	8.8	0.003	0.006
Q24 Servicios de salud	0.06	0.53	0.11	-0.03	0.76	0.07	-0.23	0.13	1.8	521	0.075	0.8	0.382	0.020
Q25 Transporte	0.02	0.27	0.11	-0.01	0.36	0.07	-0.09	0.13	-0.7	524	0.498	0.0	0.829	0.022
Q26 Sentimientos negativos	-0.01	0.85	0.11	0	0.82	0.07	0.03	0.13	0.2	510	0.837	1.5	0.228	0.018

La tabla 9 muestra el Funcionamiento Diferencial de los Ítems por reporte de enfermedad. Considerando relevante un contraste DIF mayor a igual a .5, los ítems Dependencia a medicación y Energía y Fatiga cumplen con este criterio. Para ambos ítems se obtiene significación por el método Rasch-Welch y Mantel Haenszel con valores p menores al valor BH crítico.

Los ítems Dolor y Disconfort, Sentimientos positivos, Seguridad, Sueño y Descanso y Condiciones de vivienda obtiene valores p inferiores a BH crítico pero los tamaños de DIF son menores a .5.

Funcionamiento Diferencial de los Ítems por grupos de edad

La tabla 10 muestra los resultados para el análisis de ítems por grupos de edad considerando menores de 30 años y personas con edad igual o mayor a 30 años. Con el criterio de contraste mayor o igual a .5 logitos, los ítems Sentimientos positivos, Energía y fatiga y Servicios de salud muestran DIF significativo por Rasch-Welch y Mantel Haenszel con valores p menores al valor BH crítico.

Tabla 10 Funcionamiento diferencial de los ítems por grupos de edad

Ítem	Menores 30			30 y mayores			Contraste DIF	E.E. conjunto	Rasch Welch		Mantel		BH crítico	
	promedio Obs-Esp	Medida DIF	E.E. DIF	promedio Obs-Esp	Medida DIF	E.E. DIF			t	d.f.	p	Chi-squ		p
Q3 Dolor y Disconfort	-0.01	-0.69	0.07	0.05	-0.85	0.16	0.15	0.17	0.9	210	0.379	0.6	0.455	0.018
Q4 Dependencia medicación	0.01	-1.42	0.07	-0.05	-1.24	0.17	-0.17	0.18	-1.0	216	0.344	0.2	0.628	0.019
Q5 Sentimientos positivos	0.03	0.22	0.06	-0.17	0.73	0.16	-0.51	0.17	-3.1	206	0.003	8.8	0.003	0.003
Q6 Espiritualidad	0.02	-0.28	0.06	-0.09	-0.01	0.15	-0.27	0.16	-1.7	212	0.098	1.5	0.221	0.016
Q7 Pensamiento y Memoria	0	1.34	0.07	-0.03	1.43	0.17	-0.09	0.19	-0.5	205	0.623	0.6	0.421	0.017
Q8 Seguridad	0.02	0.74	0.06	-0.11	1.09	0.16	-0.35	0.18	-2.0	204	0.045	6.0	0.014	0.007
Q9 Ambiente físico	0.02	0.81	0.06	-0.1	1.14	0.16	-0.34	0.18	-1.9	203	0.058	4.7	0.031	0.009
Q10 Energía y Fatiga	-0.04	-0.02	0.06	0.24	-0.72	0.16	0.7	0.17	4.2	206	0.000	21.4	0.000	0.001
Q11 Imagen corporal	-0.01	-1.07	0.07	0.09	-1.39	0.17	0.32	0.19	1.7	206	0.088	3.3	0.071	0.014
Q12 Finanzas	-0.01	1.02	0.07	0.03	0.93	0.16	0.09	0.17	0.5	209	0.603	0.2	0.659	0.021
Q13 Acceso a información	-0.02	0.61	0.06	0.13	0.22	0.15	0.39	0.16	2.4	211	0.018	4.5	0.034	0.010
Q14 Actividades placenteras	0.01	0.74	0.06	-0.04	0.86	0.16	-0.11	0.17	-0.7	207	0.514	0.2	0.646	0.020
Q15 Movilidad	0.01	-1.61	0.08	-0.04	-1.45	0.18	-0.16	0.19	-0.8	216	0.400	0.1	0.706	0.022
Q16 Sueño y Descanso	-0.02	0.13	0.06	0.12	-0.21	0.15	0.34	0.16	2.1	210	0.037	3.5	0.063	0.013
Q17 Actividades cotidianas	-0.02	-0.21	0.06	0.09	-0.48	0.15	0.28	0.17	1.7	209	0.097	4.5	0.034	0.011
Q18 Capacidad para trabajar	-0.02	-0.2	0.06	0.13	-0.58	0.15	0.38	0.17	2.3	208	0.024	10.1	0.002	0.002
Q19 Autoestima	0	-0.71	0.07	0	-0.71	0.16	0	0.17	0.0	212	1.000	0.0	0.944	0.024
Q20 Relaciones Personales	-0.01	-0.29	0.06	0.07	-0.51	0.15	0.22	0.17	1.3	209	0.187	1.7	0.188	0.015
Q21 Actividad sexual	0	-0.21	0.06	-0.02	-0.16	0.15	-0.04	0.16	-0.3	211	0.787	0.1	0.772	0.023
Q22 Soporte social	0.03	-0.03	0.06	-0.15	0.4	0.15	-0.43	0.16	-2.6	209	0.009	8.4	0.004	0.005
Q23 Condiciones vivienda	0.02	-0.64	0.07	-0.12	-0.28	0.15	-0.36	0.16	-2.2	214	0.029	5.5	0.019	0.008
Q24 Servicios de salud	0.03	0.61	0.06	-0.17	1.17	0.16	-0.56	0.18	-3.2	202	0.002	8.6	0.003	0.004
Q25 Transporte	0	0.3	0.06	0.02	0.26	0.15	0.04	0.16	0.2	210	0.818	0.0	0.994	0.025
Q26 Sentimientos negativos	-0.02	0.86	0.06	0.13	0.47	0.15	0.39	0.17	2.4	212	0.019	6.1	0.014	0.006

Los ítems Capacidad para trabajar y Soporte social muestran valores de significación menores al valor crítico BH. Sin embargo los contrastes DIF en ambos casos son menores a .5 logitos.

Funcionamiento Diferencial de los Ítems por nivel de calidad de vida

La tabla 11 muestra los cálculos de DIF por niveles de calidad de vida. La clasificación en dos niveles jerárquicos de puntuación se obtuvo a partir del procedimiento MA2 (dos grupos ascendentes a partir de la medida) de Winsteps. Considerando el criterio de contraste mayor o igual a .5 logitos, los ítems Espiritualidad, Imagen corporal y Autoestima muestran valores significativos por el método de Welch y menores al valor BH crítico.

Los ítems Dolor y Disconfort, Ambiente físico, Energía y fatiga, Servicios de salud y Transporte muestran valores de significación menores al valor crítico BH. Sin embargo los valores de contraste DIF son menores a .0.5.

Tabla 11 Funcionamiento diferencial de los ítems por nivel de calidad de vida

Ítem	Menores 0 logitos			mayores 0 logitos			Contraste DIF	E.E. conjunto	t	Rasch Welch		BH crítico
	promedio Obs-Esp	Medida DIF	E.E. DIF	promedio Obs-Esp	Medida DIF	E.E. DIF				d.f.	p	
Q3 Dolor y Discomfort	0.09	-0.92	0.08	-0.07	-0.48	0.08	-0.44	0.12	-3.8	775	0.000	0.005
Q4 Dependencia medicación	-0.01	-1.37	0.08	0.01	-1.44	0.11	0.07	0.14	0.5	788	0.594	0.019
Q5 Sentimientos positivos	-0.01	0.3	0.09	0.01	0.28	0.07	0.02	0.12	0.2	748	0.856	0.023
Q6 Espiritualidad	-0.13	0.11	0.09	0.1	-0.52	0.08	0.63	0.12	5.3	770	0.000	0.004
Q7 Pensamiento y Memoria	0.1	0.93	0.1	-0.08	1.53	0.08	-0.6	0.13	-4.8	729	0.000	0.001
Q8 Seguridad	-0.02	0.84	0.1	0.02	0.73	0.07	0.11	0.12	0.9	721	0.361	0.018
Q9 Ambiente físico	0.07	0.62	0.09	-0.06	1	0.07	-0.38	0.12	-3.2	734	0.002	0.007
Q10 Energía y Fatiga	-0.08	0.06	0.09	0.06	-0.3	0.08	0.36	0.12	3.1	766	0.002	0.008
Q11 Imagen corporal	-0.11	-0.88	0.08	0.08	-1.54	0.11	0.66	0.14	4.8	788	0.000	0.003
Q12 Finanzas	0.04	0.88	0.1	-0.03	1.09	0.07	-0.22	0.12	-1.8	721	0.079	0.011
Q13 Acceso a información	0.04	0.44	0.09	-0.03	0.62	0.07	-0.18	0.12	-1.6	740	0.121	0.014
Q14 Actividades placenteras	-0.05	0.89	0.1	0.04	0.64	0.07	0.24	0.12	1.9	720	0.053	0.010
Q15 Movilidad	0.02	-1.63	0.09	-0.02	-1.49	0.11	-0.14	0.14	-1.0	788	0.310	0.016
Q16 Sueño y Descanso	0.01	0.06	0.09	-0.01	0.09	0.07	-0.02	0.11	-0.2	757	0.829	0.021
Q17 Actividades cotidianas	-0.02	-0.19	0.08	0.02	-0.29	0.08	0.11	0.11	0.9	769	0.358	0.017
Q18 Capacidad para trabajar	-0.03	-0.18	0.08	0.02	-0.32	0.08	0.14	0.12	1.3	770	0.211	0.015
Q19 Autoestima	-0.13	-0.4	0.08	0.1	-1.09	0.1	0.69	0.13	5.5	786	0.000	0.002
Q20 Relaciones Personales	-0.01	-0.28	0.08	0.01	-0.34	0.08	0.06	0.12	0.5	771	0.596	0.020
Q21 Actividad sexual	0.01	-0.2	0.08	0	-0.2	0.08	0	0.11	0.0	767	1.000	0.025
Q22 Soporte social	0.01	0	0.08	-0.01	0.03	0.08	-0.02	0.11	-0.2	759	0.846	0.022
Q23 Condiciones vivienda	0	-0.58	0.08	0	-0.58	0.08	0	0.12	0.0	778	1.000	0.024
Q24 Servicios de salud	0.09	0.4	0.09	-0.07	0.83	0.07	-0.43	0.12	-3.8	741	0.000	0.006
Q25 Transporte	0.08	0.11	0.09	-0.06	0.46	0.07	-0.34	0.11	-3.1	751	0.002	0.009
Q26 Sentimientos negativos	0.04	0.67	0.1	-0.03	0.87	0.07	-0.2	0.12	-1.7	730	0.096	0.013

