



UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL
www.ufrgs.br



FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS: PSIQUIATRIA
TESE DE DOUTORADO

**Teoria de Resposta ao Item: Aplicação do modelo Rasch em
desenvolvimento e validação de instrumentos em saúde mental**

EDUARDO CHACHAMOVICH

Porto Alegre, Dezembro de 2007.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS: PSIQUIATRIA
TESE DE DOUTORADO

**Teoria de Resposta ao Item: Aplicação do modelo Rasch em
desenvolvimento e validação de instrumentos em saúde mental**

Tese apresentada ao Programa de Pós-
Graduação em Ciências Médicas:
Psiquiatria como requisito parcial para a
obtenção do grau de Doutor.

EDUARDO CHACHAMOVICH

Orientador: Prof. Dr. Marcelo Pio de Almeida Fleck

Porto Alegre, Dezembro de 2007.

Catalogação-na-Publicação

C431 Chachamovich, Eduardo
Teoria de resposta ao item: aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental / Eduardo Chachamovich. – 2007.
288 f.

Tese (doutorado) – Programa de Pós-Graduação em Ciências Médicas: Psiquiatria, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

Orientador: Prof. Dr. Marcelo Pio de Almeida Fleck.

1. Saúde mental 2. Psicometria 3. Modelo de Rasch 4. Estudos de validação I. Fleck, Marcelo Pio de Almeida II. Título

NLM WT 145

*“Most researchers use statistics the way a drunkard uses a lamp-post:
more for support than illumination”*

(Winifred Castle)

AGRADECIMENTOS

Aos meus pais Júlio e Jane, e aos meus irmãos Régis e Lúcio. Aos que representam, antes de tudo, a origem deste sonho e a possibilidade de que ele tenha o significado que possui ... e que me ensinam que a vida é música.

A minha esposa Juliana, com quem reparto cada etapa deste sonho e de todos os demais.

As famílias Luzardo Rigol, particularmente Dario Carlos, Gladys, Eliana e Marcos, pelo acolhimento carinhoso e pela alegria que trouxeram aos momentos de conquistas.

Ao meu orientador Prof. Marcelo Fleck, que vem me guiando nas sucessivas etapas de minha formação acadêmica e profissional. Por ser um modelo de cientista , pesquisador e orientador e por irrestritamente apoiar-me e incentivar-me nas trilhas do mestrado e do doutorado. É um privilegio poder contar com tua amizade durante os dez anos de desenvolvimentos de pesquisas com o Grupo WHOQOL ! Muito Obrigado ...

Aos colegas do Centro Brasileiro do Grupo de Qualidade de Vida WHOQOL, pelo convívio, pela amizade e pelas importantes contribuições ao longo da trajetória do doutorado.

Aos amigos Jacques Zimmermann e Marcelo Berlim, pelos internacionais momentos de agradável convívio e filosofias científicas (e outras nem tanto ...)

À colega e amiga Clarissa Trentini, pela colaboração inestimável na condução das etapas iniciais desse trabalho. Muito obrigado!

Ao prof. Mick Power e sua equipe (Doug, Nuno, Suzanne, Tim, Sarah e Ciara), por me abrirem as portas da Universidade de Edimburgo e por me acolherem calorosamente nesta cidade “tão sem atrativos” !! Thanks, guys !!

Ao Programa de Pós-Graduação em Psiquiatria, em especial a Patrícia Ribeiro e ao Prof. Paulo Abreu, pelo incansável apoio e incentivo na viabilização das importantes etapas deste doutorado.

À CAPES e à Pró-Reitoria de Pós-Graduação da Universidade Federal do Rio Grande do Sul, pelo prestimoso auxílio na realização do estágio de doutorado no exterior. É um orgulho poder desfrutar de tal possibilidade, proporcionada por órgãos públicos brasileiros.

Aos queridos amigos Luciana Bridi, Yuri Restano e Krisin Lewis, por repartirem momentos tão importantes e por estarem sempre presentes e dispostos. Muito grato!

Aos amigos do Grupo de Estudo em Fotografia do Projeto Contato, pelas valiosas janelas na alma a mim proporcionadas e pelas oportunidades de crescimento que elas determinam.

SUMÁRIO

Resumo	10
Abstract	13
Lista de Tabelas	15
Lista de Figuras	17
Lista de Gráficos	19
1. Apresentação	20
2. Revisão de Literatura	
2.1 Mensuração e Psicometria	21
2.2 Modelo de Rasch	52
2.3 Aplicações do Modelo de Rasch em Psicometria	69
3. Referencias Bibliográficas da Introdução	87
4. Questões de pesquisa, objetivos e hipóteses	102
5. Proteção de direitos humanos	106
6. Artigo 1 : The Brazilian WHOQOL-OLD Module version: A Rasch analysis of a new measure	107
7. Artigo 2: Development and validation of the Brazilian version of the Attitudes to Aging Questionnaire (AAQ): an example of merging classical psychometric theory and the Rasch measurement model	137
8. Artigo 3: A multi-point Likert Scale proved not to be suitable for illiterate subjects	163

9. Artigo 4: The cultural influence in measuring depression of older adults: Is GDS-15 a suitable instrument in Brazil?	197
<hr/>	
10. Considerações finais	216
<hr/>	
11. Anexos	
11.1 Artigo anexo: Impact of major depression and subsyndromal symptoms on Quality of Life and Attitudes to Aging in an international sample of older adults	218
11.2 Termo de Consentimento	254
11.3 Ficha de dados sócio-demográficos	256
11.4 Escala de avaliação sócio-demográfica	264
11.5 WHOQOL-BREF	269
11.6 WHQOL-OLD	276
11.7 Geriatric Depression Scale	278
11.8 Atitudes Frente ao Envelhecimento	287
11.9 Aprovação da Comissão Científica e Comissão de pesquisa e Ética em Saúde do HCPA	326
<hr/>	

RESUMO

Introdução: A Teoria de Resposta ao Item constitui-se em um novo grupo de modelos psicométricos de construção, validação e avaliação de instrumentos de aferição em desfechos em saúde. Dentre tais modelos, o Modelo de Rasch se destaca como uma possibilidade de testar a Teoria de Mensuração Conjunta e, portanto, desenvolver escalas de medida com estrutura intervalar. De fato, as robustas estratégias estatísticas utilizadas recentemente demandam que as variáveis apresentem propriedades intervalares para que resultados fidedignos sejam obtidos. Na área da saúde mental, o Modelo de Rasch agrega a possibilidade de que os construtos sejam avaliados quanto à dimensionalidade. Os instrumentos podem ser explorados quanto à invariância dos itens, calibragem, adequação das escalas de resposta e adequação dos itens ao modelo teórico subjacente.

Objetivos: Explorar a aplicação do Modelo de Rasch no desenvolvimento de instrumentos de qualidade de vida e atitudes frente ao envelhecimento. Utilizar o Modelo de Rasch na validação cultural de um instrumento de depressão geriátrica, assim como testar a adequação de uma escala de resposta Likert de acordo com o nível educacional, através da aplicação do Modelo de Rasch.

Método: Dados internacionais de qualidade de vida, atitudes frente ao envelhecimento e depressão foram utilizados nestes estudos. Os dados são provenientes do Projeto WHOQOL-OLD e AAQ, coordenado pela Organização Mundial de Saúde em 20 países. Uma combinação de Teoria Clássica em

Psicometria e Modelo de Rasch foi utilizada para a análise dos dados. O Modelo de Rasch foi implementado através do software RUMM2020.

Resultados: A utilização do Modelo de Rasch possibilitou um refinamento dos instrumentos testados, gerando instrumentos com novos formatos mais adequados. No instrumento WHOQOL-OLD, em sua versão em Português, dois domínios dos seis apresentaram interação item-total insuficiente. O remodelamento da escala de resposta e a exclusão de itens de performance inadequada acarretaram melhora da escala. O instrumento Atitudes Frente ao Envelhecimento também necessitou de exclusão de itens e agrupamento de categorias de resposta, em sua versão em Português, para que seu desempenho psicométrico fosse adequado. Em relação ao nível de escolaridade, a análise dos dados gerados pelo instrumento WHOQOL-BREF em uma amostra internacional de idosos por Modelo de Rasch evidenciou que uma escala Likert de cinco pontos não é adequada para sujeitos analfabetos, em oposição aos achados com sujeitos universitários. A escala de depressão geriátrica GDS também foi analisada em uma amostra brasileira. O formato original de 15 itens se mostrou inadequado, e a exclusão de itens é fundamental para que o desempenho deste instrumento seja satisfatório. Por fim, o impacto de sintomas depressivos clínicos e subclínicos na qualidade de vida de idosos foi avaliado. Mesmo níveis subclínicos de depressão mostraram-se significativamente associados a decréscimo de qualidade de vida e atitudes frente ao envelhecimento em uma amostra de 5566 idosos.

Conclusão: A aplicação do Modelo de Rasch no desenvolvimento e validação de instrumentos na área da saúde mental mostrou-se marcadamente relevante. Tal modelo proporcionou que as escalas possam ser examinadas em detalhes. A avaliação dos itens de modo individual permite que a sua adequação seja determinada e que o instrumento seja refinado de acordo com as modificações sugeridas pelas análises de Rasch. Por fim, o Modelo de Rasch permite o teste de invariância dos itens, que assume particular importância na utilização transcultural de instrumentos na área da saúde mental.

Descritores: Modelo de Rasch, psicometria, Teoria de Resposta ao Item, saúde mental.

ABSTRACT:

Background: Item Response Theory represents a novel group of psychometric models for constructing, validating and evaluating health outcome instruments. Among these models, the Rasch Model is able to test the Theory of Conjoint Measurement and, so, develop intervalar scales. In fact, the recently applied robust statistical strategies demand that the variables have an intervalar structure, so that reliable results are obtained. In the mental health field, the Rasch Model assures that the constructs may be tested for dimensionality. The instruments may be explored regarding item invariance, targetting, response scales adequacy and item fit to the underlying theoretical model.

Objectives: To explore the role of the Rasch Model for the development of quality of life and attitudes to aging instruments. In addition, to apply the Rasch Model in the cultural validation of a geriatric depression scale, as well as test the suitability of a 5-point Likert response scale according to the educational level.

Method: International data on quality of life, attitudes to aging and depression were analyzed in these studies. Data were drawn from the WHOQOL-OLD and AAQ Project, coordinated by the World Health Organization in 20 countries. A combination of Classical Test Theory and Rasch Model was used to data analysis. The Rasch Model was implemented by RUMM2020 software.

Results: The application of the Rasch Model determined a refinement of the original instruments, deriving more adequate versions. The Portuguese version of the WHOQOL-OLD instrument presented insufficient item-total interaction in two domains out of six. Rescoring the response scale and deleting items led to

improvement of the instrument. The Attitudes to Aging Questionnaire (Portuguese version) has also demanded rescoring and item deletion in order to present adequate psychometric performance. Regarding the educational level, the Rasch analysis of data obtained using the WHOQOL-BREF in an international sample of older adults indicated that a 5-point Likert scale is not suitable for illiterate subjects. In opposite, this Likert scale is suitable for graduate older adults. The Geriatric Depression Scale was explored in a Brazilian sample. Its original 15-item format has proven to be inadequate, and item exclusion is crucial to obtain a satisfactory performance. Finally, the impact of clinical and subclinical depressive symptoms in quality of life was assessed. Even subthreshold symptomatology is significantly associated to impairments in quality of life and attitudes to aging in a sample of 5566 older adults.

Conclusion: The application of the Rasch Model to the development and validation of mental health instruments has proven to be highly relevant. This model brings the possibility of examining scales in depth. The individual assessment of each item makes it possible to determine item fit and, thus, to refine the instruments according to the alterations claimed by the Rasch analysis. Finally, the Rasch Model implements the item invariance test, which is particularly important for utilizing instruments in transcultural investigations

Keywords: Rasch Model, psychometrics, Item Response Theory, mental health.

Lista de Tabelas

1. Classificação de variáveis de Stevens	29
2. Exemplificação da Teoria de Mensuração Conjunta a partir de massa e volume	56
3. Probabilidade de acerto em um encontro de habilidades vs dificuldade de itens	58
4. Ordenamento de dificuldade e habilidades	60
5. Tabela 1 (artigo 1) – Sample demographics (n=424)	132
6. Tabela 2 (artigo 1) – WHOQOL-OLD items descriptions	133
7. Tabela 3 (artigo 1) – Fit statistics for the refined WHOQOL-OLD Module	134
8. Tabela1 (artigo 2) - Socio-demographic characteristics of Brazilian and International Samples	158
9. Tabela 2 (artigo 2) - Ancova analyses including Educational level, gender and depression between Brazilian and International Samples	159
10. Tabela 3 (artigo 2) – Descriptive analysis of the set of 38 items in the Brazilian sample (n=424)	160
11. Tabela 4 (artigo 2) – Rasch Analysis of the original 24-item final version including the 5-point Likert response scale	161
12. Tabela 5 (artigo 2) - Final 22-item version, including the rescored 4-point response scale	162
13. Tabela 1 (artigo 3) - Sample characteristics (n=1048)	189
14. Tabela 2 (artigo 3) – Item-trait interaction statistics for illiterate and graduate samples	190
15. Tabela 3 (artigo 3) – Description of item fit statistics in the two subsamples (5-point response scale)	191
16. Tabela 4 (artigo 3) - Description of item fit statistics in the two subsamples (rescored into 3-point response scale)	192
17. Tabela 1 (artigo 4) – Demographics	212

18. Tabela 2 (artigo 4) – Fit of 15-item GDS to Rasch model	213
19. Tabela 3 (artigo 5) – Fit of the 10-item GDS to Rasch model	214
20. Tabela 1 (artigo anexo) – Total sample characteristics (n=4316)	238
21. Tabela 2 (artigo anexo) - Ancova results of the interaction of Gender, Education Level and Marital Status for each domain	239
22. Tabela 3 (artigo anexo) – Description of R ² changes in the hierarchical regressions for each WHOQOL-OLD, WHOQOL-BREF and AAQ domains	240
23. Tabela 4 (artigo anexo) – Standardized Beta Coefficients for each variable in the final model	241
24. Tabela 5 (artigo anexo) – Demographics of the subsyndromal sample (n=3187)	242
25. Tabela 6 (artigo anexo) - Comparison of means between subsamples	243
26. Tabela 7 (artigo anexo) – Description of R ² changes in the hierarchical regressions for each WHOQOL-OLD, WHOQOL-BREF and AAQ domains (subsyndromal sample)	244
27. Tabela 8 (artigo anexo) – Standardized Beta Coefficients for each variable in the final model (subsyndromal sample)	245

Lista de Figuras

1. Tipologia das escalas de Stevens	27
2. Ilustração gráfica da Condição de Arquimedes	38
3. Exemplo de medidas arbitrárias em duas escalas de satisfação marital	44
4. Exemplo de escalas com estrutura ordinal e intervalar	45
5. Transição entre escalas ordinais e intervalares	51
6. Exemplo de Curva Característica do Item gerada pelo software RUMM2020	65
7. Representação de simultâneas Curvas Características de Item	67
8. Distribuição adequada de categorias de resposta	77
9. Distribuição inadequada de categorias de resposta (<i>threshold disorder</i>)	77
10. Ilustração gráfica de DIF uniforme	80
11. Mapa de itens-pessoas gerado pelo software RUMM2020	82
12. Funcionamento de um teste tipo CAT	85
13. Figura 1 (artigo 1) – Item 5 Category Probability Curves in original response scale and after rescoring	135
14. Figura 2 (artigo 1) – Person item distribution graph for the WHOQOL-OLD scale	136
15. Figura 1 (artigo 2) - Scree Plots illustrating EFA Eigenvalues for Brazilian and International Samples	156
16. Figura 2 (artigo 2) - Final 24-version model tested	157
17. Figura 1 (artigo 3) – CFA of the final model	193
18. Figura 2 (artigo 3) – Threshold Map for the Physical Domain (illiterate subsample on the top and graduate subsample at the bottom)	194
19. Figura 3 (artigo 3) – Category Probability Curves for item 26 in its original 5-point scale and rescored (illiterate sample)	195
20. Figura 1 (artigo 4)- Targeting of the 10-item GDS	215

21. Figura 1 (artigo anexo) – Impact of depression levels on
QOL and AAQ scores

246

Lista de Gráficos

1. Crescimento do número de artigos indexados sobre
Modelo de Rasch na base PubMed 70

1. Apresentação

A presente Tese de Doutorado tem como foco principal a aplicação do Modelo de Rasch no desenvolvimento e validação de instrumentos de pesquisa na área da saúde. Concentra-se basicamente na investigação dos constructos qualidade de vida, depressão e atitudes frente ao envelhecimento. A partir de tais constructos, pretende-se expor a utilização de uma moderna técnica estatística na construção de novas medidas e na validação de medidas existentes consagradas. Em seu corpo, a presente Tese é composta de:

- a) uma introdução, abordando os conceitos de mensuração e psicometria, uma revisão do Modelo de Rasch e sua aplicação em psicometria.
- b) proposição das questões de pesquisa;
- c) cinco artigos científicos publicados ou submetidos a periódicos, abordando a aplicação do Modelo de Rasch e a testagem do desempenho psicométrico de uma série de instrumentos de pesquisa utilizados em idosos, bem como o impacto da depressão na qualidade de vida desta população.
- d) considerações finais
- e) anexos

2. Introdução: Revisão da Literatura

2.1 Mensuração e Psicometria

A necessidade de mensurar fenômenos psicológicos abstratos tem registros na literatura científica desde o século XIX. Na psicologia moderna, o físico alemão G.T. Fechner deu início às investigações sobre medição de sensações em 1860, com a publicação de seu livro intitulado *Elemente der Psychophysik*. Neste, o cientista, interessado principalmente em compreender as regras naturais que regem as percepções, propôs uma série de experimentos com o objetivo de mensurar a intensidade das sensações frente a estímulos pré-determinados (Michell, 2005). O “experimento das esferas de mármore” tornou-se um clássico e serviu como base para o desenvolvimento das teorias que perdurariam por cerca de meio século. Nele, os sujeitos deveriam segurar na palma da mão diferentes esferas de mármore, todas com as mesmas características externas (cor e tamanho), porém com massas diferentes. Fechner, assim como diversos psicólogos que o seguiram, buscava elucidar uma fórmula ou função matemáticas que explicassem a relação entre a massa da esfera e a sensação correspondente (Michell, 2005).