Gráficas de decisión para DIF

En conclusión a partir del análisis de funcionamiento diferencial se encuentran ocho ítems que muestran cálculos a favor de considerar el DIF en al menos una clasificación de la muestra (tabla 12). El ítem Energía y fatiga muestra DIF en dos de las cuatro formas de clasificar los participantes.

Tabla 12 Resumen de cálculos Funcionamiento Diferencial de los Ítems

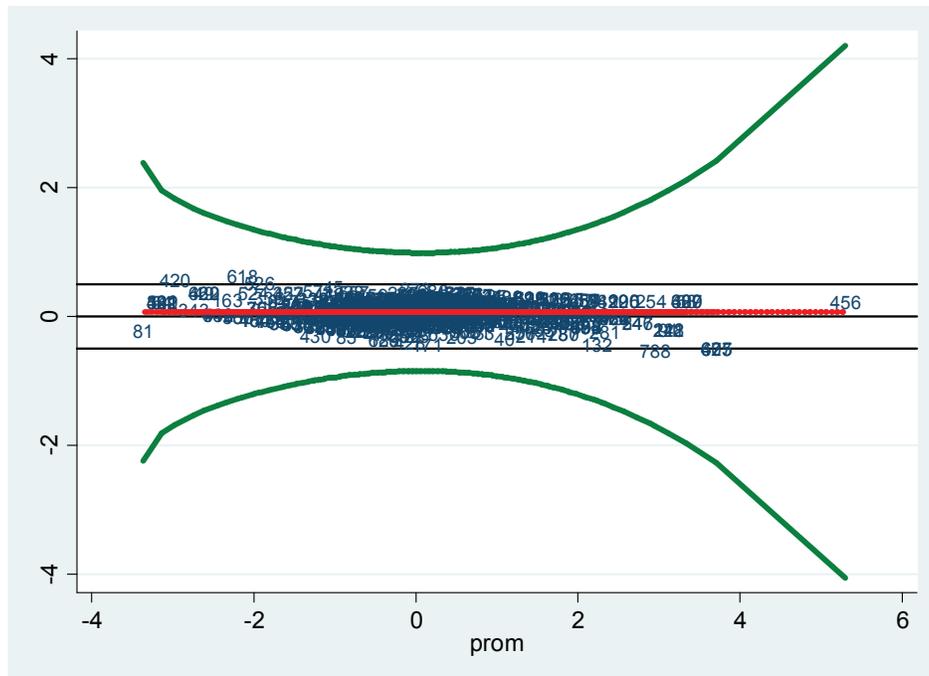
Ítem	Clasificación			
	Sexo	Enfermedad	Edad	Calidad de vida
Q4 Dependencia medicación		x		
Q5 Sentimientos positivos			x	
Q6 Espiritualidad				x
Q10 Energía y Fatiga		x	x	
Q11 Imagen corporal				x
Q19 Autoestima				x
Q24 Servicios de salud			x	
Q26 Sentimientos negativos	x			

Para corroborar estos hallazgos y concluir cuales ítems se van a considerar como relevantes según DIF se construyeron gráficos de Bland y Altman a partir de los errores calculados por Winsteps (gráfica 9).

De los ítems considerados DIF por sexo se mantiene Sentimientos negativos (9a); por enfermedad se mantienen Dependencia medicación y Energía y Fatiga (6b); por edad se mantiene únicamente energía y fatiga (9c); y por nivel de calidad de vida se mantiene únicamente Autoestima (9d).

Comparación de la medida de calidad de vida con y sin ítems DIF

A partir de los cuatro ítems para los cuales se identificó DIF relevante y significativo: Dependencia medicamentos, Energía y fatiga, Autoestima y Sentimientos Negativos, se analiza su impacto sobre la estimación del nivel de calidad de vida total. Para esto se correlacionan las puntuaciones obtenidas por los participantes en la forma original de 24 ítems y en la forma sin ítems con DIF relevante y significativo de 20 ítems obteniendo un valor de .9895. Esto indica que el cambio en las medidas es irrelevante.



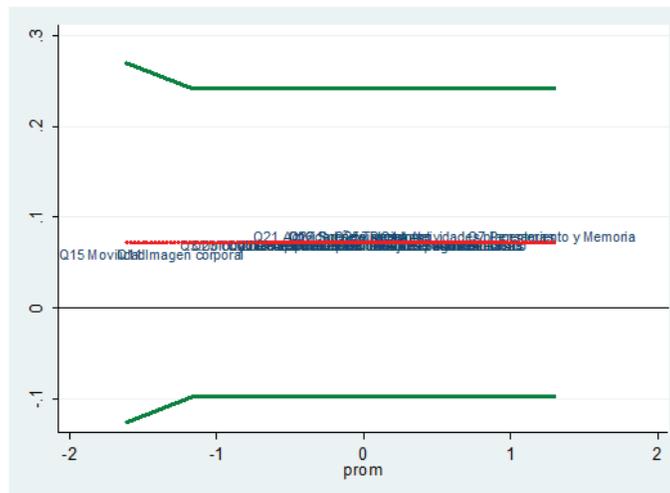
Gráfica 10 Gráfica de Bland y Altman para diferencias de medida entre formas de 24 y 20 ítems

Para corroborar la interpretación del valor de la correlación entre las medidas obtenidas por la forma de 24 ítems y la forma insesgada de 20 ítems se realizó el gráfico de Bland y Altman, el cual a su vez permite considerar los errores de medición. La gráfica 10 muestra esta relación y no evidencia casos particulares de diferencia significativa.

Tabla 13 Comparación ajuste de los ítems a partir de 24 y 20 ítems del WHOQOL-BREF

Ítem	24 ítems				20 ítems			
	Medida	E.E.	MNSQ Outfit	AKA	Medida	E.E.	MNSQ Outfit	AKA
Q3 Dolor y Discomfort	-0.71	0.06	1.5	0.34	-0.78	0.06	1.64	0.32
Q4 Dependencia medicación	-1.39	0.07	1.45	0.39				
Q5 Sentimientos positivos	0.28	0.06	1.14	0.52	0.21	0.06	1.1	0.54
Q6 Espiritualidad	-0.22	0.06	1.42	0.53	-0.29	0.06	1.4	0.54
Q7 Pensamiento y Memoria	1.34	0.06	1.21	0.38	1.26	0.06	1.16	0.41
Q8 Seguridad	0.77	0.06	1	0.56	0.7	0.06	0.96	0.58
Q9 Ambiente físico	0.86	0.06	1.03	0.5	0.79	0.06	1.02	0.51
Q10 Energía y Fatiga	-0.13	0.06	0.89	0.62				
Q11 Imagen corporal	-1.13	0.06	1.15	0.53	-1.19	0.06	1.22	0.5
Q12 Finanzas	1.02	0.06	1.15	0.47	0.95	0.06	1.12	0.48
Q13 Acceso a información	0.55	0.06	1.13	0.52	0.48	0.06	1.1	0.53
Q14 Actividades placenteras	0.73	0.06	0.94	0.6	0.65	0.06	0.92	0.61
Q15 Movilidad	-1.58	0.07	1.04	0.43	-1.64	0.07	1.07	0.42
Q16 Sueño y Descanso	0.09	0.06	0.96	0.56	0.01	0.06	0.96	0.56
Q17 Actividades cotidianas	-0.24	0.06	0.54	0.69	-0.31	0.06	0.56	0.68
Q18 Capacidad para trabajar	-0.25	0.06	0.66	0.65	-0.32	0.06	0.67	0.64
Q19 Autoestima	-0.7	0.06	0.65	0.68				
Q20 Relaciones Personales	-0.31	0.06	0.72	0.64	-0.38	0.06	0.71	0.63
Q21 Actividad sexual	-0.2	0.06	0.89	0.59	-0.28	0.06	0.86	0.59
Q22 Soporte social	0.03	0.06	0.91	0.56	-0.05	0.06	0.9	0.57
Q23 Condiciones vivienda	-0.58	0.06	0.81	0.59	-0.65	0.06	0.78	0.6
Q24 Servicios de salud	0.67	0.06	1.11	0.47	0.6	0.06	1.06	0.5
Q25 Transporte	0.32	0.06	1.19	0.47	0.24	0.06	1.15	0.48
Q26 Sentimientos negativos	0.8	0.06	1	0.48				
Media	0	0.06	1.02		0	0.06	1.02	
Desviación	0.75	0.00	0.24		0.72	0.00	0.24	

La tabla 13 compara las medidas de ajuste de los ítems al eliminar los cuatro que muestran DIF. No se observan diferencias entre los 20 ítems que son comparables. La gráfica 11 confirma estos resultados.