De fato, os escritos de Fechner são considerados o marco inicial da operacionalização da tentativa de medir fenômenos psicológicos. Entretanto, o reconhecimento da necessidade de quantificar tais fenômenos está presente em um tempo bastante mais remoto, tanto na área da psicologia quanto na da filosofia (Michell, 1999). O filósofo russo Immanuel Kant (1724-1804) ensinava que as leis da natureza conteriam mais propriedades compreensíveis na medida em que incorporassem as operações matemáticas. Mesmo filósofos mais céticos, como Nietzsche, consideravam que “o conhecimento filosófico tinha se tornado tão científico que seria capaz de aplicar números e medidas” (tradução livre, 1968, p 378). Em relação à área da psicologia, as primeiras proposições da necessidade de mensuração dos fenômenos psicológicos datam do século XIV, com o estudioso Nicole Oresmer. Todavia, o desenvolvimento do conhecimento em tal momento histórico ainda não propiciava maiores avanços.

A partir da publicação de Fechner, a exploração da necessidade e da metodologia da medição de constructos psicológicos ganhou impulso considerável. Em 1890, James McKeen Cattell, um dos primeiros professores norte-americanos da cátedra de Psicologia, escreveria que a Psicologia nunca apresentaria a exatidão e a certeza de uma ciência exata, a menos que estivesse embasada em experimentos e mensuração (Cattell, 1890, Apud Michell, 2005). Seguindo a mesma linha de Cattell, Charles Spearman (que viria a ser extensamente reconhecido como o pai das análises fatoriais e da proposição de cálculos alternativos para correlações não-paramétricas) publicou um livro em 1937 (*Psychology down the ages*) em que afirmava que os métodos

experimentais em Psicologia só poderiam alcançar o status de ciência quando estivessem substanciados por um fator vital e crucial chamado mensuração (Michell, 1999).

Tornava-se crescente o movimento científico em direção ao avanço das estratégias de mensuração de fenômenos psicológicos. Desde a publicação dos escritos de Cattell e Spearman, o desenvolvimento de medidas em Psicologia parecia ser um campo em plena consolidação (Michell, 1999). A virada do século XIX para o século XX testemunhou um acúmulo de publicações e teorizações acerca do processo de mensuração na Psicologia Moderna. Dentre estes, destacam-se os de Titchener e Baldwin. O primeiro, em 1905, definiu mensuração como a comparação de uma dada magnitude com a unidade básica da mesma característica. Determinar-se-ia, pois, o número de vezes que a unidade está contida na magnitude. Percebe-se, em tal conceituação, a presença da propriedade de aditividade, que viria a ser extensamente estudada pelas técnicas modernas de psicometria, especialmente a Teoria de Rasch (a partir dos anos 1960).

Baldwin incluiu, em 1902, o vocábulo mensuração em seu Dicionário de Filosofia e Psicologia, reforçando em sua conceituação o caráter aditivo de partes-referência para a adequada medição de um determinado atributo (Baldwin, 1902). Juntamente ao caráter aditivo de uma certa medida, ambos os autores destacaram também a possibilidade de calcular-se razões entre o atributo a ser medido e a unidade referência (Michell, 1999). Mais recentemente, Terrien (1972) discute tais princípios, indicando que quantidades são conceitos

abstratos e que possuem duas propriedades: podem ser medidos (o que significaria que se podem estabelecer razões numéricas através de experimentos) e podem ser expressados por leis da Física e por modelos matemáticos. Em notação matemática, Terrien propõe que,

$$\text{Quantidade de atributo} = \text{número puro} \times \text{unidade},$$

sendo o número puro aquele diretamente medido no experimento em questão e a unidade, a parte-referência.

Ainda que algumas das bases teóricas para a mensuração já estivessem sendo estudadas e publicadas, o processo de medição invariavelmente recai sobre a representação numérica de magnitudes de atributos psicológicos. Em outras palavras, as partes de atributos psicológicos, tais como depressão e ansiedade (ou ainda inteligência, atenção e aprendizado) devem ser adequadamente quantificadas para que se possa aplicar os conceitos de aditividade e razão.

Michell (2005) veementemente destaca a inabilidade dos psicólogos em geral para definir adequadamente mensuração, ainda que os livros em Psicologia apresentem uma grande série de tentativas de fazê-lo.

Stanley Smith Stevens publicara, em 1946, a definição de mensuração que passaria a ser o modelo conceitual mais extensamente adotado e citado. Professor do Departamento de Psicologia da Universidade de Harvard, seu

artigo intitulado “*On the theory of scales of measurement*” (publicado originalmente no periódico Science) traria contribuições que marcariam a psicometria por longo tempo (Stevens, 1946).

A primeira delas diz respeito à classificação de variáveis. Stevens descreveu esquematicamente que as variáveis medidas deveriam ser classificadas em nominais, ordinais, intervalares e de razão. Ademais, propôs, em seu artigo, definições para cada classificação, bem como informou quais estatísticas descritivas e testes poderiam ser aplicados a variáveis de cada categoria. As indicações de testes estatísticos a cada categoria de variáveis viriam a ser, posteriormente, revisadas e criticadas à luz de teorias modernas em psicometria (tais tópicos serão abordados em maior detalhe nas seções Modelo de Rasch e Aplicações do Modelo de Rasch em Psicometria).

Ainda que a divisão esquemática de variáveis de acordo com suas características seja classicamente atribuída a Stevens, as bases conceituais do entendimento das variáveis têm uma história bastante mais longa. De fato, escritos de Aristóteles e Euclides, há mais de 2000 anos, já indicavam a diferenciação da natureza de distintas variáveis. Aristóteles, em especial, propôs claramente a distinção entre variáveis com características quantitativas e qualitativas. Em relação à observação empírica de objetos, colocou-se contrário à idéia de que todas as características de um determinado objeto seriam passíveis de ser quantificadas. Propriedades, tais como cor e temperatura (de acordo com suas palavras), não seriam

propensas a adição ou soma, compondo um grupo de qualidades e não de quantidades (Michell, 1999).

Stevens ampliou sobremaneira as bases do conhecimento proposto por Aristóteles e forneceu diretrizes para a classificação da natureza de variáveis. Resumidamente, descreveu as variáveis nominais como sendo um agrupamento de características sem nenhuma ordenação de magnitude (tais como cor, nome e sexo). Tais variáveis não se prestam a operações matemáticas que necessitem de ordenamento (seja de postos ou de números reais), mas somente para descrição simples de freqüências e testes de categorias. As variáveis ordinais, de acordo com o autor, seriam caracterizadas por apresentarem ordenação, mas não magnitude invariável de incremento. Como ilustração, indicou a dureza de minerais e números de ruas, descrevendo que se podem ordenar os minerais em uma escala de dureza, mas não é possível quantificar a diferença de dureza entre eles (Stevens, 1946; Wright, 1997).

As variáveis intervalares são descritas como aquelas passíveis de serem ordenadas em intervalos conhecidos. Mais além das variáveis ordinais, a magnitude do incremento entre um ponto a e um ponto b em uma escala intervalar é conhecido e uniforme. A temperatura medida em graus Celsius é apontada como um exemplo de variável intervalar por Stevens.

Por fim, variáveis de razão são consideradas aquelas que apresentem um zero conhecido e real, ou seja, um ponto de partida conhecido que deve ser

mantido durante o processo de mensuração. Temperatura medida em graus Kelvin é apontada como um exemplo de escala de razão.

A figura 1 abaixo ilustra esquematicamente a natureza das quatro classificações de variáveis, como proposta por Stevens (1946).

Figura 1. Tipologia das escalas de Stevens.

Stevens' Scale Typology						
Number	3	575	23	12		
Order	1	2 3 4		5	6	7
Interval	-5	-2	-1	0	1	2 3
Ratio	10^{-2}	10^{-1}	10^0	10^1	10^2	10^3

Retirada de Wright (1997).

Mais do que somente descrever as categorias de variáveis, Stevens se propôs a indicar quais as estratégias de descrição e de análises estatísticas seriam adequadas às quatro categorias. A Tabela 1 abaixo, embasada na apresentada por Wright (1997) e incluindo informações oferecidas por Wilson (2005), descreve esquematicamente as diretrizes indicadas por Stevens.

Os escritos originais de Stevens já destacavam uma dúvida científica não respondida na ocasião. Os escores brutos de uma escala de medida eram considerados pelo autor como uma medida ordinal. Stevens, assim como

seus seguidores, interpretam que a simples soma de itens de uma escala fornece um indicativo de ordem de intensidade de algum traço (como depressão, ansiedade ou inteligência), mas não dá informações exatas a respeito da magnitude de tal traço. Entretanto, Stevens não descreveu conclusões acerca de considerar os escores padronizados de uma determinada escala de medida como variáveis intervalares. A tabela proposta pelo autor indica um ponto de interrogação em tal casela (reproduzido na Tabela 1 abaixo), denotando que esta permanecia uma questão a ser explorada. Este tópico seria abordado em profundidade pelos teóricos das diferentes escolas da Teoria de Resposta ao Item (Wilson, 2005).

Tabela 1. Classificação de variáveis de Stevens.

Escala	Operação	Exemplos	Localização	Dispersão	Associação	Teste
Nominal	Igualdade	Números de jogadores	Moda			Qui-quadrado
Ordinal	Maior ou menor	Dureza de minerais Números de rua Escores brutos	Mediana	Percentil	Correlação de ordem (rank)	Testes de ranking
Intervalar	Distância	Temperatura Celsius Tempo Escores padronizados ?	Média aritmética	Desvio Padrão	Correlação produto momento	t-teste F-teste
Razão	Razão	Distância Temperatura Kelvin	Média geométrica Média harmônica	% de variação		

A segunda contribuição de Stevens em tal artigo foi a definição de mensuração propriamente dita. O autor definiu mensuração como sendo “a atribuição de um numeral a objetos e eventos de acordo com uma regra” (tradução livre). O texto original em inglês refere que “*measurement is the process of assigning numerals to objects or events according to rules*” (Stevens, 1946, p 668). Em publicações subseqüentes, por vezes o autor incluiria “qualquer regra” no lugar de “uma regra” (Michell, 2005). O termo numeral difere em seu significado de número. O primeiro é definido como palavras, letras ou ainda símbolos que representem números (quantidades) reais. Já números são descritos como a quantidade de magnitude real em si (Webster’s Encyclopaedic Unabridged Dictionary, 1996; Michell, 2005)

Se a contribuição sobre a classificação das variáveis permanece sendo relevante e prestigiada por aqueles que lidam com escalas em pesquisa, a conceituação de mensuração tem rendido diversas críticas a Stevens (Bond e Fox, 2007; Wilson, 2005; Michell, 2005; Andrich, 1978; Andrich, 1988). Entre as mais contundentes, a de Duncan (1984) destaca que a definição de Stevens acarretou marcada confusão ao desenvolvimento das ciências sociais quantitativas. Bond e Fox (2007) também se aliam às ásperas críticas a Stevens e citam que os psicólogos costumam lidar com meras estatísticas descritivas como se fossem de fato mensurações científicas de fenômenos psicológicos.

Entretanto, provavelmente um dos maiores críticos da conceituação de Stevens seja o psicólogo Joel Michell, da Universidade de Sidney. Em um artigo extenso enfocando a definição de mensuração no campo das ciências quantitativas e Psicologia, Michell cunha o termo “*methodological thought disorder*” para designar aqueles que aderem à conceituação de Stevens sem refletir sobre as sérias limitações científicas que esta apresenta (Michell, 1997).

Um bloco sólido de autores teóricos, estudiosos da mensuração quantitativa em ciências sociais e da saúde, propõe críticas sistemáticas às conceituações de mensuração de Stevens. Quais são as bases científicas de tais críticas? Sobre que tópicos residem os questionamentos acerca da conceituação de mensuração, tal como descrita por Stevens?

A primeira divergência conceitual aborda a atribuição de numerais (*assigning numerals*, no texto original em inglês). Michell (2005) aponta uma clara discrepância entre a visão clássica de mensuração e aquela dada por Stevens, caracterizando as duas definições como mutuamente incompatíveis. Na visão tradicional de mensuração, nenhum numeral é atribuído; todavia, as relações métricas são exploradas e descobertas. Mais do que um mero detalhe lingüístico, a diferença de conceituações reflete um princípio psicométrico de base. Atribuir o numeral 6 ao comprimento de uma parede em função de uma regra determinada, por exemplo, não guarda relação teórica com aditividade ou razão. Ao contrario, chegar à conclusão de que a mesma parede tem 6 metros de comprimento, uma vez que ela possui seis vezes a unidade referência de medida pré-estabelecida (metro), proporciona relação de medida real com a

unidade-padrão. Assim sendo, possibilita ao cientista aplicar operações de adição e razão (por exemplo, quantas vezes a parede de seis metros é maior do que a parede de três metros). Em outros termos, Michell descreve que a mensuração relaciona-se a descobrir relações numéricas reais (razões) entre objetos (magnitude de atributos), e não à tentativa de construir relações numéricas a partir de convenções arbitrárias. Tal distinção pode ser mais marcadamente exemplificada em casos de variáveis não quantitativas. Ainda que uma regra pudesse ser a atribuição de números consecutivos a quatro diferentes cores (sendo a mais escura de valor um e a mais clara de valor quatro) – observando assim os preceitos de mensuração propostos por Stevens – não há nesta série real relação numérica entre as variáveis. Mensuração neste cenário seria impossível à luz da visão clássica de medição (Blanton e Jaccard, 2006). Em última instância, há uma oposição entre determinar regras às quais as variáveis devem se submeter (de acordo com Stevens) ou descobrir as regras reais que explicam as relações métricas de uma determinada variável (Bond e Fox, 2007). A regra pré-estabelecida pelo pesquisador, segundo a escola de Stevens, assume uma importância mais destacada do que a própria medida real. Suen (1990) coloca críticas a esta inversão, ao referir que um dos principais problemas na medição em Psicologia é que praticamente cada autor pode desenvolver sua própria regra de avaliação para atribuir números, inviabilizando comparações válidas.

O segundo ponto crucial de discordância entre as duas teorias diz respeito às possibilidades de mensuração. De acordo com Stevens, todos os atributos

psicológicos são mensuráveis, e a única regra não permitida no processo de mensuração é a atribuição aleatória de números (Stevens, 1968). Por outro lado, as teorias clássicas de mensuração ditam que somente os atributos que guardam estrutura quantitativa são passíveis de mensuração, uma vez que esta é a única que sustenta os princípios de aditividade e razão (Embretson, 2006). Assim sendo, os preceitos de Stevens não demandam que a variável seja examinada previamente em relação à sua natureza quantitativa. Qualquer objeto ou evento poderia ser medido, desde que uma regra pré-determinada seja desenvolvida e seguida, independentemente de se comportar de modo quantitativo ou não. De acordo com a escola de mensuração clássica (de Fechner, Cattell e Spearman), a definição de Stevens promove uma mudança de foco no processo de mensuração, retirando a ênfase da relação numérica entre as observações e sublinhando a importância da regra de medida (Michell, 1986). Em outros termos, a teoria de Stevens criaria confusão entre duas práticas distintas, quais sejam mensuração (no sentido clássico do termo) e codificação numérica. Basicamente, a escola clássica entende mensuração como a exploração de fatos empíricos de natureza intrinsecamente numérica, em oposição à codificação numérica de eventos ou objetos (como proposto por Stevens) (Michell, 1999).

Como supracitado, há extensas e numerosas críticas e discordâncias entre as escolas clássica e de Stevens. Psicometricistas modernos, em oposição às idéias de Stevens, tendem a embasar seus estudos nos preceitos da escola clássica. Assim sendo, que proposições de conceituação e embasamento

teórico os autores modernos oferecem? Como o entendimento do processo de mensuração, a partir da visão dos autores modernos, pode promover mudanças no campo da saúde mental? E que aprimoramentos metodológicos são necessários para acompanhar a mudança de paradigma de pensamento em relação à mensuração de fenômenos psicológicos? Tais questionamentos servirão de base para o transcorrer desta Introdução, uma vez que terão como objetivo conduzir a revisão de conhecimentos em direção à incorporação das técnicas modernas de psicometria no campo da saúde mental.

Que proposições de conceituação e embasamento teórico os autores modernos oferecem?

Talvez o mais destacado teórico no campo da mensuração, no período pós-Stevens, tenha sido Patrick Suppes. Filósofo graduado na Universidade de Columbia (em Nova Iorque, Estados Unidos), Suppes traria contribuições que causariam grande impacto no modo pelo qual a mensuração em Psicologia era vista. Em parte, o seu campo inicial de estudo (física e meteorologia) seria responsável por um posicionamento científico não comprometido pela visão de Stevens, nem tampouco pelas demandas da prática clínica como psicólogo (Michell, 2005). Luce, outro autor de destacada relevância no desenvolvimento das teorias em psicometria, escrevera que Suppes teria afetado as teorias em mensuração mais do que qualquer outra pessoa (Luce, 1979). Em 1954, Suppes propunha que axiomas eram indispensáveis, através dos quais o autor buscava aproximar os conceitos de observações qualitativas (por exemplo, a barra de ferro

A é mais longa do que a barra de ferro B) das asserções quantitativas necessárias às teorias em mensuração (por exemplo, o comprimento desta barra de ferro é exatamente 5,6 cm) (Suppes, 1954). A principal idéia da proposição de axiomas era oferecer uma série de diretrizes testáveis na prática e, em última instância, desenvolver uma especificação matemática de um modelo através do qual as observações se relacionam entre si (função matemática de relação entre as observações). Krantz (1964) viria a se associar ao desenvolvimento de axiomas para o processo de mensuração. Entretanto, ao estudar em profundidade o Teorema de Hölder, originalmente publicado em 1901, Krantz estabeleceu diretrizes claras, testáveis e reprodutíveis para a detecção de variáveis com características verdadeiramente quantitativas.

O primeiro passo para a identificação de variáveis quantitativas é testar o ordenamento destas. Espera-se que variáveis quantitativas sejam passíveis de serem ordenadas de acordo com sua magnitude. Assim sendo, três expressões devem ser satisfeitas. Imaginemos que X , Y e Z sejam quaisquer três valores de uma variável Q . Então Q é uma variável ordinal se, e somente se:

- 1 Se $X \geq Y$ e $Y \geq Z$, então $X \geq Z$ (transitividade)
- 2 Se $X \geq Y$ e $Y \geq X$, então $X = Y$ (anti-simetria)
- 3 Tanto $X \geq Y$ ou $Y \geq X$ (conectividade forte)

Uma relação que apresente essas três propriedades é chamada de ordem simples, sendo Q uma variável ordinal. Todas as variáveis quantitativas devem

necessariamente satisfazer as três condições apresentadas acima. Entretanto, nem todas as variáveis ordinais (que satisfaçam tais condições) serão quantitativas. Em outras palavras, todas as variáveis quantitativas são também ordinais, mas o inverso não é verdadeiro.

Para que uma variável seja considerada quantitativa, o princípio da aditividade tem que ser também satisfeito. Aditividade é composta de uma relação triádica, que pode ser representada por $X+Y=Z$ (Krantz, 1964). Imaginemos que Q seja uma variável ordinal tal que, para quaisquer valores de X , Y e Z ,

- 4 $X + (Y + Z) = (X + Y) + Z$ (associatividade)
- 5 $X + Y = Y + X$ (comutatividade)
- 6 $X \geq Y$ se e somente se $X + Z \geq Y + Z$ (monotonicidade)
- 7 Se $X > Y$ então há um valor de Z tal que $X = Y + Z$ (solubilidade)
- 8 $X + Y > X$ (positividade)
- 9 Existe um número natural n tal que $nX \geq Y$ (onde $1X=X$ e $(n+1)X = nX+X$)
(condição de Arquimedes)

Na hipótese de a variável Q satisfazer as condições acima, então a relação triádica envolvida é de aditividade, e a variável é quantitativa (Krantz, 1964; Michell, 1999).

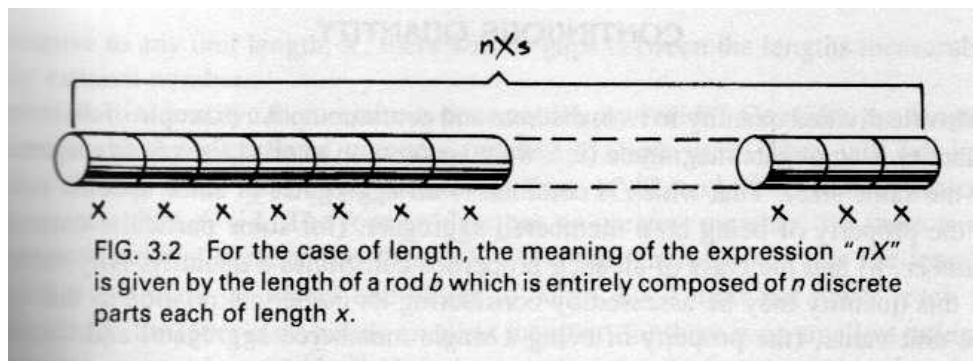
Em termos não matemáticos, a condição 4 indica que o valor composto pelas partes discretas X e $(Y + Z)$ é o mesmo de o composto pelas partes discretas Z e $(X + Y)$. A condição 5 reflete que o valor da parte composta por $(X + Y)$ é igual à

soma dos valores de X e de Y . Assim, tomadas em conjunto, as condições 4 e 5 apontam que a ordem das partes que constituem um valor de uma variável quantitativa é irrelevante.

A condição 6 afirma que se $X \geq Y$, então um valor inteiramente composto das partes discretas X e Z é pelo menos igual ao valor inteiramente composto pelas partes discretas Y e Z . A condição 7 prevê aditividade completa, uma vez que se $X > Y$, então há necessariamente uma parte Z complementar. Já a condição 8 se destina a garantir a positividade de todos os valores ou partes.

A condição 9 foi nomeada em homenagem ao matemático Grego Arquimedes (nascido em 287 AC e falecido em 212 AC). Basicamente, esta condição afirma que, independentemente da magnitude de um determinado valor Y , sempre haverá um número discreto n de partes X compondo a magnitude Y . A figura 2 abaixo ilustra graficamente a função matemática subjacente a esta condição.

Figura 2. Ilustração gráfica da condição de Arquimedes.



Retirada de Michell (1999, p 55)

A partir da operacionalização de parâmetros para a aferição da característica quantitativa de uma variável, a atribuição de números a um fenômeno ou objeto passou a ser considerado um processo mais minucioso e crucial do que Stevens propunha. De fato, os psicometricistas passavam a focar sobremaneira o processo de investigação do comportamento da variável, e não mais a regra arbitrária pela qual um pesquisador atribuiria numerais a um objeto ou evento. Ao contrário do sugerido por Stevens, os números atribuídos não mais representariam símbolos indicativos de uma determinada magnitude (numeral), mas sim relações reais entre grandezas.

De acordo com Michell (1999), mensuração é definida como o procedimento para identificar valores de variáveis quantitativas a partir de sua relação numérica com outros valores. Ou seja, mensuração seria o reconhecimento de relações numéricas entre magnitudes distintas. Ao medir determinada variável quantitativa, o valor a ser conhecido seria indicado como r vezes a unidade conhecida de referência (relação entre a quantidade a ser medida X e o valor de referência Y , assim que $X / Y = r$).

O processo de mensuração necessita que sejam desenvolvidos procedimentos para que os valores X e Y sejam comparados, e sua razão seja aferida. Estes procedimentos são genericamente chamados de *quantificação*, e requerem que uma condição seja satisfeita: a variável tem de necessariamente apresentar estrutura quantitativa e, assim sendo, ser adequada ao princípio da aditividade. Qualquer mensuração que envolva variáveis não quantitativas (por exemplo, ordinais) não deve ser aceita como tal (Michell, 1999; Pallant et al, 2006; Karabatsos, 2001).

Seguindo a definição de mensuração e a necessidade da presença de aditividade no processo de quantificação, Luce e Tuckey (1964) tiveram especial destaque a partir da publicação da Teoria de Mensuração Conjunta (*Theory of Conjoint Measurement*) especificamente para quantificação em ciências sociais. Esta teoria oferece uma via alternativa para a detecção da estrutura quantitativa de uma determinada variável, não sendo necessário que as nove condições descritas por Krantz sejam matematicamente testadas. Alternativamente, a Teoria de Mensuração Conjunta possibilita que a estrutura quantitativa seja detectada a partir de relações ordinais acerca de uma variável. Brevemente, a teoria testa a estrutura de uma variável P como sendo uma função não-iterativa (ou seja, aditiva ou multiplicativa) entre duas outras variáveis A e B . A teoria propõe testes com relações aditivas entre as três variáveis ($P = A + B$) e/ou multiplicativa ($P = A \times B$). De modo geral, a aplicação da Teoria de Mensuração Conjunta tem seu papel quando nenhuma das três variáveis é conhecida (a variável P a ser testada, e as variáveis A e B a serem auxiliares nesse teste). Através de uma série de testes aditivos e multiplicativos (não tomados em conjunto), o princípio da teoria é

demonstrar que se as variáveis A e B são transformadas simultaneamente, então a variável P também deve sofrer a mesma transformação para que uma estrutura quantitativa e aditiva seja mantida (Andrich, 1988). Os testes aditivos e multiplicativos se dividem em três categorias básicas (cancelamento duplo, solubilidade e condição de Arquimedes).

A operação de cancelamento duplo indica que se um certo par de valores de P são ordenados de forma crescente, então quaisquer outros pares de valores devem obedecer a mesma regra. O princípio do cancelamento duplo acaba por ratificar a condição de que $P = A + B$.

Solvabilidade diz respeito à interação de A e B . Este princípio requer que as duas variáveis sejam complexas o suficiente para compor diferentes valores de P . Ademais, exige que A e B apresentem distribuição homogênea.

Por fim, a condição de Arquimedes, como anteriormente citado, faz referência ao fato de que as variáveis devam ser compostas de n partes discretas, e que nenhuma das partes deve ser infinitamente maior do que as outras partes discretas. Os três princípios estão embasados em equações matemáticas e axiomas (Krantz, 1964).

O resultado da aplicação da Teoria de Mensuração Conjunta a uma certa variável é a determinação de sua estrutura quantitativa e aditiva. Assim sendo, este constitui um passo indispensável para qualquer tentativa de mensuração de fenômenos psicológicos. Faz-se necessário que as variáveis a serem estimadas em investigações científicas tenham sua natureza conhecida pelo pesquisador, como modo de garantir que as operações estatísticas a serem aplicadas estão de acordo com as possibilidades oferecidas pela característica da variável (Pallant e

Tenant, 2007; Tenant et al, 2004; Karabatsos, 2001). Este tópico será novamente abordado no capítulo 2.2 (Modelo de Rasch).

Como o entendimento do processo de mensuração, a partir da visão dos autores modernos, pode promover mudanças no campo da saúde mental?

Investigações na área da saúde mental baseiam-se fundamentalmente em informações advindas de escalas de medida (Pallant e Tenant, 2007). A mensuração é considerada como um dos fundamentos da pesquisa em Psicologia, e as medidas são aplicadas na testagem de teorias, no treinamento e balizamento das práticas clínicas (Blanton e Jaccard, 2006). Os fenômenos psicológicos não são passíveis de mensuração direta (através de unidades de medida concretas, tais como metro ou litro) (Pallant e Tenant, 2007; Chachamovich, 2007). Pesquisas em disciplinas de laboratório apresentam dificuldades significativamente menores em relação à medição, uma vez que os processos subjetivos não estão presentes em grande extensão, e as questões de validade e reprodutibilidade são aprimoradas com auxílio de novas ferramentas tecnológicas (Streiner e Norman, 2003). O crescimento da complexidade das pesquisas científicas nas últimas duas décadas e a incorporação de parâmetros subjetivos em investigações têm sido notáveis.

Por exemplo, o Food and Drug Administration (FDA), órgão norte-americano responsável por regulamentar e aprovar a utilização de novos fármacos nos Estados Unidos, exige que os protocolos de investigação de novas drogas incluam medidas de eficácia e tolerabilidade subjetivas, referidas a partir da visão dos

pacientes (os chamados *Patient Reported Outcomes* ou *PROs*). Assim sendo, as medidas objetivas de desfecho, tais como mortalidade e morbidade, não são mais suficientes para que um novo medicamento ou intervenção seja avaliado (Wiklund, 2004; Cella et al, 2007; Reeve et al, 2007; The PROMIS Group, 2007; Chang, 2007).

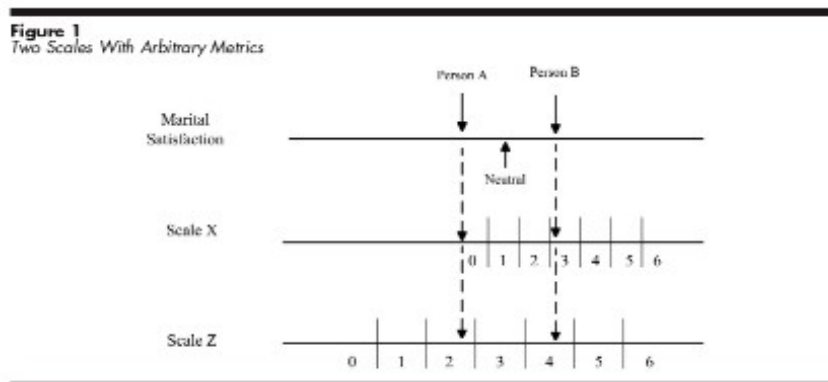
A valorização crescente de medidas não-objetivas em várias áreas da saúde, e a dificuldade de se obter mensurações válidas e confiáveis nesse campo específico do conhecimento compõem um cenário desafiador. Grande parte do conhecimento gerado nas áreas da saúde mental é desenvolvida a partir de escalas de medida (Streiner e Norman, 2003; Bowling, 1991). De fato, as escalas podem ser decisivas na obtenção e interpretação dos resultados em diversos desenhos de investigações.

Marshall et al (2000) exemplificam tal impacto ao relatar uma análise de 300 ensaios clínicos com pacientes esquizofrênicos. Os autores indicam que a utilização de escalas não publicadas previamente confere uma chance mais elevada (RR 1,94; CI: 1,35-2,79) de os fármacos testados demonstrarem efetividade, quando comparados aos ensaios que incluem escalas validadas e publicadas. Seguindo a mesma linha de investigação, Charmann et al (2003) demonstraram que distinções na composição de escalas e na metodologia de construção de medidas de severidade em dermatite atópica são responsáveis por diferenças significativas nos resultados de ensaios clínicos e estudos abertos. Tais diferenças têm grande impacto nas práticas clínicas baseadas em evidências em dermatologia, e o uso de medidas não adequadamente validadas é fonte de vieses significativos.