Gráfica 11 Gráfica de Bland y Altman para dificultades de 24 y 20 ítems.

Confiabilidad de los Ítems y de las Personas

Del total de los 792 participantes, solo uno muestra la puntuación máxima posible. La tabla 14 muestra los datos de modelo para los 791 casos no extremos. Se obtiene una confiabilidad de .88 y una separación de 2.67; las personas exhiben niveles de calidad de vida entre -3.48 logitos y 3.45 logitos. A partir de estos datos se establece un índice de separación de 3.9 estratos. Esto indica una capacidad de discriminación aceptable.

Tabla 14 Confiabilidad de las personas

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	49.4	24.0	.18	.35	1.01	-.1	1.02	.0
S.D.	9.3	.0	1.08	.06	.38	1.5	.48	1.4
MAX.	70.0	24.0	3.45	.73	2.48	4.7	6.99	4.7
MIN.	26.0	24.0	-3.48	.31	.19	-5.5	.21	-5.0
REAL RMSE	.38	TRUE SD	1.02	SEPARATION	2.67	PERSON RELIABILITY	.88	
MODEL RMSE	.35	TRUE SD	1.03	SEPARATION	2.91	PERSON RELIABILITY	.89	
S.E. OF PERSON MEAN = .04								

A partir de los 24 ítems del WHOQOL-BREF se obtiene una confiabilidad para Los ítems (tabla 15) de .99 y un valor de separación de 12.19; los ítems representan niveles de calidad de vida entre -1.58 logitos y 1.34 logitos. A partir de estos datos se obtiene un índice de separación de 16.6 estratos. Lo cual indica una discriminación adecuada.

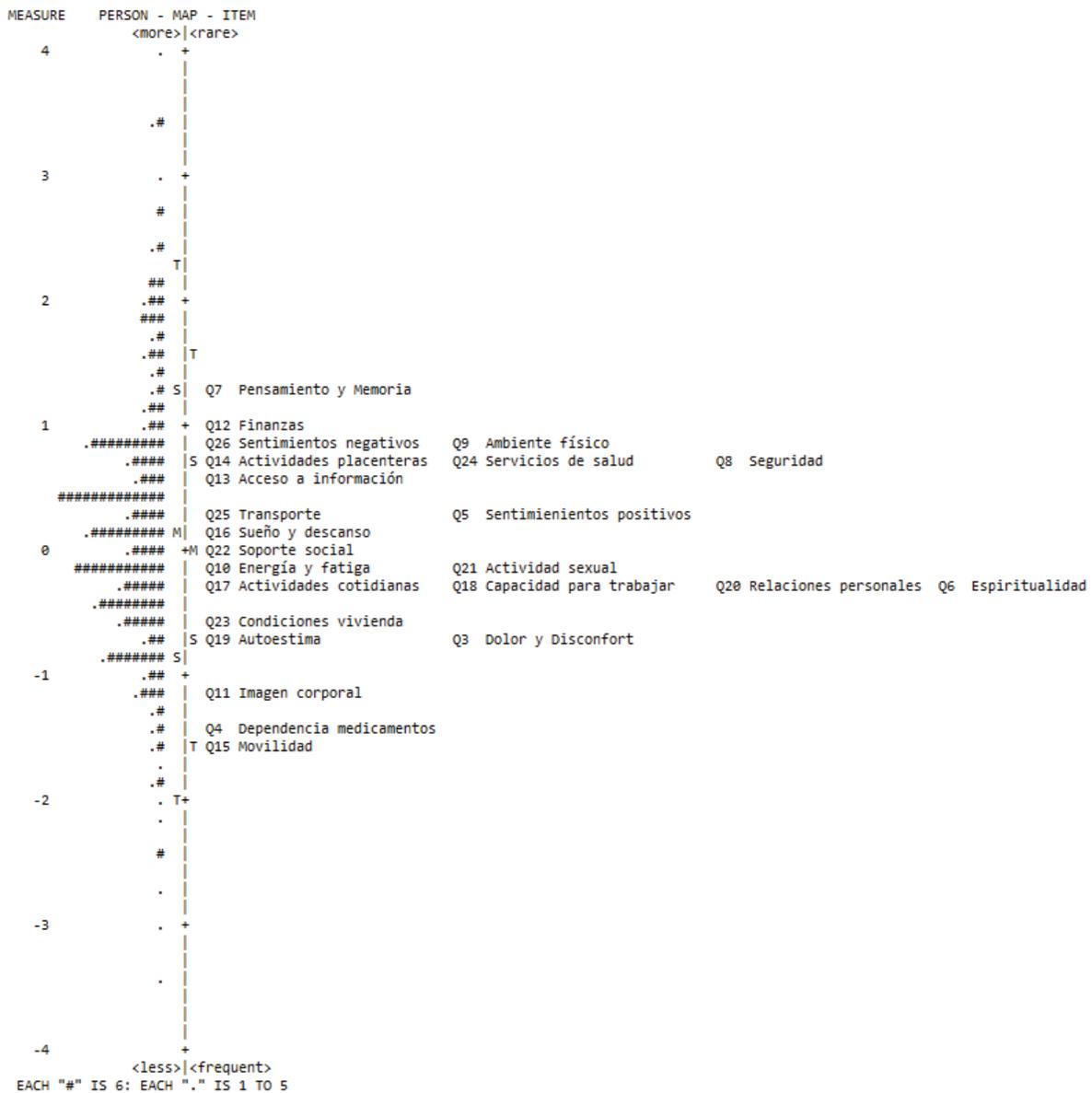
Tabla 15 Confiabilidad de los Ítems

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1632.3	792.0	.00	.06	1.00	-.2	1.02	-.1
S.D.	221.9	.0	.75	.00	.22	4.9	.24	4.5
MAX.	2067.0	792.0	1.34	.07	1.43	9.0	1.50	7.8
MIN.	1241.0	792.0	-1.58	.06	.54	-9.9	.54	-9.9
REAL RMSE	.06	TRUE SD	.75	SEPARATION	12.19	ITEM	RELIABILITY	.99
MODEL RMSE	.06	TRUE SD	.75	SEPARATION	12.74	ITEM	RELIABILITY	.99
S.E. OF ITEM MEAN = .16								

Jerarquía de los ítems y Targeting

La gráfica 12 muestra ordenados por logitos el nivel de calidad de vida expresado por los participantes hacía el lado izquierdo de la línea punteada. De forma similar, la gráfica muestra en igual unidad de medida el nivel de calidad de vida evidenciado por cada ítem hacía el lado derecho de la línea punteada. Tanto para los ítems como para las personas

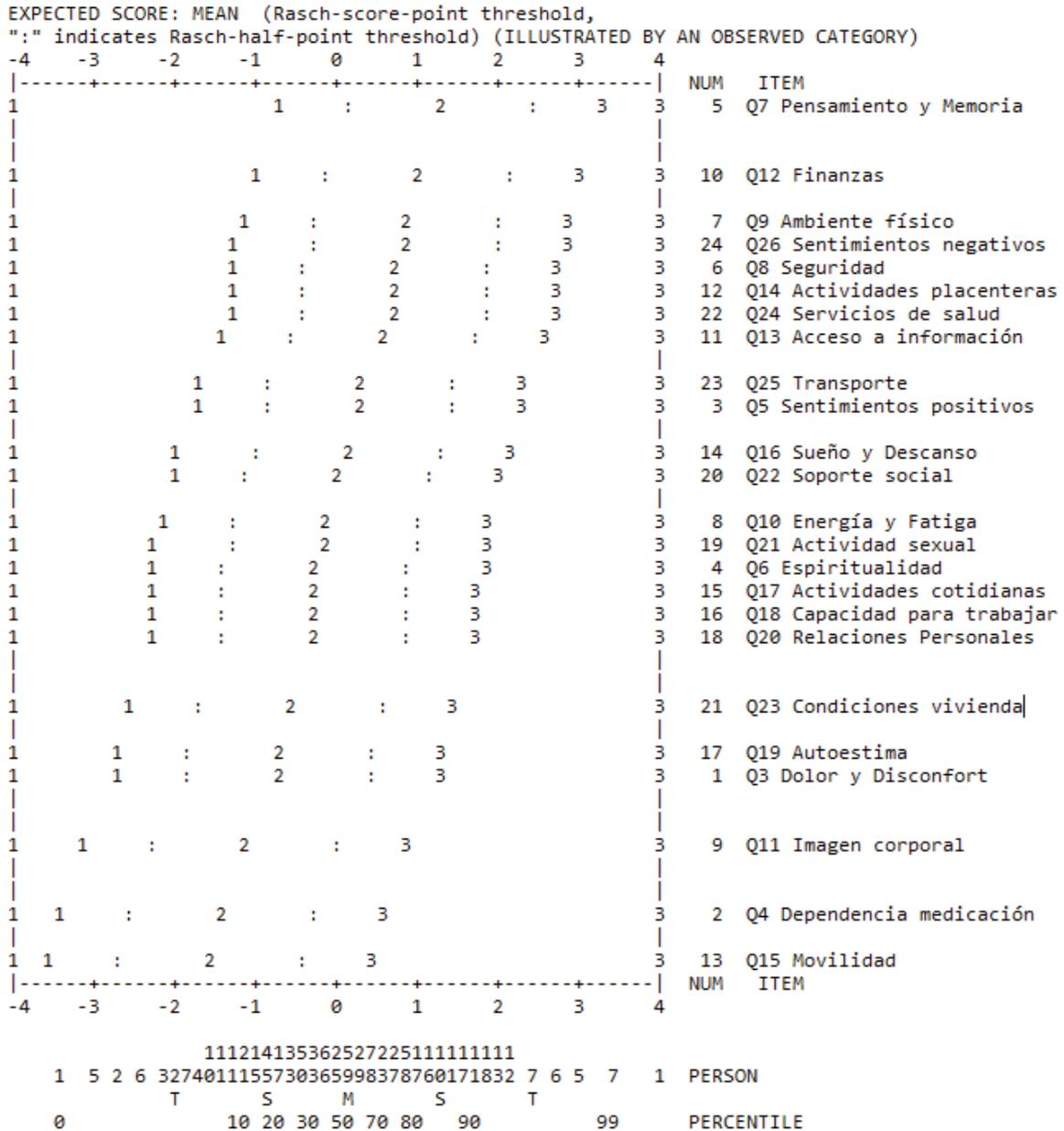
están marcados el promedio (M) una desviación estándar (S) y dos desviaciones estándar (T).