A construção de escalas de medida em desfechos de saúde exige que uma série de parâmetros sejam adequadamente testados. Entre eles, destacam-se a validade, confiabilidade e estabilidade (Streiner e Norman, 2003; Wilson, 2005). Resumidamente, tais verificações têm como objetivo detectar se a medida é estável (ou seja, se fornece resultados semelhantes frente a uma mesma quantidade de traço latente), se é confiável (se reflete consistentemente o constructo a ser medido) e se é válida (ou seja, se o teste mede o que ele teoricamente se destina a medir) (Field, 2005; Streiner e Norman, 2003; Banton e Jaccard, 2006; Kazdin, 2006).

Ainda que a avaliação de confiabilidade e de validade seja indispensável para a utilização de novas medidas em saúde mental, o pesquisador desta área depara-se com outro desafio: a arbitrariedade dos escores gerados (Blanton e Jacard, 2006; Bond e Fox, 2007). O sistema métrico adotado por uma nova medida é geralmente arbitrário e, por consequência, de difícil interpretação. Duas medidas hipotéticas válidas e confiáveis (destinadas a medir, por exemplo, depressão) podem gerar escores em diferentes parâmetros métricos, o que impossibilita a comparação entre ambas. A figura 3 ilustra o exemplo destas escalas:

Figura 3. Exemplo de medidas arbitrárias em duas escalas de satisfação marital



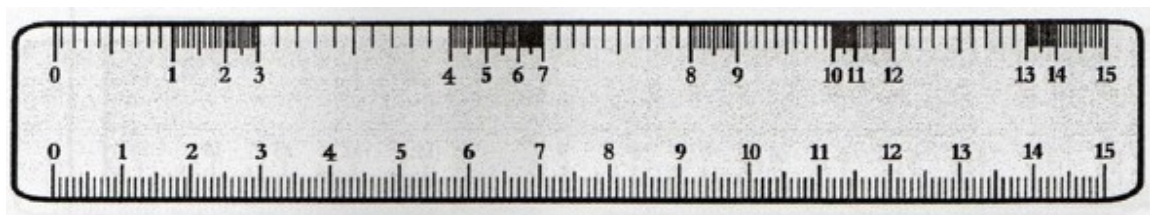
Retirada de Blanton e Jaccard, 2006

Nota-se, no exemplo acima, que a linha superior indica a quantidade real de satisfação marital, enquanto as escalas X e Z destinam-se a estimar tal constructo. Ambas variam de 0-6, mas os escores são diferentes para uma mesma quantidade de satisfação marital, como evidenciam as duas pessoas ilustradas na figura. Um outro ponto a ser destacado nesta figura é a inferência da magnitude da mudança real do traço latente a partir das escalas. Ainda que ambas as escalas apresentem 7 pontos, a distância entre cada ponto não é semelhante entre elas. Deste modo, uma mudança de um ponto na escala Z não é comparável à mudança de um ponto na escala X. Um ponto da escala Z representa mais quantidade de traço latente do que um ponto da escala X.

A segunda questão, também de extrema importância para a interpretação dos escores gerados a partir de escalas, diz respeito à composição interna dos escores. Não somente a interpretabilidade dos escores entre escalas deve ser assegurada, mas também as variações intra-escala necessitam ser avaliadas.

Tomemos como exemplo uma escala hipotética que avalia depressão, variando de 0 a 15 pontos. A variação de um ponto (por exemplo, de 3 para 4) significa a mesma variação de um ponto em outra localização da mesma escala (por exemplo, de 10 para 11)? Em outras palavras, os pontos são homogeneamente distribuídos ao longo da escala? A figura 4 exemplifica duas escalas hipotéticas para medir um determinado constructo. Ainda que ambas variem de 0-15, observa-se que a escala posicionada na parte superior da régua apresenta ordenamento, mas os intervalos não são equidistantes. Já na escala localizada na metade inferior da régua, os números são ordenados e simétricos.

Figura 4. Exemplo de escalas com estruturas ordinal e intervalar.



Retirada da apostila do curso de Introduction to Rasch Analysis, *Psychometric Laboratory for Health Sciences, Universidade de Leeds*.

As interpretações dos resultados gerados pelas duas escalas não são idênticas. Os intervalos da escala superior, dependendo da localização em que se posicionam, não representam a mesma quantidade de traço latente.

Como citado anteriormente, a conceituação moderna de mensuração inclui a exploração das relações métricas entre a magnitude de uma quantidade e uma unidade padrão da mesma quantidade (Michell, 2005). Ademais, requer que a

quantidade (ou constructo, na área da saúde mental) seja quantitativa e permita o princípio da aditividade. No exemplo acima, somente a escala inferior permite aditividade, uma vez que os intervalos são regulares (equidistantes) (Karabatsos, 2001; Bond e Fox, 2007; Fischer e Molenaar, 1995).

De fato, a régua da Figura 4 ilustra exemplos de uma escala ordinal (posicionada na parte superior) e uma escala do tipo intervalar (na parte inferior). Tal como descrito por Stevens (1946), a escala ordinal apresenta numerais alocados ordenadamente que refletem a ordem crescente de traço latente ao variar de 0 a 15. Entretanto, como destacado pelo autor, os numerais são representações de ordem, mas não expressam quantidade. Os numerais são indicativos de ranking, e poderiam ser substituídos por letras ou símbolos (como, por exemplo, variando de *a* até *o*) sem prejuízos (Embretson e Reise, 2000).

Já a escala intervalar indica equidistância entre intervalos. Assim sendo, possibilita aditividade e comporta-se como uma estrutura quantitativa. Nesse caso, os números não são simples símbolos, mas de fato permitem operações matemáticas e estatísticas paramétricas (Tenant, 2004; Pallant e Tenant, 2007; Cicchetti et al, 2006).

Wright (1989) dedicou-se ao estudo do comportamento dos dados obtidos por escalas, em comparação aos esperados em relação a uma medida intervalar. De modo categórico, o autor intitula seu artigo de “*Observations are always ordinal; measurements, however, must be interval*”. No desenvolvimento de seu posicionamento, Wright indica que há uma confusão vigente entre a soma de um score (o que chama de contagem) e o processo de mensuração de determinado traço ou habilidade. Mais além, o autor (entusiasta do Modelo de Rasch), lista

algumas características que credenciam tal modelo como capaz de realizar a transição de dados observados a medidas intervalares.

Seguindo a linha de Wright, diversos autores têm se dedicado a demonstrar as diferenças nas estimativas de escores quando escalas ordinais são tratadas como intervalares (Kelly et al, 2007; Conrad et al, 2004; McHorney et al, 1997; Andrich, 2004; Tenant et al, 1996; Tutz, 2003; Cicchetti et al, 2006; Wright e Linacre, 1989; Tastle e Wierman, 2006). De modo geral, tais autores realizaram comparações entre estimativas obtidas por escores brutos (*raw scores*) e por Análise de Rasch, após a transformação das escalas ordinais em intervalares. Os achados são unânimes em destacar as vantagens advindas das operações com escalas intervalares, ou seja, maior precisão, menor erro padrão e menor impacto de *missing values* na performance das estimativas.

A indicação do uso de estatísticas paramétricas em variáveis ordinais é largamente criticada na literatura internacional (Svensson, 1998; Claesson e Svensson, 2001; Kuzon et al, 1996). Merbitz et al (1989) destacam que variáveis ordinais não suportam as operações matemáticas necessárias para cálculos de média, desvio-padrão ou tamanho de efeito. Wittkowsi et al (2004) enfatizam o uso indiscriminado de operações paramétricas em investigações médicas e ressalta que as suposições necessárias para a utilização destas estatísticas não raro não são observadas. Em relação às análises multivariadas de dados, Hair et al (2006) também salientam que regressões paramétricas (tais como lineares ou hierárquicas) requerem variáveis intervalares ou de razão.

Song et al (2006), ao explorar métodos estatísticos alternativos na análise de pacientes pós-AVC avaliados pela Escala de Barthel, indicam que, por sua

natureza de *ranking* e pela assimetria dos intervalos entre categorias, variáveis ordinais são adequadas somente para análises não-paramétricas. Outros autores também se alinham à noção da limitação do uso de estatísticas paramétricas a variáveis intervalares (Dombret et al, 2003; Svensson, 1998; Bandelow et al, 1998).

Ainda que escalas de mensuração em saúde mental ofereçam medidas ordinais, e que estas não se prestem ao uso de estatísticas paramétricas, a utilização de operações não-paramétricas não parece ser larga. Munzel e Bandelow (1998) investigaram a opção pela aplicação de análises paramétricas e não-paramétricas em ensaios clínicos de tratamentos de Transtorno de Pânico diagnosticados pelo DSM-III e DSM-III-R. Os achados revelaram que somente 16,7% dos artigos incluíram estratégias não-paramétricas, tendo o restante optado pela indicação de estratégias paramétricas.

Que aprimoramentos metodológicos são necessários para acompanhar a mudança de paradigma de pensamento em relação à mensuração de fenômenos psicológicos?

Os avanços no estudo teórico do processo de mensuração, bem como o reconhecimento da inadequação do uso de variáveis ordinais em estatísticas paramétricas, apontam a necessidade de que o campo da saúde mental possa desenvolver medidas mais fidedignas. Ademais, tais medidas devem também ser adequadas às suposições de testes estatísticos mais refinados.

A noção de que as medidas no campo da saúde mental serão tanto mais fortes quanto mais próximas dos números que expressam quantidade não é recente. Em 1908, Hermann Ebbinghaus escrevera que “os brilhantes resultados alcançados nas ciências naturais através de mensuração e cálculos sugerem a idéia de que algo similar pode ser feito pela psicologia” (tradução livre, Michell, 1999, p7). Mais além de Ebbinghaus, Thomson escrevera, em 1891, que um conhecimento que não pode ser medido e expressado em forma de números está em um estágio inicial de desenvolvimento e, provavelmente, representa um esboço de ciência. De fato, Thomson foi um dos maiores representantes da escola teórica de Pitágoras, que entendia que todos os componentes da natureza eram compostos de números (Michell, 2005; Michell, 1999). O impulso de Pitágoras para a quantificação matemática de virtualmente todos os fenômenos naturais norteou o rumo das investigações científicas durante séculos, tendo sido influência para importantes pensadores como Platão, Descartes e Galileu (Michell, 2005). Entre outras contribuições, Pitágoras desenvolveu modelos matemáticos de exploração dos sons, o que resultou nas escalas musicais ainda atualmente utilizadas (Sirvent, 2003; Tiburcio, 2002).

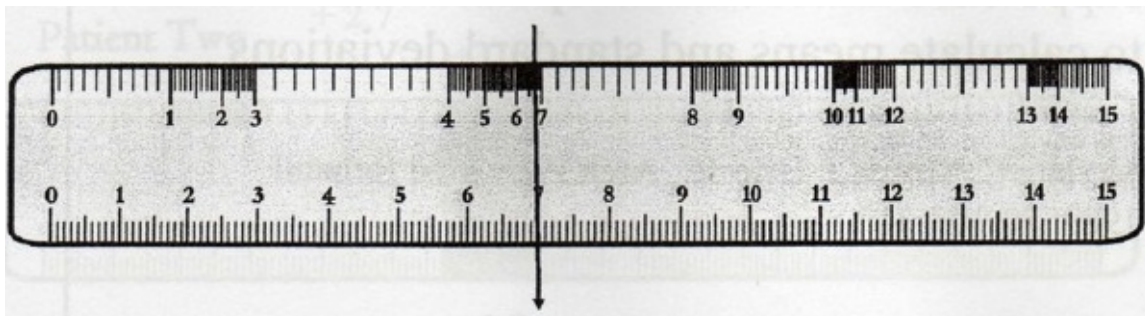
Estratégias estatísticas de análise de dados na área da saúde têm sofrido marcada evolução recentemente. Modelos de análises multivariadas de dados, particularmente úteis em pesquisas envolvendo constructos complexos e multideterminados, têm sido refinados (Hair et al, 2006). O desenvolvimento de softwares para executar tais análises também contribui para o crescente uso de tais estratégias. Kline (2005) destaca a importância da Modelagem de Equações

Estruturais (*Structural Equation Modeling - SEM*) no exame de uma série de relações de dependência simultaneamente. Ademais, destaca que a habilidade de incorporar variáveis latentes ao modelo traz uma contribuição marcante para as técnicas de análise multivariada de dados. Hair et al (2006) corroboram a visão de Kline e sustentam que a incorporação da SEM no armamentário estatístico moderno representa um importante ganho científico. A disponibilização de programas de computador com interfaces mais acessíveis (tais como AMOS® e MPlus®), e que não exigem que o usuário produza as sintaxes dos testes, expandiu sobremaneira o uso de tais análises (Kline, 2005).

Entretanto, a oportunidade de incorporar métodos mais robustos de análises estatísticas às investigações em saúde mental pressupõe a existência de medidas quantitativas confiáveis (Merbitz et al, 1989). Se esta condição não for adequadamente observada, corre-se o risco de aplicar técnicas estatísticas avançadas a dados inconsistentes, o que produziria resultados insustentáveis.

Retomando a figura 4 como exemplo, a seguinte questão se impõe: Como fazer a transição de uma escala ordinal (na parte superior da régua) para uma intervalar? A figura 5 ilustra tal tópico.

Figura 5. Transição entre escalas ordinais e intervalares.



Retirada da apostila do curso de Introduction to Rasch Analysis, *Psychometric Laboratory for Health Sciences, Universidade de Leeds*

Uma recente e importante contribuição das áreas da matemática e estatística para o desenvolvimento científico da saúde mental é a possibilidade da construção de medidas quantitativas, aditivas e intervalares. Os métodos chamados de Teoria de Resposta ao Item (TRI) constituem uma coleção de técnicas de modelagem para a análise de dados em nível de itens (Edelen e Reeve, 2007). Os diferentes modelos de TRI propiciam métodos para que algumas das tradicionais limitações das escalas em saúde mental sejam reparadas, especialmente a impossibilidade de comparação de escores inter-escalas, a obtenção de escores intervalares e o desenvolvimento de versões abreviadas de instrumentos originais (mantendo uma performance psicométrica desejada) (Embretson e Reise, 2000).

Os reflexos da utilização de medidas quantitativas intervalares (após transformação por TRI), ao invés de medidas ordinais, têm sido demonstrados na literatura internacional. Norquist et al (2004) reportaram as diferenças entre a computação de dados por escore bruto (*raw score*) e por Modelo de Rasch. O

segundo modelo conferiu ganhos de precisão e de estabilidade de itens ao longo do tempo em pacientes avaliados um ano após realização de cirurgia de quadril.

Stucki et al (1996) investigaram as diferenças entre estimar a magnitude da mudança em escores da subescala de atividade física do instrumento SF36 de acordo com a estratégia de análise de dados. Os achados obtidos pelo Modelo de Rasch diferiram significativamente dos obtidos por escore bruto. Os autores concluem que variações na escala ordinal refletem magnitudes distintas em relação ao traço subjacente (atividade física). Achados semelhantes foram descritos por Fitzpatrick et al (2004) em pacientes submetidos a cirurgia de quadril e avaliados em 3 e 12 meses no período pós-operatório. A comparação dos escores ordinais obtidos por escala Likert com os intervalares, obtidos por Modelo de Rasch, indica que os últimos apresentam maior precisão na discriminação de grupos em relação à satisfação com o procedimento.

Entre os diversos modelos de Teoria de Resposta ao Item, o Modelo de Rasch constitui um dos mais extensamente utilizados em investigações em saúde (Bond e Fox, 2007). Uma das suas principais aplicações é a capacidade de desenvolver e/ou refinar medidas em escalas intervalares e, assim, potencialmente candidatas às várias estratégias de análise paramétricas.

2.2 Modelo de Rasch

O Modelo de Rasch foi assim denominado em homenagem ao matemático Dinamarquês Georg William Rasch. Nascido a 21 de setembro de 1901 em Odense (Dinamarca), Georg teve seu interesse por matemática despertado aos

doze anos com a leitura dos livros de navegação de seu pai, que incluíam capítulos de geometria plana e esférica. Após a conclusão do ensino médio (*high school*), Rasch deu início aos estudos formais na Faculdade de Matemática da Universidade de Copenhague em 1919. Até a finalização de sua Tese de Doutorado, Georg Rasch não teria nenhum contato com análises estatísticas. Seu trabalho até então se restringia ao campo das funções matemáticas básicas, especialmente as funções *Gama* (Andersen, 2001).

De fato, Rasch passou a ter envolvimento incipiente com as ciências da saúde por acaso. Em 1934, Rasch foi convidado a dar algumas opiniões acerca de curvas fisiológicas de cérebros de coelhos por alguns colegas. A partir de então, Rasch seria convidado regularmente a coordenar grupos de estudos de cientistas interessados em estatística e matemática.