Gráfica 12 Mapa Ítem-Persona WHOQOL-BREF

Se observa que tanto personas como ítems acercan su media al valor teórico del modelo de cero logitos. Sin embargo, la dispersión de las personas excede a la dispersión de los ítems. Esto implica que los ítems del WHOQOL BREF no muestran valores de dificultad que representen el nivel de calidad de vida mostrado por las personas por encima de una desviación estándar en logitos, ni por debajo de dos desviaciones estándar en logitos.

El ordenamiento jerárquico de los ítems permite conceptualizar la forma como el WHOQOL-BREF mide la calidad de vida.



Gráfica 13 Valores en logitos para cada categoría de respuesta por ítem del WHOQOL-BREF

La gráfica 13 permite complementar el análisis jerárquico de los ítems al representar el valor en lógitos de cada categoría de respuesta para cada ítem del WHOQOL-BREF. De esta forma el rango de valores oscila entre ± 4 . Vale la pena recalcar que la gráfica 10

Tabla 16 Estadísticos descriptivos de la medida Rasch por categorías de autovaloración de calidad de vida

Medidas Rasch	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al		Mínimo	Máximo
					95%			
					Límite inferior	Límite superior		
Muy pobre	6	-.7600	1.87737	.76643	-2.7302	1.2102	-3.03	1.54
Pobre	11	-.4291	1.66031	.50060	-1.5445	.6863	-2.70	2.68
Ni pobre ni buena	73	-.9827	.83954	.09826	-1.1786	-.7869	-3.03	.81
Buena	472	.0357	.88921	.04093	-.0447	.1161	-3.48	3.45
Muy buena	220	.9131	1.04820	.07067	.7739	1.0524	-2.21	5.41
Total	782	.1748	1.10006	.03934	.0976	.2520	-3.48	5.41

Según análisis de varianza, estas diferencias son estadísticamente significativas ($F=64.69$; sig. $<.001$).

Tabla 17 Análisis de Bonferroni medida Rasch por autovaloración de calidad de vida

(I) Calidad de vida		Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Pobre	Muy pobre	.33091	.48479	1.000	-1.0338	1.6956
Ni pobre ni buena	Muy pobre	-.22274	.40568	1.000	-1.3648	.9193
	Pobre	-.55365	.30895	.735	-1.4234	.3161
Buena	Muy pobre	.79568	.39244	.430	-.3091	1.9004
	Pobre	.46477	.29135	1.000	-.3554	1.2849
Muy buena	Ni pobre ni buena	1.01842*	.12014	.000	.6802	1.3566
	Muy pobre	1.67314*	.39525	.000	.5605	2.7858
	Pobre	1.34223*	.29512	.000	.5114	2.1730
	Ni pobre ni buena	1.89588*	.12902	.000	1.5327	2.2591
	Buena	.87746*	.07798	.000	.6579	1.0970

*. La diferencia de medias es significativa al nivel .05.

La tabla 17 muestra los resultados de los análisis Post Hoc mediante Bonferroni. Se evidencia que la categoría muy buena se diferencia significativamente de las demás categorías; la categoría Buena, se diferencia significativamente únicamente de la categoría ni pobre ni buena. Las categorías inferiores de autovaloración de calidad de vida no se diferencian significativamente entre sí.

Relación de la medición Rasch del WHOQOL-BREF con autovaloración de la satisfacción con la salud

Al analizar el nivel de calidad de vida obtenido del WHOQOL-BREF en logitos por las categorías jerárquicas de autovaloración de satisfacción con la salud, se observa un

aumento de la media en logitos con excepción de la categoría Muy insatisfecho. Esta categoría guarda la relación jerárquica de la variable y estos resultados sugieren que quienes optan por esta categoría de respuesta muestran un nivel de calidad de vida mayor aún que quienes optan por Satisfecho.

Tabla 18 Estadísticos descriptivos de la medida Rasch por categorías de autovaloración de satisfacción con la salud

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al		Mínimo	Máximo
					95%			
					Límite inferior	Límite superior		
Muy insatisfecho	45	.6811	.92410	.13776	.4035	.9587	-1.14	2.42
Insatisfecho	42	-.9229	.98366	.15178	-1.2294	-.6163	-3.03	1.68
Ni insatisfecho ni satisfecho	94	-.6466	.98689	.10179	-.8487	-.4445	-3.03	1.84
Satisfecho	396	.0753	.89530	.04499	-.0132	.1637	-3.48	3.01
Muy satisfecho	198	.8971	1.05955	.07530	.7486	1.0456	-1.68	5.41
Total	775	.1788	1.10108	.03955	.1011	.2564	-3.48	5.41

Según análisis de varianza, estas diferencias son estadísticamente significativas ($F=63.5$; sig. $<.001$).

Tabla 19 Análisis de Bonferroni medida Rasch por autovaloración de satisfacción con la salud

(I) Satisfacción salud		Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Insatisfecho	Muy insatisfecho	-1.60397*	.20538	.000	-2.1822	-1.0258
	Insatisfecho	.27626	.17767	1.000	-.2239	.7764
Satisfecho	Muy insatisfecho	-.60586*	.15059	.001	-1.0298	-.1819
	Insatisfecho	.99811*	.15535	.000	.5608	1.4354
	Ni insatisfecho ni satisfecho	.72185*	.10983	.000	.4127	1.0310
Muy satisfecho	Muy insatisfecho	.21601	.15809	1.000	-.2290	.6611
	Insatisfecho	1.81998*	.16262	.000	1.3622	2.2778
	Ni insatisfecho ni satisfecho	1.54372*	.11990	.000	1.2062	1.8813
	Satisfecho	.82187*	.08332	.000	.5873	1.0564

La tabla 19 muestra los resultados de los análisis Post Hoc mediante Bonferroni. Se evidencia que la categoría muy satisfecho se diferencia significativamente de las demás categorías con excepción de muy insatisfecho; la categoría Satisfecho, se diferencia

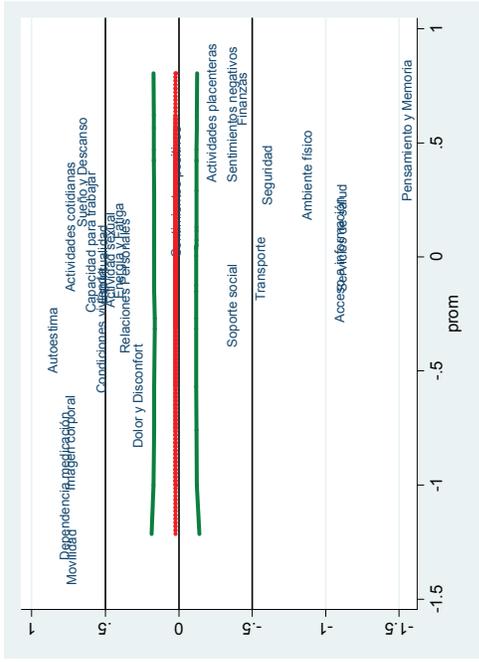
significativamente de las demás categorías inferiores con excepción de Muy insatisfecho. La categoría Muy insatisfecho muestra diferencia significativa con las demás categorías.

Comparación de los resultados obtenidos con los antecedentes

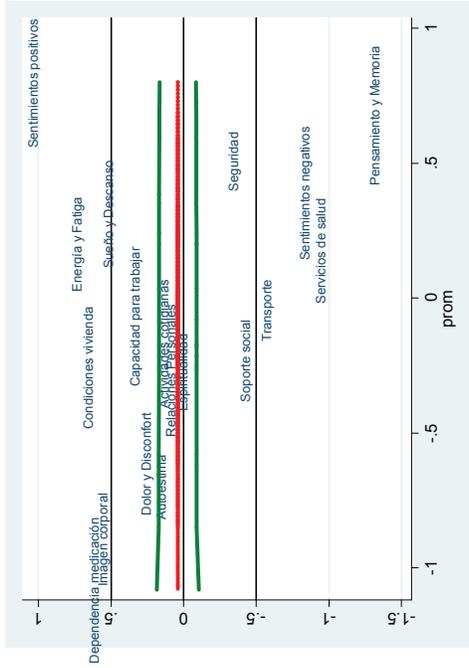
En la tabla 1 se reportaron los valores de dificultad reportados en los antecedentes revisados. Mediante las gráficas de Bland y Altman (gráfica 15) se comparan las dificultades, con sus respectivos errores, obtenidas en la muestra estudiada en este proyecto con los resultados de los antecedentes encontrados en la literatura. La tabla 20 resume los resultados mostrando con marcas (x) los ítems en los cuales las diferencias en la dificultad no son relevantes al comparar los antecedentes con los resultados de el presente estudio. De los 24 ítems, 3 son comparables con los 3 antecedentes; 4 con 2 antecedentes; 11 con un antecedente; y 6 no son comparables con ningún antecedente.

Tabla 20 Ítems dentro de los intervalos de acuerdo comparación Rojas 2013 con antecedentes

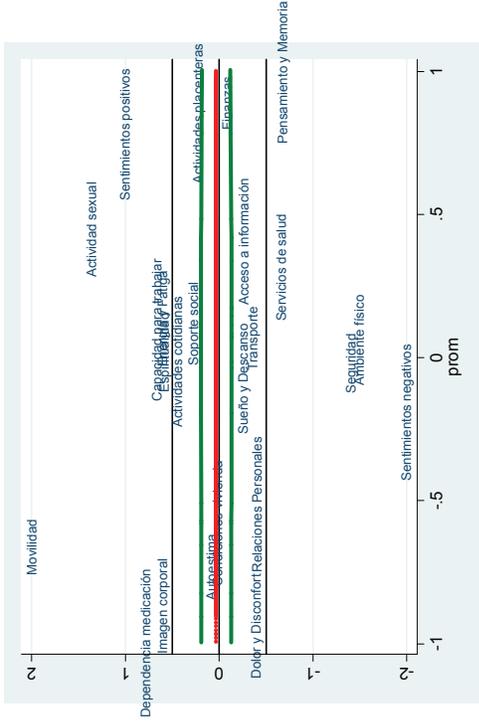
Ítem	Rocha y otros, 2012 Depresivos	Liang y otros, 2009 Mayor 65 años	Wang y otros, 2006 20 - 65 años
Pensamiento y Memoria			
Finanzas	x	x	---
Ambiente físico			---
Sentimientos negativos	x		
Seguridad			x
Actividades placenteras	x	x	---
Servicios de salud			
Acceso a información		x	---
Transporte		x	
Sentimientos positivos	x		
Sueño y Descanso		x	
Soporte social	x	x	x
Energía y Fatiga	x		
Actividad sexual	x		---
Espiritualidad			x
Actividades cotidianas		x	x
Capacidad para trabajar			x
Relaciones Personales	x	x	x
Condiciones vivienda		x	
Autoestima		x	x
Dolor y Disconfort	x	x	x
Imagen corporal			
Dependencia medicación			
Movilidad			---



a. Rocha y otros, 2012 Depresivos



c. Wang y otros, 2006 20-65 años



b. Liang y otros, 2009 Adultos mayores

Gráfica 15 Gráficas de Bland y Altman para la comparación de las dificultades obtenidas con las reportadas en los antecedentes

Discusión

En este apartado se presentan aspectos interpretativos relacionados por cada objetivo específico. A partir de estas conclusiones e inferencias se discuten los aspectos relevantes de la discusión en relación al objetivo general de la investigación.

En relación al objetivo “Analizar las características de las categorías de respuesta a los ítems del WHOQOL-BREF.”

En todos los ítems el desorden de categorías se da entre los valores 1, 2 y 3 con excepción del ítem “Dependencia a la medicación” que muestra desorden desde el valor 1 hasta el valor 4, esto probablemente debido a las características de salud de la muestra. Una vez se colapsaron las categorías al modelo 11123, los problemas de orden se resolvieron para todos los ítems con excepción de los ítems “Dependencia a la medicación” y “Dolor y Disconfort”. Estos dos ítems podrían ser calificados como “sí o no” (11112) en personas sanas, sin embargo, con el propósito de no generar complejidad innecesaria al modelo de análisis, se optó por mantener la misma forma de calificación para todos los ítems sin que esto afectase las propiedades de la escala.

Cuando se comparan los resultados con análisis desde modelos de Rasch al WHOQOL-BREF realizados en los antecedentes encontrados en la literatura, el presente trabajo es el primero, hasta donde el autor pudo constatar, que genera una única forma de graduar las respuestas de los ítems. En los antecedentes especialmente en Rocha y otros (2012) al no haber uniformidad en los inconvenientes con el ordenamiento y distancia entre categorías deben optar por diferentes formas de gradación de respuesta para los ítems, lo que los lleva al análisis de crédito parcial añadiendo complejidad innecesaria a un instrumento que originalmente no la requiere.

Estos hallazgos en relación con las categorías de respuesta permiten inferir la no necesidad de alta discriminación mediante las categorías de los ítems (a través de una mayor cantidad de categorías) para establecer el nivel de calidad de vida en personas de población general.

En relación al objetivo “Determinar el ajuste individual de los ítems del WHOQOL-BREF y las personas al Modelo de Rasch.”

A diferencia de todos los antecedentes consultados, el presente trabajo no encuentra la necesidad de eliminar ítems (después de ser colapsados a tres categorías)

por inconvenientes con el ajuste de estos a los criterios del modelo Rasch. En los antecedentes, los análisis relacionados con el ajuste mantienen entre 12 y 17 de los 24 ítems calificables del WHOQOL-BREF.

El ajuste de los ítems (y de las personas) se establece a partir de los valores de Infit y Outfit. Los Infit son sensibles a comportamientos inesperados que afectan las respuestas a los ítems cercanos al nivel de "habilidad" exhibido por los individuos; por su parte los Outfit son sensibles a comportamientos inesperados que afectan las respuestas a los ítems lejanos al nivel de "habilidad" exhibido por los individuos (Bond & Fox, 2001). El no encontrar problemas de ajuste de los ítems sobre la muestra estudiada permite inferir que la locación (dificultad) de los ítems dentro de la escala de calidad de vida del WHOQOL-BREF es coherente con el nivel de calidad de vida (habilidad) exhibido por los sujetos.

En relación al ajuste de las personas, no se reporta información en los antecedentes. En este trabajo se encuentran indicadores óptimos de ajuste de las personas y no se da la necesidad de eliminar casos para el análisis. El 91% de los participantes mostraron ajuste aceptable al modelo.

En relación al objetivo "Determinar la unidimensionalidad, confiabilidad y separabilidad de la escala total del WHOQOL-BREF."

El modelo original del WHOQOL-BREF no está planteado para ser unidimensional, el cuestionario fue ideado para medir cuatro dimensiones: Salud física, Salud psicológica, Relaciones sociales y Variables del entorno (World Health Organization, 1998). Este trabajo se planteó como objetivo, validar una aproximación unidimensional a la medición de la calidad de vida a través del WHOQOL-BREF.

Los análisis de unidimensionalidad no arrojaron resultados favorables. La medida extraída de la información explicó el 38.8% de la varianza un 1% menos que la varianza explicada en la base de datos simulada. Se consideran aceptables valores de varianza explicada iguales o mayores al 40% (Bond & Fox, 2001; J. M. Linacre, 2002). No obstante, el análisis de residuales no demostró que los contrastes obtenidos tuviesen capacidad estadística relevante con capacidad de explicar varianza inferior al 10% y un autovalor máximo de 2.6. Más allá de los aspectos técnicos de la unidimensionalidad, los ítems que cargaron el primer contraste agrupados en tres clústeres no guardan entre sí una relación conceptual clara tomando como referencia su procedencia original desde la clasificación original de cuatro dimensiones.

La confiabilidad de los ítems arroja un valor ideal de .99. Sin embargo, cuando se calcula su separabilidad se obtienen 12.19 estratos. El WHOQOL-BREF se compone de 24 ítems lo cual sugiere que algunos de estos generan información diferencial en términos de “dificultad”. Este resultado se puede corroborar gráficamente con el mapa de ítems-personas de la gráfica 12. Estos ítems se distribuyen aproximadamente entre más (Pensamiento y memoria) o menos (Movilidad) un logito y medio a partir de la media teórica de cero logitos.

En relación al objetivo “Evaluar el Funcionamiento Diferencial de los Ítems del WHOQOL-BREF por sexo, reporte subjetivo de enfermedad, edad y nivel de calidad de vida.”

En los resultados se demuestra como ocho ítems muestran funcionamiento diferencial uniforme a partir de los cálculos realizados en Winsteps. De igual forma se demuestra como siguiendo el procedimiento de Benjamini-Hochberg para controlar los falsos positivos en las comparaciones múltiples, se consideran únicamente cuatro ítems como significativos y relevantes para funcionamiento diferencial.

Estos cuatro ítems fueron Dependencia a medicación para la comparación entre enfermos y no seguros versus no enfermos; energía y fatiga para esta misma comparación y de forma similar para los grupos de edad con corte a 30 años. Autoestima mostró funcionamiento diferencial por niveles alto y bajo de calidad de vida y Sentimientos negativos al comprar las dificultades entre hombres y mujeres.