O treinamento formal em estatística se daria em 1935 e 1936, a partir do recebimento de uma bolsa para que Rasch estudasse em Londres com o importante estatístico Sir Ronald Aylmer Fisher. A intensa influência de Fischer, aliada a uma liberdade que Rasch teria em seu futuro emprego no *Military Psychology Group* na Dinamarca, comporia um cenário propício para o desenvolvimento da Teoria que tornaria Georg Rasch um autor de extrema importância no campo da estatística aplicada à área da saúde. Após grandes contribuições na área da psicometria, Rasch morreu aos 79 anos de idade, no ano de 1980 (Olsen, 2003).

As bases do Modelo de Rasch foram inicialmente publicadas em 1960, em seu livro intitulado "*Probabilistic models for some intelligence and attainment tests*" (Rasch, 1960). Em síntese, o novo modelo romperia com as bases da Teoria de

Testes Clássica (TCC) e proporia novos métodos para o desenvolvimento e análise de testes. Entre as características do novo modelo, os itens e testes não dependeriam mais da amostra de sujeitos aos quais eles se destinam. Ademais, a estimativa da quantidade de habilidade (no caso do livro pioneiro de Rasch, inteligência e grau de aquisição acadêmica) também seria independente do conjunto de itens a serem ministrados (Andrich, 1988; Hays et al, 2000; Baker, 2001).

A principal contribuição do Modelo de Rasch diz respeito ao entendimento de que um sujeito A, com habilidade maior do que um sujeito B, deve ter uma probabilidade superior de responder acertadamente a um determinado item. Paralelamente, um item com grau de dificuldade C deve ser mais freqüentemente acertado por um sujeito com habilidade D, quando comparado a um item E de maior dificuldade (Rasch, 1960). Em outros termos, Wright e Panchapakesan (1969) descreveram que o resultado do encontro entre um sujeito X e um item Z é o produto da habilidade do sujeito e da dificuldade do item, e nada mais. Tal noção reforça a propriedade de independência da amostra e dos itens que caracteriza o Modelo de Rasch.

O Modelo de Rasch representa uma estratégia matemática capaz de testar a Teoria de Mensuração Conjunta de Luce e Tuckey e, assim sendo, proporciona a derivação de medidas quantitativas, intervalares e aditivas (Bond e Fox, 2007). Como anteriormente citado, o pioneiro trabalho de Luce e Tuckey (1964) abriu caminho para a verificação matemática do comportamento quantitativo de variáveis abstratas, tão freqüentes nas ciências da saúde mental. Ainda que o trabalho original de Luce e Tuckey enfocasse variáveis concretas e diretamente

mensuráveis (tais como massa e volume), o embasamento teórico sólido é passível de ser aplicado a constructos complexos e intangíveis.

Em relação ao exemplo de massa e volume, a Teoria da Mensuração Conjunta propõe a derivação de uma terceira variável (densidade) que, ao se relacionar diretamente com as duas iniciais, apresentaria um comportamento quantitativo e atestaria o mesmo comportamento em relação a massa e volume (Bond e Fox, 2007). Assim sendo, comporia uma terceira variável não diretamente aferível, mas possível de ser estimada a partir da concatenação de duas variáveis conhecidas. A Tabela 2 descreve a relação entre massa, volume e densidade em um determinado experimento, ilustrando a inferência da variável densidade.

Tabela 2. Exemplificação da Teoria de Mensuração Conjunta a partir de massa e volume.

Mass	0.2 k	0.4 k	0.6k	0.8k	1.0k	1.2k
Volume						
0.5 l	.4	.8	1.2	1.6	2.0	2.4
1.0 l	.2	.4	.6	.8	1.0	1.2
1.5 l	.13	.27	.4	.53	.67	.8
2.0 l	.1	.2	.3	.4	.5	.6
2.5 l	.08	.16	.24	.32	.4	.48
3.0 l	.07	.13	.2	.27	.33	.4

Retirada de Bond e Fox (2007, p7)

As variáveis massa e volume estão explicitamente descritas a partir de mensurações diretas, enquanto densidade é representada pela estimativa calculada a partir das duas primeiras. Entretanto, como a tabela ilustra, densidade é definida como uma função linear direta de massa e volume. Segundo Luce e Tuckey (1964), o padrão de relação observável entre as duas variáveis é o indicador crucial da estrutura quantitativa aditiva da terceira.

Os axiomas utilizados para testar os requisitos necessários à Teoria de Mensuração Conjunta são numerosos e complexos. Entretanto, Bond e Fox (2007) oferecem uma sistematização simplificada para auxiliar o entendimento das bases de tais axiomas. Em síntese, é esperado que o valor de uma determinada casela na tabela 2 seja: a) menor do que o valor da casela imediatamente acima; b) menor do que o valor imediatamente à direita, e; c) menor do que o valor da casela posicionada na diagonal à direita e acima. Se as três variáveis examinadas

mantiverem essas relações estáveis ao longo de suas magnitudes, os requisitos da Teoria de Mensuração Conjunta estão satisfeitos e estamos, pois, diante de três variáveis quantitativas e aditivas (Bond e Fox, 2007; Karabatsos, 2001; Tenant et al, 2004).

A Teoria de Mensuração Conjunta passou, inicialmente, a ser aplicada no contexto de educação e de determinação de habilidade (inteligência) frente a testes de aferição de aprendizado (Andrich, 1988). Basicamente, testes compostos de uma série de itens eram apresentados a alunos de diferentes graus de habilidade, e o resultado desse encontro era estudado como uma terceira variável a ser avaliada. Se os requisitos da Teoria de Mensuração Conjunta se mantêm, uma hipotética tabela de probabilidade de acerto para cada item deve seguir o mesmo padrão apresentado na Tabela 2 (Bond e Fox, 2007).

A Tabela 3 abaixo descreve o padrão esperado para um teste de habilidade em uma série de alunos.

Tabela 3. Probabilidades de acerto em um encontro de habilidade vs dificuldade de itens.

Items	<i>p</i>	<i>q</i>	<i>r</i>	<i>s</i>	<i>t</i>	<i>u</i>	<i>v</i>	<i>w</i>	<i>x</i>	<i>y</i>	<i>z</i>	Ability
Persons												
P	.500	.866	.924	.949	.963	.973	.980	.986	.991	.995	.999	1/99
Q	.134	.500	.653	.741	.801	.847	.884	.915	.942	.968	.995	10/90
R	.076	.347	.500	.603	.682	.746	.801	.851	.896	.942	.991	20/80
S	.051	.259	.397	.500	.585	.659	.726	.789	.851	.915	.986	30/70
T	.037	.199	.318	.415	.500	.578	.653	.726	.801	.884	.980	40/60
U	.027	.153	.254	.341	.422	.500	.578	.659	.746	.847	.973	50/50
V	.020	.116	.199	.274	.347	.422	.500	.585	.682	.801	.963	60/40
W	.014	.085	.149	.211	.274	.341	.415	.500	.603	.741	.949	70/30
X	.009	.058	.104	.149	.199	.254	.318	.397	.500	.653	.924	80/20
Y	.005	.032	.058	.085	.116	.153	.199	.259	.347	.500	.866	90/10
Z	.001	.005	.009	.014	.020	.027	.037	.051	.076	.134	.500	99/1
Facility	1/99	10/90	20/80	30/70	40/60	50/50	60/40	70/30	80/20	90/10	99/1	

Retirada de Bond e Fox (2007, p11)

Como ilustrado na Tabela 3, o aumento de um atributo (a probabilidade de acerto) está relacionado ao aumento das outras duas variáveis (facilidade do item e habilidade do aluno). A relação entre as três variáveis é indicadora de suas naturezas quantitativas. Uma vez que a relação entre as três variáveis está estabelecida, pode-se estimar a magnitude de qualquer delas a partir do conhecimento das outras duas. Ou seja, a dificuldade de um determinado item pode ser estimada se a habilidade dos alunos e a probabilidade de acerto para cada dificuldade forem conhecidas. Ainda mais importante para as investigações na área da saúde mental, se a dificuldade dos itens e a probabilidade de acerto para cada grau de habilidade forem conhecidas, então a real habilidade pode ser estimada (Andrich, 1988). A habilidade (denominada de theta na literatura acerca

do tema e denotada com a letra grega ϕ) diz respeito a um determinado atributo de um sujeito. Assim como no campo da educação o interesse freqüentemente é determinar o grau de conhecimento de um aluno, a habilidade pode ser entendida no campo da saúde mental como a quantidade de um sintoma ou prejuízo. Em realidade, no contexto de desenvolvimento de escalas e instrumentos de medida em saúde mental, ϕ passa a ser referida como a intensidade de depressão (Chachamovich et al, 2007), qualidade de vida (Tenant et al, 2004), ansiedade (Pallant e Tenant, 2007), etc.

O Modelo de Rasch representa um modo de aplicação e testagem da Teoria de Mensuração Conjunta, uma vez que possibilita o ordenamento de habilidades e dificuldades. Deste modo, propicia que a probabilidade de acerto de cada item seja estimada. Adotemos como exemplo um teste com 12 itens que seja aplicado a um conjunto de 14 sujeitos. Com um grau diferente de dificuldade entre os itens, espera-se que os mais fáceis sejam acertados pela maioria dos sujeitos, enquanto somente os mais hábeis acertem os mais difíceis . A tabela 4 ilustra a situação hipotética mencionada, com os itens e sujeitos ordenados por dificuldade e habilidade, respectivamente.

Tabela 4. Ordenamento de dificuldade e habilidade

Selected Data Matrix With Persons Arranged According to Ability (From Top to Bottom) and Items Arranged by Facility (From Left to Right)													
Persons	Items											Ability	
	c	i	a	l	b	h	k	d	f	j	e		g
N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
J	✓	✓	✓	X	✓	✓	✓	✓	✓	X	X	✓	10
C	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	X	9
E	✓	✓	X	✓	✓	✓	✓	X	✓	✓	X	X	8
L	✓	✓	X	✓	✓	✓	✓	X	✓	✓	✓	X	8
I	✓	✓	✓	✓	✓	✓	X	✓	X	X	X	X	7
F	✓	✓	✓	✓	✓	✓	X	✓	X	X	✓	X	7
K	✓	✓	✓	✓	X	X	✓	X	✓	X	X	X	6
A	✓	✓	✓	✓	✓	✓	X	X	X	X	X	X	6
G	✓	✓	✓	✓	X	X	✓	X	✓	X	X	X	6
D	✓	✓	✓	✓	X	X	X	✓	X	X	X	X	5
B	✓	✓	✓	X	X	X	X	✓	X	X	X	X	4
H	✓	X	✓	✓	X	X	X	X	X	X	X	X	3
M	X	X	X	X	X	X	X	X	X	X	X	X	0
Facility	13	12	11	11	8	8	7	6	6	3	3	2	

Retirada de Bond e Fox (2007, p20)

No exemplo ilustrado na Tabela 4, percebe-se que os itens de dificuldade menor estão situados à esquerda, e os mais difíceis à direita. Assim sendo, espera-se que os itens *c*, *i*, *a* sejam respondidos corretamente pelos sujeitos mais hábeis e pelos sujeitos com graus menores de habilidade. Entretanto, à medida que os itens vão exibindo graus maiores de dificuldade (tais como os itens *f*, *j*, *e*, *g*), somente os sujeitos com maior habilidade os respondem corretamente. Paralelamente, os sujeitos também estão posicionados na Tabela 4 de acordo com graus crescentes de habilidade. Assim sendo, os sujeitos mais hábeis (tais como os *N*, *J* e *C*) acertam mais do que os de menor habilidade (como os *M*, *H* e *B*).

Entretanto, a Tabela 4 também evidencia alguns padrões de resposta inesperados, destacados pelas áreas sombreadas. Tomemos como exemplo o sujeito *K*. Se examinarmos o padrão de respostas desse sujeito, perceberemos que os itens *b* e *h* foram incorretamente respondidos. A probabilidade de acerto desse mesmo sujeito (com o mesmo nível de habilidades) ao encontrar um item de dificuldade maior é teoricamente menor do que as esperadas para os itens *b* e *h*. Ainda assim, o sujeito *K* acertou os itens *k* e *f*.

O Escalograma de Gutmann, desenvolvido em 1950, viria a servir de base para a posterior construção do Modelo de Rasch. O Escalograma descreve o padrão esperado de probabilidades de acerto do encontro de um item *X* com um sujeito de habilidade *Y* (Van der Linden, 1998). Esquemáticamente, Gutmann definiu um padrão de resposta ideal determinístico para esperar a probabilidade de acerto entre itens e sujeitos (Andrich, 1988). Seguindo as bases teóricas do Escalograma, o modelo assume a existência de um traço latente unidimensional, a ser definido por um padrão ideal de respostas completamente determinado pela relação entre habilidade e dificuldade (Van der Linden e Hambleton, 1997). Como o modelo de Gutmann é eminentemente determinístico, prevê que as respostas sigam o padrão de acertos esperado pelo encontro habilidade vs dificuldade.

O padrão ideal determinístico, entretanto, não é verificado em situações reais de testes. Alguns padrões não previstos pelo Escalograma de Gutmann (tais como os ilustrados pelas áreas sombreadas da Tabela 6) são encontrados. Ou seja, há fatores aleatórios que podem interferir no resultado do encontro de um sujeito com um item, que não somente a dificuldade e a habilidade (Andrich, 1988). Entre outras estatísticas envolvidas no Modelo de Rasch, este modelo

conduz uma verificação da adequação dos dados observados (coletados a partir de um teste) com o padrão esperado pelo Escalograma. Assim sendo, o Modelo de Rasch é considerado um modelo probabilístico, e não determinístico (Van der Linden e Hambleton, 1997).

Através de transformações logarítmicas, o Modelo de Mensuração de Rasch é fundamentalmente um modelo teórico probabilístico em relação ao qual os dados coletados serão confrontados (Bond e Fox, 2007; Fischer e Molenaar, 1995). Originalmente, o Modelo de Rasch foi desenvolvido para ser adequado a dados binários (Rasch, 1960) e, posteriormente, expandido para uso em variáveis politômicas (Andrich, 1978). Por incluir somente um parâmetro, o Modelo de Rasch é um dos modelos mais aplicados da família dos modelos de Teoria de Resposta ao Item de 1 Parâmetro (Embretson e Reise, 2000). Como, por definição, o Modelo de Rasch entende que o resultado do encontro de um item com um sujeito depende somente da habilidade do sujeito e da *dificuldade* do item (e nada mais), este modelo considera somente o parâmetro *dificuldade*. Outros parâmetros componentes de outros modelos de TRI (tais como discriminação e guessing) não são parte do Modelo de Rasch (Edelen e Reeve, 2007).

Para o modelo original de Rasch, a variável dependente é a probabilidade de acerto/erro de um determinado item de uma certa pessoa. As variáveis independentes são o traço de habilidade do sujeito (ϕ) e o nível de dificuldade do item (parâmetro δ). As variáveis independentes são aditivamente combinadas, e a dificuldade do item é subtraída da habilidade do sujeito. A variável dependente (probabilidade de acerto) é modelada pelo logaritmo da chance de o sujeito

responder ao item corretamente, e é expressa pela relação entre a probabilidade de acerto (P_{is}) sobre a probabilidade de erro ($1 - P_{is}$). A notação matemática de tal relação segue:

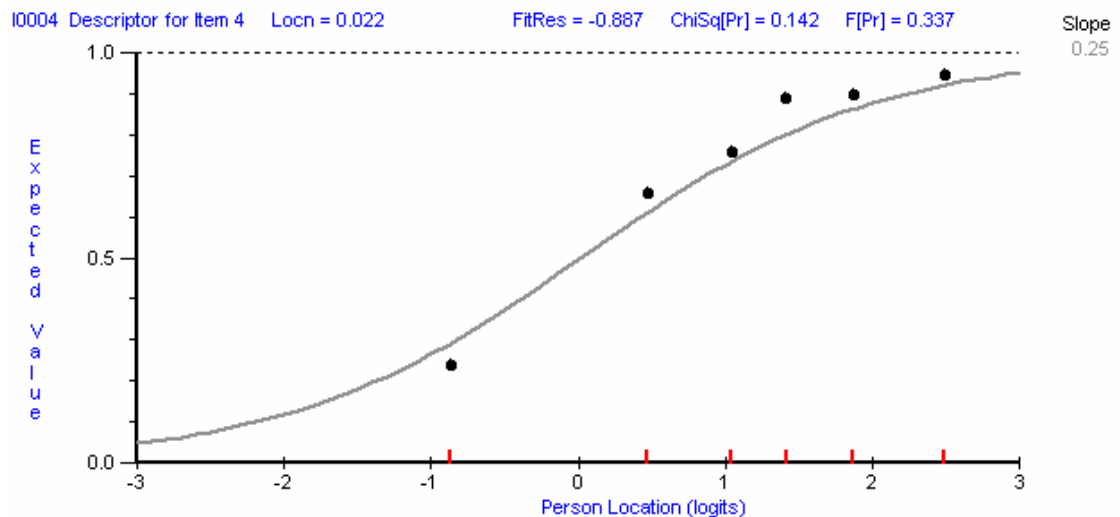
$$\ln [P_{is} / (1 - P_{is})] = \phi_s - \delta_i$$

Se a magnitude do traço (ϕ) é igual a magnitude da dificuldade (δ), então a probabilidade de acerto e de erro serão idênticas ($P = 0,50/0,50$). A relação entre ϕ e δ proporciona que a habilidade possa ser inferida a partir de um item com dificuldade conhecida (Embretson e Reise, 2000).