Estas diferencias se desestiman mediante la correlación de la forma original y la forma sin los ítems que presentaron funcionamiento diferencial, lo cual se corrobora con los límites de acuerdo. No obstante esta demostración de la no afectación de estos sesgos sobre la medición Rasch de calidad de vida, los funcionamientos diferenciales en los ítems deben ser considerados en términos de interpretabilidad del constructo a partir de su composición en ítems.

A partir de la demostración de funcionamiento diferencial no se puede concluir que un grupo muestre un nivel promedio significativamente inferior que el otro. Lo que indica el funcionamiento diferencial, es una desventaja en la medición ya que un grupo debe mostrar un nivel de “habilidad” mayor para lograr suplir la “dificultad” impuesta por un ítem (Bond & Fox, 2001; Fischer & Molenaar, 1995; Wilson, 2005). En el caso particular del WHOQOL-BREF, esto sería reinterpretado, como si el grupo desfavorecido por el funcionamiento

diferencial necesitase mostrar un mayor nivel de calidad de vida para evidenciar la cantidad de calidad de vida que representa un ítem dentro de la escala.

De los cuatro ítems que evidencian funcionamiento diferencial tanto por el método matemático como gráfico, dos son conceptualmente razonables y dos son discutibles en posibles razones teóricas.

El ítem Dependencia a medicación¹³ ubicado en el extremo bajo de la jerarquía representada en el mapa de ítems-personas (gráfica 12) muestra una desventaja en contra de los enfermos y quienes no están seguros de estarlo. Esto es lógico ya que es más fácil para una persona sana lograr el bajo nivel de calidad de vida evidenciado por este ítem entre todas las personas. De forma contraria, este ítem de dependencia a medicación requiere que las personas enfermas y quienes no están seguras de estarlo muestren un nivel de calidad de vida mayor para lograr la dificultad exigida por ellos en conjunto con las personas sanas. Si bien los resultados demostraron, como se ha dicho en ocasiones anteriores, que los ítems con funcionamiento diferencial no afectan la estimación total, lo más probable es que la posición del ítem Dependencia a Medicación en el “metro” de calidad de vida para el WHOQOL-BREF en personas enfermas se ubique en un lugar diferente.

Una lógica similar a la expuesta para dependencia a medicación aplica para el ítem Energía y fatiga, el cual también es desfavorable para los participantes enfermos y quienes no están seguros de estarlo y para las personas del grupo de mayores de 30 años para quienes es esperado que sea más difícil demostrar el nivel de Energía para las actividades diarias exigido por este ítem si se consideran en conjunto con los sujetos jóvenes.

A diferencia de estos dos ítems, los funcionamientos diferenciales observados en Sentimientos negativos y Autoestima no son tan claros en términos conceptuales.

El ítem Sentimientos negativos¹⁴ muestra una mayor dificultad entre las mujeres. Esto implica que ellas deban mostrar un mayor nivel de calidad de vida para lograr el nivel exigido por el ítem si se consideran las mujeres en conjunto con los hombres. Podría argumentarse que los hombres son menos susceptibles a los sentimientos negativos logrando evidenciar el nivel de calidad de vida que estos exhiben con mayor facilidad

¹³ En este momento cabe recordar que el ítem es invertido antes de obtener la medición, por lo cual realmente representa no dependencia a medicación.

¹⁴ Invertido de forma similar al ítem Dependencia a medicación.

(menor cantidad de “habilidad”) que las mujeres(Johnson & Whisman, 2013). Sin embargo también se encuentra evidencia que sugiere que estas inferencias son imprecisas y que se debe a definiciones sesgadas de la emocionalidad femenina(Else-Quest, Higgins, Allison, & Morton, 2012).

El caso del funcionamiento diferencial del ítem Autoestima en contra de las personas con bajo nivel de autoestima es extraño, ya que se espera invarianza del parámetro de dificultad a lo largo de la escala de habilidad de las personas(Bond & Fox, 2001; Karabatsos, 2001; Wilson, 2005). No se encuentra alguna razón conceptual particular que explique porque únicamente este ítem incumple este supuesto.

En relación al objetivo “Analizar las relaciones entre la medición de calidad de vida del WHOQOL-BREF y autovaloración de variables de salud.”

El cuestionario WHOQOL-BREF incluye preguntas no calificables que indagan por reporte subjetivo de enfermedad, percepción de calidad de vida y satisfacción con la salud. Al comparar la medida Rasch con estas preguntas se encontraron dos aspectos relevantes de discutir.

La concordancia entre la medida y la percepción de calidad de vida es alta. Con respecto a esta variable solo se afecta el aumento del promedio de medida de calidad de vida en función de la jerarquía conceptual de las categorías en la opción “Ni buena ni mala”. Este resultado respalda la posición de los modelos Rasch de no considerar automáticamente una variable ordinal en función de sus categorías conceptuales(J. M. Linacre, 2002). Es común en las opciones de respuesta en cuestionarios psicométricos permitir opciones intermedias tipo “Ni lo uno ni lo otro”, “ni de acuerdo ni en desacuerdo”, etc. Los cuales en términos técnicos de construcción de medidas no es una práctica favorable(Wilson, 2005).

Los resultados obtenidos con la variable satisfacción con la salud, son aún más extraños. En esta variable se logra un aumento progresivo de la medida Rasch a través de la jerarquía conceptual de las categorías. Sin embargo, la categoría “Muy Insatisfecho” muestra un promedio de medida superior a las categorías semánticamente superiores ubicándose en segundo lugar por debajo de “Muy satisfecho”. Ante este resultado, cabe considerar cierta independencia entre la percepción de salud, la percepción de calidad de vida y el nivel de satisfacción con lo percibido. Sin embargo, esto no puede ser demostrado

a partir de los resultados y existe la posibilidad de que factores extraños no controlados expliquen esta anomalía en los resultados.

En relación al objetivo “Comparar las dificultades obtenidas en los ítems del WHOQOL-BREF con las reportadas por los antecedentes”

Los antecedentes documentados que realizaron validaciones desde el modelo de Rasch se realizaron en diferentes contextos de medición.

Rocha y otros (2012) reporta los resultados de medir la calidad de vida con el WHOQOL-BREF en una muestra internacional de personas con depresión. Al comparar los resultados por ellos reportados con los obtenidos en esta investigación cabe destacar que se encontró acuerdo en ítems del cuestionario de la calidad de vida conceptualmente relacionados con la depresión. Estos ítems son: Sentimientos positivos y negativos, Energía y fatiga y Actividad sexual, los cuales no muestran acuerdo con los otros dos antecedentes analizados (adultos y adultos mayores orientales). Si bien estos resultados no son concluyentes, permiten inferir generalizabilidad de las facetas de la calidad de vida conceptualmente asociadas a la depresión entre personas con y sin depresión. Sin embargo, se parte de la presunción de no depresión entre los participantes del presente estudio, dado que no la reportan al indagar por enfermedades. No se midió depresión en ellos lo cual debilita la interpretación. Estos son los ítems en los que se encontró acuerdo únicamente con este antecedente, en total 9 ítems lo muestran con los resultados obtenidos.

Liang y otros (2009) reportan los resultados de validación sobre una muestra de adultos orientales mayores de 65 años. Al comparar los resultados por ellos reportados con los obtenidos en esta investigación se encuentran cuatro ítems que muestran acuerdo únicamente con este antecedente: Acceso a la información, Transporte, Sueño y descanso y Condiciones de la vivienda; tres de estos cuatro ítems miden variables del entorno. Incluidos estos ítems en total 11 muestran acuerdo con los resultados obtenidos.

Wang y otros (2006) reportan los resultados obtenidos sobre una muestra de adultos orientales entre 20 y 65 años. La comparación con los resultados originales evidencia tres ítems que muestran acuerdo únicamente con este antecedente: Seguridad, Espiritualidad y Capacidad para trabajar. En total 8 ítems muestran acuerdo con los resultados obtenidos.

Los ítems finanzas y actividades placenteras muestran acuerdo tanto con la muestra de depresivos, como con la muestra de adultos mayores. Los ítems Actividades cotidianas

muestran acuerdo con las muestras orientales de adultos y adultos mayores. Solo los ítems Soporte social, Relaciones personales y Dolor y disconfort muestran acuerdo con los tres antecedentes. Tres de estos ítems miden variables de Relaciones sociales.

En relación al objetivo general “Validar el Constructo de Calidad de vida en el WHOQOL-BREF desde el Modelo de Rasch y siguiendo los lineamientos de validación de constructo de Messick estructurados en la metodología de Wolfe y Smith”

En términos generales se considera que hay suficiente evidencia para considerar válida la aproximación unidimensional mediante el Modelo Rasch al cuestionario de calidad de vida WHOQOL-BREF en sujetos provenientes de comunidad. A continuación se dan los argumentos a partir de los lineamientos de validación de Messick siguiendo la metodología de Wolfe y Smith que respaldan esta conclusión al objetivo general de la investigación.

Contenido.

Esta faceta de la validez hace referencia a la “Evidencia de la pertinencia, representatividad y calidad técnica del contenido de los ítems”. A este respecto la literatura reporta una importante cantidad de estudios que demuestran la pertinencia de los ítems del cuestionario para medir calidad de vida. De igual forma, el instrumento cuenta con el respaldo del trabajo del comité de Calidad de vida de la Organización Mundial de la Salud el cual fue el encargado de su construcción (World Health Organization, 1998). En relación a la calidad técnica de los ítems, esta fue demostrada en los análisis y discusión reportados sobre las características de las categorías de respuesta de los ítems con valores de ajuste para el modelo de tres opciones cercanos a uno, aumento monótonico del atributo y coherencia entre las medidas y los puntajes. Con la re categorización a tres opciones de respuesta se garantiza la validez de contenido en términos conceptuales y técnicos.

Sustantiva.

Esta faceta de la validez hace referencia a la “Racionalidad teórica y empírica de la consistencia observada en las respuestas del test”. Los resultados relacionados con el ajuste de los ítems soportan la calidad de estos para considerar que el WHOQOL-BREF en su calificación unidimensional es consistente. Los valores medios de ajuste Infit y Outfit son cercanos a uno y la menor correlación puntual observada fue de .34. Ningún ítem debió eliminarse por problemas de consistencia o ajuste al modelo.

El favorable ajuste de las personas también respalda la validez sustantiva del WHOQOL-BREF ya que el hecho de que las personas respondan de la forma esperada por

el modelo conceptual del cuestionario, es una evidencia de consistencia de los ítems que lo componen. Sin embargo, ya que el modelo original del instrumento no plantea un ordenamiento jerárquico de los ítems no es posible hacer inferencias sobre este aspecto de la validez sustantiva.

Para intentar suplir este vacío teórico de la formulación del WHOQOL para la calidad de vida se muestra en la gráfica el ordenamiento de los ítems del cuestionario según la lógica sistémica propuesta por Shye (1989). Los colores reflejan nivel de dificultad del ítem según convención de colores. Soporte social se ubica en 0 logitos.

		Subsistema			
		Personalidad	Físico	Social	Cultural
Facet: Función	Expresiva	Sentimientos negativos Sentimientos positivos	Actividades cotidianas Capacidad para trabajar Movilidad		Acceso a información
	Adaptativa	Actividades placenteras	Ambiente físico Transporte Condiciones vivienda	Servicios de salud Finanzas	
	Integrativa	Pensamiento y Memoria Autoestima	Energía y Fatiga Dolor y Discomfort Sueño y Descanso Dependencia medicación	Actividad sexual Relaciones Personales	
	Conservativa	Imagen corporal	Seguridad	Soporte social	Espiritualidad

Mayor 1 logito	
media - más 1 logito	
media - menos 1 logito	
menos 2 logitos	

Gráfica 16 Clasificación de los ítems del WHOQOL-BREF según el marco sistémico de Shye (1989)

El esquema permite identificar como los ítems que muestran mayor dificultad en términos de medición de la calidad de vida se agrupan según la clasificación de Shye (1989) en las funciones adaptativas de los subsistemas de personalidad, físico y social con excepción del ítem Condiciones de vida que muestra menor nivel de dificultad. Soporte social, el ítem cuya dificultad se ubica en el punto cero relativo, es el único clasificado como conservativo-social. De forma análoga se observa que los ítems con valores de dificultad menor a 0 se concentran principalmente en las funciones integrativas de los subsistemas físico y social y en la función expresiva del subsistema físico. Las funciones conservativas con excepción de las del subsistema físico muestran en general baja dificultad de los ítems.

Salvando las excepciones del componente expresivo-físico, se puede inferir que las funciones expresivas y adaptativas de la calidad de vida muestran mayores niveles de este atributo que las funciones integrativas y conservativas. En términos de subsistemas, se

observa una distribución similar con facetas que evidencian mayor calidad de vida en lo relacionado con personalidad.

Cabe destacar que el WHOQOL-BREF no mide suficientes aspectos del subsistema cultural (valores), y en baja medida aspectos del subsistema social.

Estructural.

Esta faceta de la validez hace referencia a la “Fidelidad entre la estructura del puntaje y la estructura del dominio del constructo.” A este respecto se intenta una aproximación unidimensional al WHOQOL-BREF, la cual se logra de forma aceptable. Sin embargo, como se plantea en la discusión del objetivo específico pertinente los residuales obtenidos son estadística y conceptualmente irrelevantes. Esto en conjunción con bondad de ajuste favorable hace que se considere suficiente la evidencia a la aproximación unidimensional.

De igual forma, conceptualmente se consideran como irrelevantes estructuras multidimensionales que muestran correlaciones muy altas entre sí por la posibilidad de considerarlas redundantes.(Wilson, 2005). Este suele ser el caso reportado en la literatura para el WHOQOL-BREF y otros instrumentos similares, sin que los investigadores, ni los autores y desarrolladores del instrumento hayan dado razones suficientes que sustente los beneficios de tal forma de medir la calidad de vida.

Generalizabilidad

Esta faceta de la validez “Examina si las propiedades de los puntajes y su interpretación se generalizan en grupos poblacionales.”. En este aspecto se lograron valores importantes de separabilidad tanto para los ítems como para las personas. En relación a la invarianza de la calibración de los ítems se encontró que esta falla en cuatro ítems sin que afecte la medida total obtenida, razón por la cual se considera que se cumple con esta faceta de la generalizabilidad.

En relación a los resultados que comparan las propiedades con grupos poblacionales externos, a través de los antecedentes, se infiere que la generalizabilidad del WHOQOL-BREF, un cuestionario pensado para ser una medida internacional de la calidad de vida, es baja. Principalmente ítems de relaciones sociales se asemejan en sus propiedades a los tres antecedentes; ítems principalmente de entorno a los adultos mayores orientales; y principalmente ítems conceptualmente asociados a la depresión se asemejan

en sus propiedades al antecedente referente a depresivos. Sin embargo, el tipo de análisis y las comparabilidad de las muestras limitan esta interpretación.

Validación externa

Esta faceta de la validez se refiere al “grado con el cual la medida se relaciona con variables externas que miden constructos similares”. Como se evidenció en los resultados y la discusión sobre la relación entre las variables de autovaloración de salud y la medida obtenida con el WHOQOL-BREF, la evidencia de esta faceta de validez se considera aceptable. Los inconvenientes con la graduación ordinal de estas variables tiene una explicación técnica y es poco probable que se deban a una relación inconsistente con la medida obtenida del cuestionario.

Conclusiones

La evidencia permite considerar que a partir del WHOQOL-BREF se puede obtener una medida unidimensional de la calidad de vida que genera una escala de intervalo.

El presente trabajo logra uniformidad en la escala de categorías de respuesta a los ítems luego de colapsarlas en tres opciones, lo cual permite aplicar el modelo Rasch para datos politómicos más sencillo. Los antecedentes reportados en la literatura deben recurrir al modelo de crédito parcial.

En comparación con los antecedentes, la dispersión de las dificultades de los ítems en la escala de logitos genera un rango más amplio, lo cual repercute en mejores propiedades de discriminación.

Si bien la proporción de varianza explicada por la medida Rasch no es muy alta, los componentes obtenidos a partir de los residuales estandarizados no muestran capacidad explicativa relevante. Esto respalda la aproximación unidimensional a la medición de la calidad de vida.

Si bien se observa funcionamiento diferencial en algunos ítems, se demuestra que este no tiene un impacto relevante en la estimación de la medida total de calidad de vida.

Se logra evidencia de criterio externo para la medida de calidad de vida en relación a la percepción de calidad de vida y la satisfacción con la salud. Sin embargo, en relación a la segunda variable se observa una inconsistencia en el ordenamiento.

Se observa para el análisis de la medida con la percepción de calidad de vida que la alternativa de respuesta “Ni pobre ni buena” no respeta la lógica ordinal de medición. Estas opciones intermedias son comunes en las etiquetas semánticas de las escalas Likert y son analizadas de forma ordinal sin demostrar que esta es realmente su naturaleza.

La lógica conceptual que sustenta al WHOQOL-BREF es multidimensional y no jerárquica. Los resultados plantean un modelo unidimensional y jerárquico de las facetas de la calidad de vida que permitirá fortalecer los análisis conceptuales sobre el constructo de calidad de vida.

Se recomienda revisar con mayor detalle la generalizabilidad de la medida del WHOQOL-BREF a partir de análisis Rasch. Este instrumento está pensado para hacer

comparaciones internacionales. Estas comparaciones pueden llevar a conclusiones erróneas si no se garantiza la invarianza de las propiedades de los ítems.

Referencias

- Anderson, K. L., & Burckhardt, C. S. (1999). Conceptualization and measurement of quality of life as an outcome variable for health care intervention and research. *Journal of Advanced Nursing*, 29(2), 298-306.
- Asamblea Mundial de la Salud. (2009a). Resolución A62/9: Comisión sobre Determinantes Sociales de la Salud. Informe de la Secretaría.
- Asamblea Mundial de la Salud. (2009b). Resolución WHA62/R14: Reducir las inequidades sanitarias actuando sobre los determinantes sociales de la salud.
- Balakrishnan, N., & Nevzorov, V. B. (2004). *A Primer on Statistical Distributions*. John Wiley & Sons.
- Ballard, P. B. (1918). The distribution and relations of educational abilities. *The Eugenics Review*, 10(1), 45-47.
- Bartholomew, D. J., & Knott, M. (1999a). *Latent Variable Models and Factor Analysis: Kendall's Library of Statistics 7*. John Wiley & Sons.
- Bartholomew, D. J., & Knott, M. (1999b). *Latent Variable Models and Factor Analysis: Kendall's Library of Statistics 7*. John Wiley & Sons.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. New York: Addison-Wesley.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307-310.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459. doi:10.1007/BF02293801
- Bodek, H. (2013). Facilitating the provision of quality spiritual care in palliative care. *Omega*, 67(1-2), 37-41.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Erlbaum.
- Clayton, D. (1993). *Statistical models in epidemiology*. Oxford ; New York: Oxford University Press.
- Colombia, Republica de. Ley No 100, del 23 de diciembre de 1993. Crea el Sistema de Seguridad Social Integral y se dictan otras disposiciones, Pub. L. No. 100 (1993).