Uma característica de extrema importância para o Modelo de Rasch é a estimação independente da habilidade e da dificuldade. A separabilidade destes parâmetros constitui uma propriedade crucial para o modelo, qual seja a *especificidade objetiva* (Andrich, 1988; Embretson e Reise, 2000; Fischer e Molenaar, 1995). Rasch descreveu este princípio em 1977, e por ele expressa que as comparações obtidas por sujeitos são independentes dos itens utilizados. Por outro lado, as comparações entre itens também não dependem dos sujeitos que os respondem (Rasch, 1977; Embretson e Reise, 2000). Seguindo o conceito de *especificidade objetiva*, o Modelo de Rasch é considerado como um modelo não dependente de amostras (*sample-free*). Em outros termos, a estimativa da probabilidade de acerto de um item e da dificuldade do mesmo item não depende da amostra de sujeitos que responde o teste e, deste modo, o teste não necessita ser calibrado para cada amostra (Andrich, 1988).

O *princípio da invariância dos itens* responde pelo fenômeno da independência da amostra no Modelo de Rasch. Segundo aquele, o comportamento de um item é resultado somente da quantidade de habilidade de um sujeito que o responda, independentemente de qualquer outra característica que esse sujeito possa apresentar (Embretson e Reise, 2000). Características tais como sexo, idade, escolaridade, presença ou ausência de doença ou estado civil não devem ter reflexos no comportamento do item. Por outro lado, a variância do item frente a sujeitos com a mesma magnitude de ϕ indica problemas de performance deste item (também chamado de *item bias* ou *differential item functioning*) (Tenant e Pallant, 2007). Assim sendo, o Modelo de Rasch propicia que as estimativas de habilidade e de dificuldade de itens sejam realizadas de modo independente, e que a curva característica do item informe a probabilidade de acerto do item para cada magnitude de habilidade. A figura 6 ilustra uma curva característica do item, extraída das análises dos dados do Artigo 6 da presente Tese. Observa-se que o software utilizado para as análises de Rasch foi o *Rasch Unidimensional Measurement Model 2020* (RUMM2020), que fornece uma interface gráfica de fácil entendimento (Andrich et al, 2003).

Figura 6. Exemplo de Curva Característica do Item gerada pelo software RUMM 2020



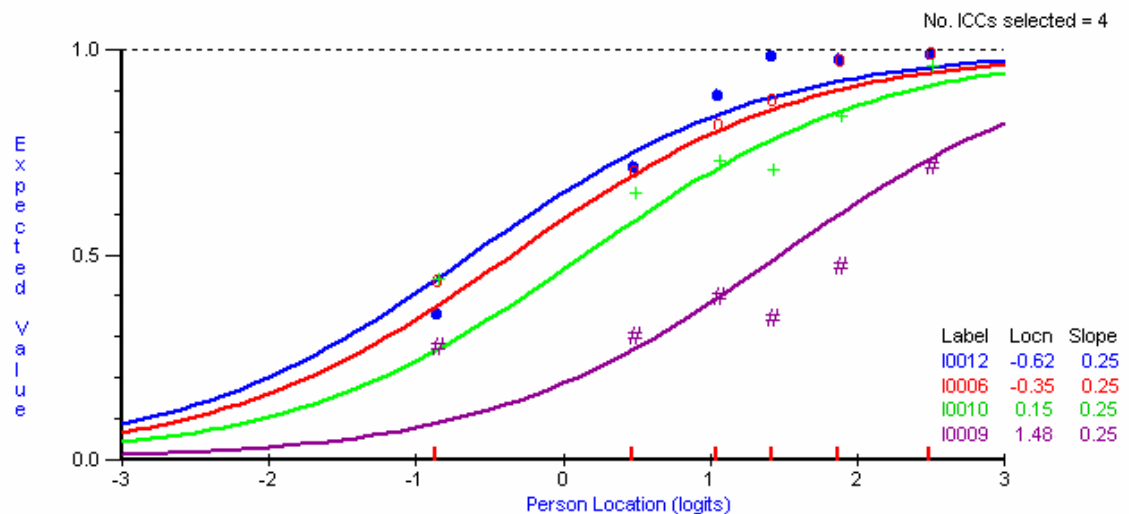
A figura 6 descreve informações importantes acerca da performance individual do item. Inicialmente, a curva desenhada com linha cheia representa a função esperada das probabilidades de resposta a este item, de acordo com a habilidade dos sujeitos (representada no eixo X pela localização das pessoas ao longo do eixo). Os seis pontos do gráfico indicam a probabilidade observada no banco de dados analisado, distribuídos de acordo com intervalos de classe. O parâmetro dificuldade é explicitado pela localização do item ($Locn = 0,022$). Complementarmente, o teste da adequação do item (*Item fit*) está indicado pelo resultado do teste Qui-quadrado (entre a curva observada e a esperada) e pela probabilidade do teste, assim como pela descrição dos resíduos obtidos na comparação curva observada – curva esperada. Como o desejado é que as duas

curvas não apresentem diferenças significativas, o valor de p deve ser superior ao valor crítico de 0.05 (Kline, 2005).

A Curva Característica do Item explicita também uma das fundamentais propriedades do Modelo de Rasch, qual seja a possibilidade de posicionar a dificuldade do item e a habilidade do sujeito em uma mesma escala métrica (Bond e Fox, 2007). A partir de transformações logarítmicas das estimativas de habilidade e dificuldade (em vários ciclos iterativos), os valores são representados em uma escala de logit (*log odd unit*). A escala de logit é uma escala intervalar, na qual os intervalos entre dois pontos têm um significado estável e regular (Embretson e Reise, 2000). Por exemplo, um sujeito com θX tem 50% de probabilidade de acertar um item de dificuldade Y . Sabemos, pois, que se a habilidade desse sujeito aumentar em 1 logit, a probabilidade de acerto no mesmo item passa a ser de 73%. Por outro lado, se o mesmo sujeito, com o mesmo θX , responder a um item com dificuldade 1 logit menor do que a do item Y , a probabilidade de acerto é também 73%. Se o item for 2 logits mais fácil, a probabilidade aumenta para 88%, e chega a 95% para itens com dificuldade 3 logit menor do que o item original Y (Smith, 2000; Wong et al, 2007).

A partir das Curvas Características de Itens, o investigador pode ter uma representação de vários itens simultaneamente. A Figura 7, retirada da mesma série de análises estatísticas da Figura 6, ilustra a representação gráfica de diferentes itens pelo programa RUMM2020.

Figura 7. Representação de simultâneas Curvas Características de Itens.



A Figura 7 tem como objetivo exemplificar o comportamento das diferentes probabilidades de acerto de um mesmo sujeito (e, portanto, com um mesmo nível de theta) frente a quatro itens. A localização da dificuldade dos quatro itens (descrita numericamente pelos *Locn* à direita da figura) indica que o item azul é o que possui menor grau de dificuldade, e o lilás tem o maior nível de dificuldade. Assim sendo, um sujeito necessita possuir habilidade de -0.62 logit para ter uma probabilidade de 50% de acertar o item azul, enquanto que são necessários 1,48 logits para que ele tenha a mesma probabilidade de acertar o item lilás.

O Modelo de Rasch apresenta duas suposições indispensáveis para que os resultados guardem credibilidade. Outros modelos de TRI compartilham tais suposições também. Unidimensionalidade é um requisito para a construção de variáveis segundo o Modelo de Rasch. A segunda condição requerida pelo

Modelo é a ausência de dependência local entre os itens (Pallant e Tenant, 2007; Andrich, 1988; Embretson e Reise, 2000; Fischer e Molenaar, 1995).

Ainda que os fenômenos psicológicos sejam eminentemente complexos e multidimensionais, as variáveis a serem medidas por um determinado teste ou instrumento devem apresentar unidimensionalidade (Andrich, 1988). Um fenômeno amplo pode demandar que seja aferido por uma série de instrumentos unidimensionais; entretanto, a variável a ser analisada deve guardar unidimensionalidade para que o Modelo de Rasch possa ser adequadamente aplicado (Tenant e Pallant, 2006). O princípio da unidimensionalidade indica que a variável refere-se a somente um atributo. Assim sendo, a interpretabilidade dos achados fica garantida, uma vez que o significado destes é conhecido (Bond e Fox, 2007). A inclusão de diversos atributos em um único instrumento impossibilita a aplicação do Modelo de Rasch, assim como cria grande confusão no significado das estimativas derivadas.

O princípio da *independência local* demanda que a probabilidade de acerto ou erro em um determinado item não deve ser dependente do sucesso ou falha em outro item. Em outras palavras, os itens devem ser independentes entre si e as probabilidades de acerto ou erro não devem ter relação entre itens. O instrumento SF36 (Ware et al, 1992; Ciconelli, 1999) exemplifica a ausência de independência local. Com uma escala de resposta tipo Likert de 3 pontos, os itens do bloco 3 do referido instrumento questionam, em seu enunciado: “Devido a sua saúde, você teria dificuldade para fazer essas atividades? Neste caso, quanto?” O subitem “d” menciona: “Subir **vários** lances de escada”, enquanto o subitem “e” pergunta: “Subir **um lance** de escada” (grifos originais do instrumento). Percebe-se, entre os

dois itens, uma relação de quantidade de atributo, de modo que o sujeito que responder afirmativamente ao item “e” tem uma probabilidade aumentada de responder afirmativamente também ao item “d”. Deste modo, os dois itens rompem a suposição de independência local e apenas um deve ser mantido para que o Modelo de Rasch seja adotado na análise (Andrich, 1988; Pallant e Tenant, 2007; Van der Linden e Hambleton, 1997).

Uma vez respeitados os princípios da unidimensionalidade e independência local, o Modelo de Rasch pode ser adequadamente executado. As principais contribuições deste modelo para as análises psicométricas de instrumentos de medida são a aferição da estrutura intervalar da medida, a consistência interna da medida, o exame do comportamento das categorias de resposta (*thresholds*), o comportamento invariante dos itens (*differential item functioning*) e a calibragem (ou *targetting*) dos itens de um determinado instrumento. Tais características são especialmente relevantes no desenvolvimento e adaptação de instrumentos de psicometria e serão descritos a seguir.

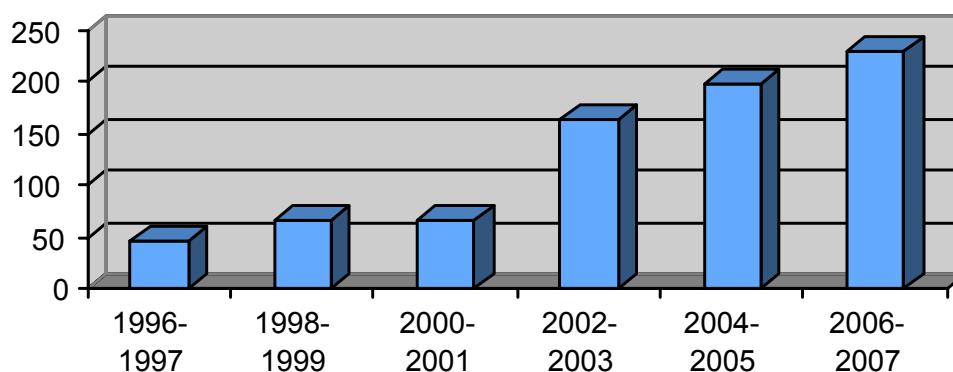
2.3 Aplicações do Modelo de Rasch em Psicometria

O Modelo de Rasch teve seu nascimento e desenvolvimento inicial predominantemente na área da educação. A terminologia adotada em relação aos seus parâmetros (habilidade e dificuldade) indica a sua origem. Entretanto, os diversos campos da saúde têm utilizado de modo crescente o Modelo de Rasch em suas investigações recentes, especialmente impulsionados pela necessidade de construção de medidas mais refinadas e adequadas a testes estatísticos complexos (Embretson e Reise, 2000).

Conrad e Smith (2004) destacaram a crescente utilização do Modelo de Rasch. De fato, o *International Conference on Objective Measurement (ICOM): Focus on Health Care*, realizado em Chicago em Outubro de 2001, teve como ponto de destaque a contribuição do Modelo de Rasch para o desenvolvimento dos estudos em psicometria. Entre os pontos enfatizados, a possibilidade de equacionar escores de diferentes instrumentos e a aplicação do Modelo de Rasch em Testes Computadorizados Adaptáveis (*Computerized Adaptive Tests* ou *CAT*) tiveram destaque (Conrad e Smith, 2004).

Uma busca sistematizada na base de dados PubMed evidencia o crescimento sustentado da utilização do Modelo de Rasch nas áreas da saúde. Utilizando-se o termo *Rasch* limitado aos campos de título ou *abstract*, o número de artigos indexados apresenta incremento constante na última década. O Gráfico 1 descreve tal crescimento.

Gráfico 1. Crescimento do número de artigos indexados sobre Modelo de Rasch na base PubMed



Qualquer aplicação do Modelo de Rasch demanda que suas duas assunções sejam previamente testadas. Primeiramente, a variável em análise deve apresentar unidimensionalidade. Ainda que vários métodos tenham sido historicamente aplicados para a determinação de fatores unidimensionais (tais como análise fatorial exploratória e análise fatorial confirmatória), o Modelo de Rasch tem sido considerado recentemente um dos métodos mais potentes para assegurar unidimensionalidade (Tenant e Pallant, 2006; Doward et al, 2004). Dentre as diversas estratégias estatísticas possíveis, uma das mais utilizadas se baseia na análise de componentes principais (ACP) dos resíduos. Resumidamente, os resíduos localizados no primeiro fator (determinado pela ACP) constituem dois subgrupos de itens (o primeiro composto pelos itens com carga mais positiva e o segundo pelos itens com carga mais negativa). Em seguida, as estimativas de habilidade de cada sujeito (calculadas a partir dos dois subgrupos de itens) são comparadas através de teste *t*-pareado. A ausência de diferença significativa entre as duas estimativas indica unidimensionalidade. Em tese, a comparação dos resíduos dos itens mais distantes representaria a maior possibilidade de se encontrar discrepâncias, e a semelhança entre as estimativas neste teste provavelmente atesta a semelhança entre quaisquer outros subgrupos de itens (Smith, 2002; Pallant e Tenant, 2007).

Kelly et al (2007) investigaram as potenciais discrepâncias entre análise fatorial confirmatória (AFC) e análise de Rasch em um instrumento para aferir crenças relacionadas ao estado de saúde (*Multidimensional Health Locus of Control*). Utilizando dados de três estudos prévios, com um total de 1206 sujeitos, os autores demonstraram que os resultados iniciais (*model fit*) pobres advindos de

análise pura de AFC apresentaram significativa melhora após a incorporação dos resultados gerados pelo Modelo de Rasch. Robitail et al (2007) propõem a combinação das metodologias psicométricas clássica e moderna (Modelo de Rasch) como uma possível alternativa de potencializar a credibilidade dos resultados. Os autores indicam que a utilização dos escores gerados por Modelo de Rasch são mais adequados para a exploração de parâmetros tais como invariância de itens e estabilidade do instrumento entre culturas.