- Condon, B. B. (2013). Honoring Quality of Life. *Nursing Science Quarterly*, 26(2), 124-124. doi:10.1177/0894318413477147
- Darrell Bock, R., & Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika*, 35(2), 179-197. doi:10.1007/BF02291262
- De Leeuw, J., & Verhelst, N. (1986). Maximum Likelihood Estimation in Generalized Rasch Models. *Journal of Educational Statistics*, 11(3), 183. doi:10.2307/1165071
- Departamento Nacional de Planeación. (2011). Plan Nacional de Desarrollo 2010- 2014 «Prosperidad para todos». Tomo I. Imprenta Nacional de Colombia.
- Eggen, T. J. H. M. (2000). On the loss of information in conditional maximum likelihood estimation of item parameters. *Psychometrika*, 65(3), 337-362. doi:10.1007/BF02296150
- Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technology Assessment (Winchester, England)*, 5(4), 1-157.
- Else-Quest, N. M., Higgins, A., Allison, C., & Morton, L. C. (2012). Gender differences in self-conscious emotional experience: a meta-analysis. *Psychological Bulletin*, 138(5), 947-981. doi:10.1037/a0027930
- Embretson, S. E. (2000). *Item response theory for psychologists*. Mahwah, N.J: L. Erlbaum Associates.
- Evans, M., Hastings, N. A. J., & Peacock, J. B. (2000). *Statistical Distributions*. Wiley.
- Fischer, G. H., & Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. Springer.
- Gurin, G., Feld, S., Veroff, J., & University of Michigan. (1975). *Americans view their mental health: March-August, 1957*. Ann Arbor: ISR Social Science Archive, Institute for Social Research. Recuperado a partir de <http://catalog.hathitrust.org/Record/000699200>
- Haas, B. K. (1999). Clarification and integration of similar quality of life concepts. *Image--the Journal of Nursing Scholarship*, 31(3), 215-220.
- Hart, B., & Spearman, C. (1912). General Ability, Its Existence and Nature. *British Journal of Psychology, 1904-1920*, 5(1), 51-84. doi:10.1111/j.2044-8295.1912.tb00055.x
- Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. SAGE Publications.
- Hotelling, H. (1933). *Analysis of a Complex of Statistical Variables Into Principal Components*. Warwick & York.

- Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist*, 17(7), 475-483. doi:10.1037/h0041550
- Johns, S. A. (2013). Translating dignity therapy into practice: effects and lessons learned. *Omega*, 67(1-2), 135-145.
- Johnson, D. P., & Whisman, M. A. (2013). Gender differences in rumination: A meta-analysis. *Personality and Individual Differences*, 55(4), 367-374. doi:10.1016/j.paid.2013.03.019
- Kaasa, S., & Loge, J. H. (2003). Quality of life in palliative care: principles and practice. *Palliative Medicine*, 17(1), 11-20.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of applied measurement*, 2(4), 389-423.
- Kelley, T. L. (1928). *Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities*. Stanford University Press.
- Kleinpell, R. M. (1991). Concept analysis of quality of life. *Dimensions of Critical Care Nursing: DCCN*, 10(4), 223-229.
- Krägeloh, C. U., Kersten, P., Rex Billington, D., Hsu, P. H.-C., Shepherd, D., Landon, J., & Feng, X. J. (2013). Validation of the WHOQOL-BREF quality of life questionnaire for general use in New Zealand: confirmatory factor analysis and Rasch analysis. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 22(6), 1451-1457. doi:10.1007/s11136-012-0265-9
- Linacre, J. (2012a). *A user's guide to WINSTEPS*. winsteps.com.
- Linacre, J. (2012b). WINSTEPS (Versión 3.75.1).
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of applied measurement*, 3(1), 85-106.
- Linacre, M. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, 7, 328.
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2), 247-264. doi:10.1007/BF02291471
- Mandzuk, L. L., & McMillan, D. E. (2005). A concept analysis of quality of life. *Journal of Orthopaedic Nursing*, 9(1), 12-18. doi:10.1016/j.joon.2004.11.001
- Mateus Steffens, C., & Rojas Gualdrón, D. F. (2009). *Del tratamiento moral de las enfermedades mentales a la salud mental comunitaria: Una aproximación histórico-*

epistemológica al concepto de salud mental durante el periodo de higiene mental.
UNAB, Bucaramanga.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. doi:10.1037/0003-066X.50.9.741
- Nunnally, J. C. A., Bernstein, I. H. A., & Arellano, J. A. V. (1995). *Teoría psicométrica*. McGraw-Hill.
- Pearson, K. (1901). *On Lines and Planes of Closest Fit to Systems of Points in Space*. University College.
- Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. En *Contributions to Mathematical Psychology, Psychometrics, and Methodology* (pp. 249–263). Springer.
- Plummer, M., & Molzahn, A. E. (2009). Quality of life in contemporary nursing theory: a concept analysis. *Nursing Science Quarterly*, 22(2), 134-140. doi:10.1177/0894318409332807
- Protección Social, Ministerio de la. (2011). *Indicadores Básicos 2010 Situación de salud en Colombia*.
- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 14, 58-94.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press.
- Roff, M. (1936). Some properties of the communality in multiple factor theory. *Psychometrika*, 1(2), 1-6. doi:10.1007/BF02287999
- Rojas Gualdrón, D. F. (2012). WHOQOL-BREF: ¿Calidad de vida o otra forma de medir la salud percibida? En *Memorias I Congreso Internacional de Psicología: Investigación y Responsabilidad Social*. Bucaramanga.
- Scientific Advisory Committee of the Medical Outcomes Trust. (1995). Instrument review criteria. *Med Outcomes Trust Bull*, 3, i-iv.
- Shye, S. (1989). The systemic life quality model: A basis for urban renewal evaluation. *Social Indicators Research*, 21(4), 343-378. doi:10.1007/BF00303952
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8(1), 33. doi:10.1186/1471-2288-8-33

- Spearman, C. (1904). «General Intelligence,» Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201-292. doi:10.2307/1412107
- Spearman, C. (2008). *The Abilities of Man: Their Nature and Measurement*. Pierides Press.
- Spearman, C. (2010). The proof and measurement of association between two things. *International Journal of Epidemiology*, 39(5), 1137-1150. doi:10.1093/ije/dyq191
- Taillefer, M.-C., Dupuis, G., Roberge, M.-A., & May, S. L. (2003). Health-Related Quality of Life Models: Systematic Review of the Literature. *Social Indicators Research*, 64(2), 293-323.
- Taylor, R. M., Gibson, F., & Franck, L. S. (2008). A concept analysis of health-related quality of life in young people with chronic illness. *Journal of Clinical Nursing*, 17(14), 1823-1833. doi:10.1111/j.1365-2702.2008.02379.x
- The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. (1995). *Social Science & Medicine*, 41(10), 1403-1409. doi:10.1016/0277-9536(95)00112-K
- Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *Journal of Educational and Behavioral Statistics*, 27(1), 77-83. doi:10.3102/10769986027001077
- Thomson, G. H. (1920). General versus group factors in mental activities. *Psychological Review*, 27(3), 173-190. doi:10.1037/h0068628
- Thorndike, E. L., Lay, W., & Dean, P. R. (1909). The Relation of Accuracy in Sensory Discrimination to General Intelligence. *The American Journal of Psychology*, 20(3), 364-369. doi:10.2307/1413366
- Thurstone, L. (1936). A new conception of Intelligence and a new method of measuring primary abilities. *Educational Record*, 17 (Suppl. 10), 124-138.
- Thurstone, L. (2011). *The Vectors of Mind Multiple Factor Analysis for the Isolation of Primary Traits*. BiblioBazaar.
- Thurstone, L. L. (1947). *Multiple-factor analysis: a development and expansion of The vectors of the mind*. The University of Chicago Press.
- Velicer, W. F., & Jackson, D. N. (1990). Component Analysis versus Common Factor Analysis: Some issues in Selecting an Appropriate Procedure. *Multivariate Behavioral Research*, 25(1), 1-28. doi:10.1207/s15327906mbr2501_1
- Vernon, P. E. (1950). *The structure of human abilities*. Methuen.

- Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Routledge.
- Wolfe, E. W., & Smith, E. V., Jr. (2007a). Instrument development tools and activities for measure validation using Rasch models: part I - instrument development tools. *Journal of Applied Measurement*, 8(1), 97-123.
- Wolfe, E. W., & Smith, E. V., Jr. (2007b). Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement*, 8(2), 204-234.
- World Health Organization. (1946). Constitution of the World Health Organization.
- World Health Organization. (1991). World Health Statistics Annual.
- World Health Organization. (1997). Measuring Quality of Life.
- World Health Organization. (1998). WHOQOL. Users Manual. Division of Mental Health and Prevention of Substance Abuse.
- Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis: [Rasch Measurement]*. Mesa Press.