O princípio da independência local dos itens também demanda ser testado e assegurado. O software RUMM2020 incorpora uma ferramenta direta para seu teste, baseada na determinação de correlação entre os itens. As correlações superiores ou iguais a 0.3 são consideradas inadequadas e indicam dependência entre dois itens (Pallant e Tenant, 2007; Pallant et al, 2006). Em casos de dependência local, recomenda-se que um item seja selecionado para exclusão, de acordo com o julgamento teórico do pesquisador e com o *item fit* de ambos (Pallant e Tenant, 2007; Embretson e Reise, 2000; Bond e fox, 2007).

De modo geral, a superioridade dos escores gerados pelo Modelo de Rasch sobre os escores brutos tem sido consistentemente reportada na literatura médica. Fitzpatrick et al (2003) exploraram a comparação entre o cálculo de escores brutos e derivados do Modelo de Rasch em 1221 pacientes realizando seguimento pós-cirúrgico em intervenções de quadril. Através da aplicação do instrumento *Oxford Hip Score* (composto de 12 itens com escala de respostas de 5 pontos), os autores demonstraram que a utilização de escores baseados no Modelo de Rasch foram mais capazes de demonstrar alterações na severidade dos sintomas (sensibilidade). Esta diferença é ainda mais evidente em escores distantes das

medidas centrais, evidenciando que as estimativas de pacientes com níveis mais brandos ou muito mais intensos de sintomas eram mais sistematicamente afetadas.

Nijsten et al (2006) realizaram um estudo comparativo entre o desempenho da Teoria de Psicometria Clássica e do Modelo de Rasch no refinamento de um instrumento de aferição de qualidade de vida em pacientes com psoríase (*Impact of Psoriasis Questionnaire*). Os autores conduziram análises independentes para desenvolver dois formatos de instrumentos abreviados (sendo um através de teoria clássica e outro por Modelo de Rasch). Por fim, o desempenho psicométrico de ambos foi comparado. Ainda que os dois instrumentos derivados tenham apresentado semelhança em relação à maioria dos itens retidos, a versão desenvolvida por Modelo de Rasch demonstrou superioridade em relação a unidimensionalidade, ordenação correta das categorias de resposta e consistência interna mais apropriada.

A propriedade de fornecer instrumentos abreviados com performance superior, assim como de manter essa superioridade em todo o espectro da intensidade da doença, são atribuídos primariamente à capacidade do Modelo de Rasch de lidar com variáveis intervalares (Norquist et al, 2003; Prieto et al, 2003; Nijsten et al, 2006; Bond e Fox, 2007; Tenant, 2004). Por outro lado, algumas limitações da teoria clássica de testes também são imputadas ao fato de esta não envolver necessariamente medidas lineares e traços latentes unidimensionais (Kelly et al, 2007).

De fato, o Modelo de Rasch constitui uma alternativa para a obtenção de variáveis com estrutura intervalar, e esta propriedade é considerada uma de suas

principais contribuições ao campo da psicometria (Pallant e Tenant, 2007; Andrich, 1988; Embretson e Reise, 2000; Fischer e Molenaar, 1995; Wilson, 2005). A inspeção individual dos itens (*item fit*) proporciona que cada questão de um determinado instrumento seja avaliada em relação à sua adequação ao modelo intervalar (Bond e Fox, 2007; Van der Linden e Hambleton, 1997). Assim sendo, os itens de pior desempenho podem ser individualmente excluídos e/ou modificados. Os itens restantes são, então, novamente analisados com o intuito de checar a adequação destes ao Modelo de Rasch (quanto à sua unidimensionalidade, independência local e propriedade intervalar) (Tenant, 2004; Pallant e Tenant, 2007; Pallant et al, 2006). Diversos ciclos de exclusão de itens (ou alteração de sua estrutura, como agrupamento de categorias de resposta) e de reanálise do desempenho dos demais itens levam ao formato final do instrumento. Este, então, será composto de itens com estrutura intervalar ao longo de todo o espectro de habilidade. Como supracitado, o Modelo de Rasch representa uma ferramenta que intrinsecamente incorpora os preceitos da Teoria de Mensuração Conjunta de Luce e Tuckey (1964), e, assim, gera variáveis genuinamente quantitativas, intervalares e aditivas (Pallant e Tenant, 2007; Karabatsos, 2001). A derivação de uma medida intervalar propiciará, pois, que uma série de testes estatísticos refinados possam ser aplicados sem o risco de resultados enviesados por informações de pouca qualidade (Bond e Fox, 2007).

Diversos modelos estatísticos estão disponíveis para que a propriedade intervalar da medida seja testada (Fischer e Molenaar, 1995). Uma das estratégias mais freqüentemente utilizada é a aplicação de teste de Qui-quadrado de aderência entre a estimativa esperada e a observada ao longo de diversos pontos

do traço latente (Andrich, 1988). De fato, o software RUMM2020 adota tal estratégia, e uma medida intervalar deve apresentar resultados não-significativos quando os valores esperado e observado são comparados (Pallant e Tenant, 2007; McKenna et al, 2007).

No Modelo de Rasch, a consistência interna é avaliada através do *Person Separation Index (PSI)*. Semelhantemente ao Coeficiente α de Cronbach, o PSI indica a replicabilidade dos resultados das estimativas. Em outros termos, o PSI indica a fração da resposta observada que é reproduzível em outras observações (Bond e Fox, 2007). Ademais, o PSI representa a capacidade de um conjunto de itens de “separar” ou diferenciar dois grupos distintos de sujeitos (Elhan et al, 2005; Fischer, 1992; Andrich 1982). A estatística de PSI varia de 0 a 1, sendo 0,7 aceitável para grupos e 0,9 para indivíduos (Lawton et al, 2004). Wright e Stone (1999) apontam que o fato de o PSI ser calculado a partir de medidas lineares (em contraste ao cálculo do Coeficiente α de Cronbach) confere a este uma credibilidade maior, assim como interpretabilidade. Os autores destacam que a linearidade do PSI permite inferir que um aumento de 0,05 pontos (como, por exemplo, de 0,65 para 0,7) representa metade do aumento de 0,1 ponto (por exemplo, de 0,8 para 0,9). Esta interpretação, entretanto, não é permitida para o Coeficiente α de Cronbach em função de sua estrutura não linear.

Outro implemento do Modelo de Rasch diz respeito à possibilidade de examinar o desempenho das categorias de resposta em casos de instrumentos com itens politômicos. O ponto de transição entre duas categorias de resposta em uma escala do tipo Likert é chamado *threshold*. Cada *threshold* representa o ponto

no qual há uma probabilidade de 50% de que o sujeito com um grau de habilidade X responda a categoria A ou B. Sujeitos com habilidades abaixo de X tenderão a assinalar a categoria A, enquanto sujeitos com habilidades superiores a X terão probabilidade maior que 50% de assinalar a categoria B (Linacre, 2001).

Assim sendo, espera-se que uma escala de resposta politômica possua um ordenamento lógico de categorias de resposta. A inversão da ordem esperada constitui o chamado *threshold disorder* (Pallant e Tenant, 2007; Osborn et al, 2000). O Modelo de Rasch oferece técnicas para averiguação do desempenho original das categorias de resposta, assim como permite que eventuais problemas sejam corrigidos (Hopman-Hok et al, 2000; Zhu et al, 1997). Uma das estratégias amplamente citada na literatura é o agrupamento das categorias disfuncionais, com o objetivo de compor uma categoria única e com ordenamento adequado de *thresholds* (Andersen, 1977; Andrich, 1978; Pallant et al, 2006). Como o Modelo de Rasch aborda variáveis latentes unidimensionais, os softwares são capazes de obter uma estimativa da habilidade de cada sujeito e comparar o comportamento das respostas em relação à quantidade de habilidade latente. Assim sendo, faz-se possível detectar se, em um determinado item, sujeitos com maior habilidade tendem a assinalar a categoria de resposta 2, enquanto outros com habilidade menor assinalam a categoria 3. Este exemplo hipotético comporia um *threshold disorder* e demandaria medidas de correção da escala ou exclusão deste item particular. As curvas de probabilidade de resposta são um meio de representação da interação entre o nível de habilidade e a probabilidade de respostas em cada categoria. A Figura 8 ilustra uma distribuição adequada de categorias, enquanto a

Figura 9 evidencia a presença de *threshold disorder*. Ambas foram retiradas das análises estatísticas do artigo 4 da presente Tese.

Figura 8. Distribuição adequada de categorias de resposta

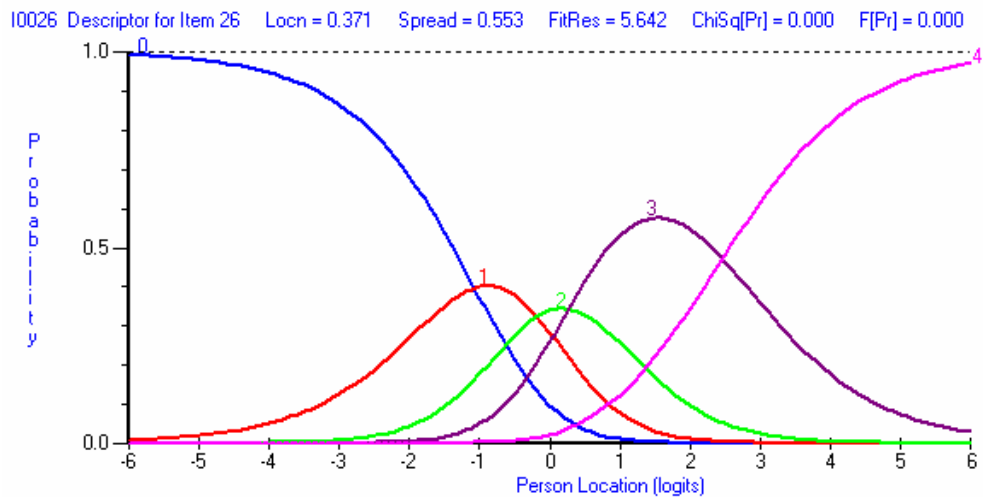
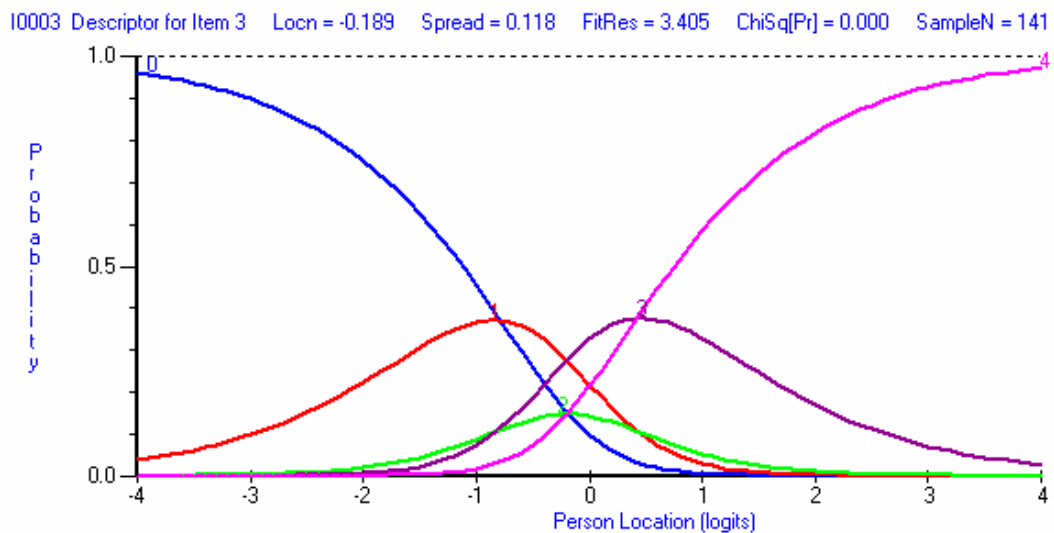


Figura 9. Distribuição inadequada de categorias de resposta (*threshold disorder*)



A figura 9 explicita a inadequação da categoria de resposta 2, uma vez que esta não assume probabilidade superior às outras categorias em nenhum ponto ao longo do traço de habilidade. Deste modo, o item demanda que alguma alteração seja procedida para que a qualidade da informação fornecida pelo item seja aprimorada (Andersen, 1977; Andrich, 1978).

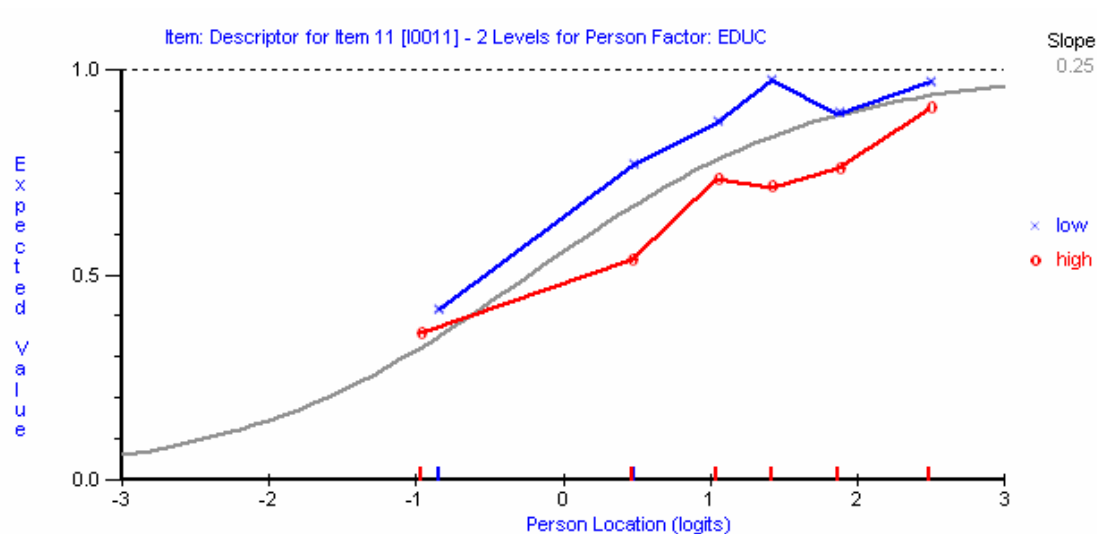
A possibilidade da determinação do princípio da invariância dos itens constitui outra importante contribuição do Modelo de Rasch. Kelderman (1989) destaca que a separação das estimativas de habilidade do sujeito e dificuldade do item confere ao Modelo de Rasch superioridade na investigação da invariância, comparado à teoria de psicometria clássica. Segundo tal princípio, o comportamento do item deve variar somente de acordo com a quantidade de habilidade que o sujeito apresenta (Wright e Panchapakesan, 1969). Nenhum outro fator externo deve, em tese, influenciar a probabilidade de um sujeito responder acertadamente a um item (Wright, 1989). A ausência da invariância determina o chamado *Differential Item Functioning* (DIF), definido por Embretson e Reise (2000) como a situação em que um item não tem uma relação estável com um mesmo nível de traço latente entre dois ou mais grupos de examinandos (tradução livre, p 251). Em outros termos, sujeitos com o mesmo nível de habilidade devem ter a mesma probabilidade de responder acertadamente a um item. Se, a despeito de um mesmo nível de traço latente, a probabilidade for diferente para dois sujeitos em função de uma determinada característica (tais como sexo, religião, cultura ou estado civil), este item é considerado como variante e, pois, enviesado (Pallant e Tenant, 2007, McKenna et al, 2007; Cauffman e MacIntosh, 2006). Em um exemplo hipotético de uma escala de

depressão, dois sujeitos com a mesma intensidade de sintomas depressivos devem ter a mesma chance de responder “sim” a um item que investigue nível de afeto deprimido. A não observância da invariância do item acarreta a impossibilidade de comparar os escores de tal escala entre os dois grupos (homem/mulher, cultura A/cultura B, etc.) (Hungj, 2005; Lai et al, 2005). Diversas características têm sido examinadas com a aplicação do Modelo de Rasch na investigação de DIF. Além de aspectos demográficos (tais como sexo e idade), o papel da transculturalidade constitui tópico de interesse no desenvolvimento de novos instrumentos, ou de instrumentos consagrados a serem aplicados em pesquisas multicêntricas (McKenna et al, 2007; Nijsten et al, 2006; Kong et al, 2007; Scott et al, 2007; Ramirez et al, 2007; Teresi, 2006).

A ausência de funcionamento invariante do item (DIF) divide-se em uniforme e não uniforme. *Differential Item Functioning* do tipo uniforme indica que a probabilidade distinta entre dois grupos de sujeitos se mantém homogênea ao longo de todas as intensidades do traço latente. O viés do item tem a mesma direção em todos os níveis de habilidade (Pallant e Tenant, 2007; Pallant et al, 2006; Lai et al, 2005; Hungj, 2005). Entretanto, o significado da mesma resposta entre os grupos não é uniforme. Vários métodos estatísticos podem ser aplicados na detecção de DIF (Hungj, 2005; Wilson, 2005). Entre eles, um dos mais utilizados é a Análise de Variância (ANOVA) em diversos pontos de habilidade ao longo do traço latente (McKenna et al, 2007). O software RUMM2020 disponibiliza a testagem de DIF por tal método (Andrich, Luo e Sheridan, 2003). Além da disponibilização dos resultados numéricos, o software também oferece uma representação gráfica do efeito do DIF no item. A Figura 10, retirada das análises

estatísticas do artigo 4, demonstra a presença de DIF uniforme em relação ao nível de escolaridade.

Figura 10. Ilustração gráfica de DIF uniforme



A presença de DIF do tipo não-uniforme, por outro lado, aponta o comportamento instável da probabilidade de resposta do item ao longo do traço latente (Pallant e Tenant, 2007). Assim sendo, a probabilidade de resposta apresenta-se maior para um grupo A em um determinado ponto do nível de habilidade, mas maior para um grupo B em outro ponto (Embretson e Reise, 2000; Van der Linden e Hambleton, 1997; Crane et al, 2006; Morales et al, 2006). Gráficamente, a presença de DIF não-uniforme resulta em linhas que se cruzam nos pontos em que a probabilidade inverte-se, acarretando graves dificuldades de interpretação das estimativas de habilidade (Edelen et al, 2006).

A uniformidade de DIF possibilita que medidas sejam adotadas para corrigir a variância do item. Uma das alternativas mais frequentemente utilizadas na

literatura é a divisão do item (*splitting*) de acordo com a característica que determina DIF uniforme. Assim sendo, o item contribuirá de forma distinta para a estimativa de habilidade em cada grupo, refletindo os diferentes níveis de dificuldade deste em contextos distintos (McKenna et al, 2007; Pallant e Tenant, 2007; Pallant et al, 2006). Entretanto, a heterogeneidade do comportamento do item com DIF não-uniforme determina sérias complicações quando estratégias de controle deste efeito são aplicadas. Conseqüentemente, a sugestão de que tais itens sejam deletados do instrumento predomina entre as estratégias de adequação (McKenna et al, 2007; Edelen et al, 2006).

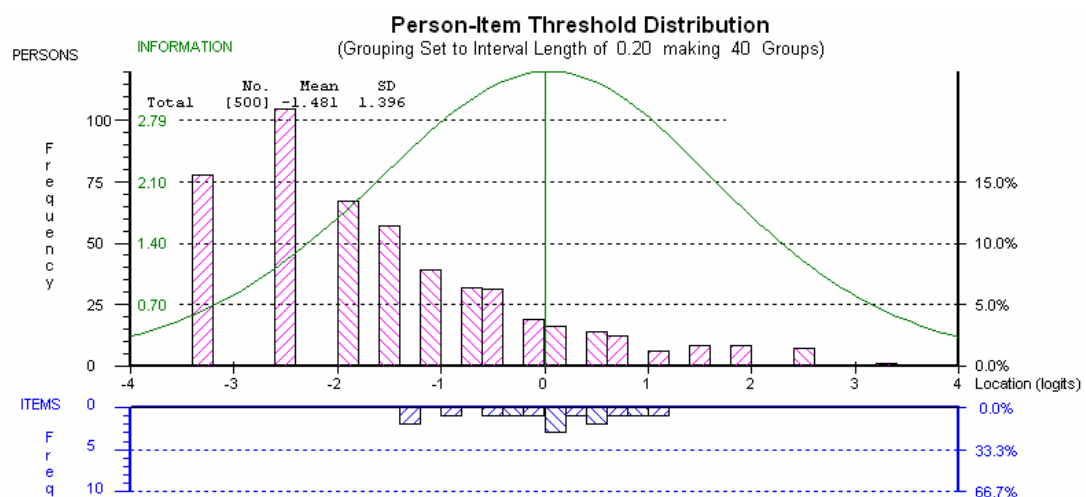
Estudos que envolvam mais de uma cultura devem necessariamente examinar a presença de DIF nos itens dos instrumentos utilizados (Tenant et al, 2004; Skevington et al, 2004). Anteriormente negligenciado na literatura, o funcionamento diferencial do item tem recebido crescente atenção devido ao seu marcado impacto na mensuração de um determinado atributo em diferentes populações (Teresi, 2006).

A propriedade da especificidade objetiva, baseada na separação das estimativas de δ e ϕ de modo independente, proporciona ao Modelo de Rasch determinar a calibragem (ou *targetting*) dos itens do instrumento (Wilson, 2005). O posicionamento de habilidade e dificuldade em uma mesma unidade métrica possibilita indicar em que ponto do traço latente cada item fornece mais informação (Pallant e Tenant, 2007). O mapa resultante desta análise é também conhecido como Mapa de Wright, em homenagem ao seu idealizador Benjamin D. Wright. Entre suas importantes funções, o mapa itens-pessoas possibilita ao

pesquisador detectar em que área de habilidade os itens são capazes de medir (ou seja, em que nível do traço latente os itens fornecem informação) (Pallant et al, 2006). Outra destacada função do mapa é a possibilidade de agregar itens de diferentes instrumentos, uma vez que todos os itens são posicionados em uma mesma “régua” (Bond e Fox, 2007). Esta função a Teoria de Resposta ao Item denomina de equalização de instrumentos.

A Figura 11 apresenta o mapa itens-pessoas gerado pelo software RUMM2020. A parte superior da escala ilustra o posicionamento dos sujeitos, enquanto a porção inferior apresenta a localização dos itens.

Figura 11. Mapa de itens-pessoas gerado pelo software RUMM2020



A figura 11 descreve uma predominância de informação (descrita pela curva em verde) próxima ao logit 0 da escala. A interpretação de tal achado

depende fundamentalmente do objetivo específico do instrumento (Teresi, 2006). Em casos de instrumentos desenhados para serem utilizados na estimativa de níveis de depressão (desde intensidades leves à graves), por exemplo, seria desejável desenvolver um instrumento capaz de fornecer informação em pontos mais amplos da escala. Para tal fim, o exemplo ilustrado na Figura 11 provavelmente não possuiria desempenho adequado, uma vez que não é suficientemente informativo em relação a sujeitos com depressão grave. Entretanto, em casos de testes do tipo aprovado-reprovado (em que o objetivo é determinar se o sujeito possui mais habilidade do que um determinado ponto de corte, mas não estimar o quanto mais hábil o sujeito é), a convergência de informação no ponto de corte é de interesse do pesquisador. Neste segundo exemplo, o teste ilustrado na Figura 11 seria, pois, adequado.

A possibilidade de equalização de itens a partir de diferentes escalas se constitui em um destacado ponto forte do Modelo de Rasch. A partir do Mapa de Wright, itens com o mesmo grau de dificuldade (e provenientes de diferentes instrumentos) podem ser ancorados em um mesmo ponto da escala métrica. Então, os itens restantes passam a ser posicionados de acordo com o δ específico de cada um, e estimativas de ϕ de cada sujeito podem ser realizadas a partir de combinações de itens dos diferentes instrumentos (Pallant et al, 2006; Tenant, 2004; Bond e Fox, 2007).

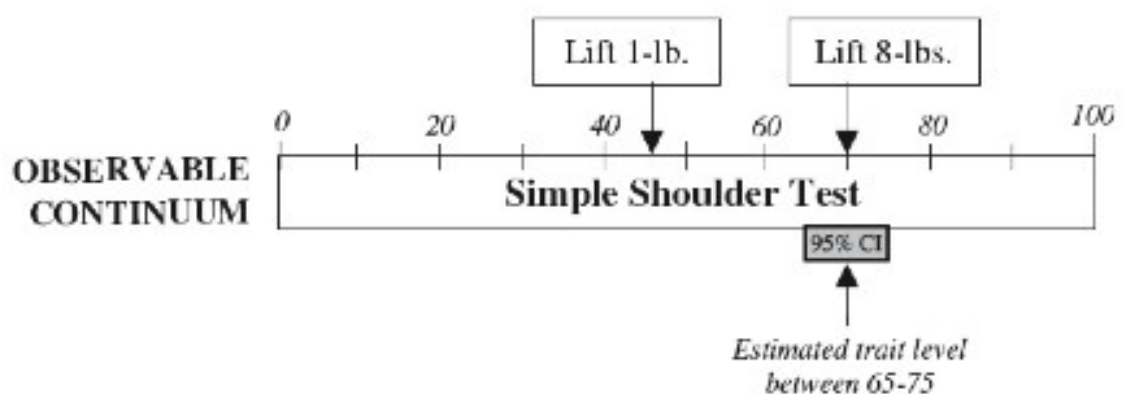
Por fim, os avanços proporcionados pela inclusão do Modelo de Rasch (particularmente equalização de instrumentos e especificidade objetiva) determinam que os chamados Testes Adaptados por Computador (*Computerized*

Adaptive Tests - CAT) possam ser aplicados (Bjorner et al, 2007; Feyers, 2007; Haley et al, 2006; Revicki e Sloan, 2007). Originalmente implementados na área de educação, a utilização de bancos de itens e o direcionamento de itens de acordo com o padrão de respostas do sujeito representam a possibilidade de administrar instrumentos significativamente mais curtos e eficientes (Bjorner et al, 2007; Ware et al, 2000).

Brevemente, os CAT são testes baseados em bancos de itens previamente validados e calibrados de acordo com o Modelo de Rasch (ou de outras modalidades de Teoria de Resposta ao Item). O caráter adaptativo faz com que o computador ofereça ao respondente itens de acordo com sua habilidade (aferida pelo padrão de respostas aos itens prévios). Se as respostas iniciais indicam que o sujeito tem alta probabilidade de responder afirmativamente ao um item de menor dificuldade, então o algoritmo computadorizado seleciona itens com maior grau de dificuldade. Cook et al (2005) apresentam um exemplo clínico de limitação funcional em um elegante artigo de revisão do funcionamento de CAT e banco de itens. Se o primeiro item apresentado fornece informação a respeito do grau 44 de um traço latente que varia de 0 a 100, e o sujeito responde que não há limitações funcionais de ombro para tarefas com esse grau de dificuldade, então o algoritmo computadorizado apresentará itens com δ superiores e, conseqüentemente, não aplicará itens com graus menores de dificuldade. Os últimos não forneceria informação adicional ao teste.

A Figura 12 ilustra o funcionamento de um CAT baseado em banco de itens previamente calibrado.

Figura 12. Funcionamento de um teste tipo CAT



Retirado de Cook et al, 2005

A utilização de CATs determina que algumas vantagens possam ser incorporadas às rotinas de testagens na área da saúde. A primeira delas é o incremento da eficiência de medida, definida como a relação entre a adequação da medida (validade, confiabilidade) e o fardo da aplicação de instrumentos. Os testes computadorizados permitem que a adaptação dos itens a cada sujeito diminua o número de itens necessários, mantendo a precisão da medida. Outra vantagem interessante é a definição *a priori* da precisão da medida. O erro padrão pode ser previamente definido, e a aplicação de itens cessa quando a estimativa obtiver a precisão desejada. Dependendo do objetivo do instrumento, o erro padrão pode ser ajustado e, pois, determinar um ainda maior encurtamento do

número de itens (Cook et al, 2005; Hart et al, 2006; Haley et al, 2006; Hahn et al, 2006).

Os testes adaptáveis por computador têm sido crescentemente desenvolvidos em diversas áreas. Por se tratar de um teste complexo, necessitam de uma estrutura complexa para que possam ser adequadamente implementados (Cook et al, 2005). Medidas de *Patient Reported Outcomes* estão entre as que já contam com CATs (Fries et al, 2005), assim como depressão (Fliege et al, 2005) e ansiedade (Walter et al, 2007).

3. Referências Bibliográficas da Introdução

- 1 Andersen E (1977). Sufficient statistics and latent trait models. *Psychometrika*, 42(1):69-81.
- 2 Andersen E (1997). Sufficient statistics and latent trait models. *Psychometrika*, 42(1):69-81.
- 3 Andersen E (2001). Rasch Symposium in honor of Georg Rasch. *Rasch Measurement Transactions*, 2001, 15:1 p.794
- 4 Andrich D (1982). An Index of Person Separation in Latent Trait Theory, the Traditional KR-20 Index, and the Guttman Scale Response Pattern.
□ *Education Research and Perspectives*, 9:1: 95-104.
- 5 Andrich D (1988). Rasch Models for measurement. Londres: SAGE Publications.
- 6 Andrich D, Lyne A, Sheridan B and Luo G (2003). RUMM 2020. Perth: RUMM Laboratory.
- 7 Andrich D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4):561-573.
- 8 Andrich D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care*, Jan; 42(1 Suppl):17-16
- 9 Baker F. (2001). Basics of Item Response Theory. Maryland: ERIC Clearinghouse on Assessment and Evaluation.
- 10 Baldwin JM (1902). Dictionary of Philosophy and Psychology (vol 2). Londres: McMilliam.

- 11 Bjorner JB, Chang CH, Thissen D et al (2007). Developing tailored instruments: item banking and computerized adaptive assessment. *Quality of Life Research*, 16(Suppl 1): 95-108.
- 12 Blanton H e Jaccard J (2006). Arbitrary metrics in psychology. *American Psychology*, 61(1):27-41.
- 13 Bond TG e Fox CM (2007). Applying the Rasch Model. Fundamental Measurement in the Human Sciences. New Jersey: Lawrence Erlbaum Associates Publishers.
- 14 Bowling A (1991). Health care research: measuring health status. *Nursing Practice* ; 4(4):2-8
- 15 Cauffman E e MacIntosh R (2006). A Rasch Differential Item Functioning Analysis of the Massachusetts Youth Screening Instrument: Identifying Race and Gender Differential Item Functioning Among Juvenile Offenders. *Educational and Psychological Measurement* 2006; 66; 502
- 16 Cella D, Yount S, Rothrock N et al (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, 45(5 Suppl 1): S3-S11.
- 17 Chang CH (2007). Patient-reported outcomes measurement and management with innovative methodologies and technologies. *Quality of Life Research*, 16 (Suppl 1):157-66
- 18 Charmann C, Chambers C, Williams H (2003). Measuring atopic dermatitis severity in randomized controlled clinical trials: what exactly are we measuring? *Journal of Investigative Dermatology*, 120:932 -941.

- 19 Chachamovich E, Fleck MP, Trentini CM et al (2007). The Brazilian WHOQOL-OLD Module version: A Rasch analysis of a new measure. *Revista de Saúde Pública, no prelo.*
- 20 Cichetti D, Bronen R, Spencer S et al (2006). Rating scales, scales of measurement, issues of reliability: resolving some critical issues for clinicians and researchers. *Journal of Nervous and Mental Disease.*, Aug;194(8):557-64.
- 21 Ciconelli RM, Ferraz FB, Santos W, Meinão I, Quaresma MR (1999.) Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Revista Brasileira de Reumatologia.* 39 (3), 143-149.
- 22 Claesson L e Svensson E (2001). Measures of order consistency between paired ordinal data: application to the Functional Independence Measure and Sunnaas index of ADL. *Journal of Rehabilitation Medicine,* Mar;33(3):137-44.
- 23 Conrad KJ e Smith EV (2004). International Conference on Objective Measurement *Applications of Rasch Analysis in Health Care. Medical Care • Volume 42, 1(1 suppl):I1-I6.*
- 24 Conrad KJ, Wright BD, McKnight P, McFall M, Fontana A e Rosenbeck M (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: revealing limitations of reverse-scored items. *Journal of Applied Measurement,* 5(1):15-30.

- 25 Cook KF, O'Malley KJ e Roddey TS (2005). Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Service Research*, 40 (5 pt 2):1694-711.
- 26 Crane PK, Gibbons L, Jolley L e Belle GV (2006). Differential Item Functioning Analysis with Ordinal Logistic Regression Techniques. *Medical Care*, 44:S115-S123.
- 27 Dombret B, Matthijs S e Sabzevar MM (2003). Interexaminer reproducibility of ordinal and interval-scaled plaque indices. *Journal of Clinical Periodontology*, Jul;30(7):630-5.
- 28 Doward LC, Meads D, Thorsen H. (2004). Requirements for Quality of Life instruments in Clinical Research. *Value in Health*, 7(S1):S13-S16.
- 29 Duncan OD (1984). Notes on social measurement: Historical and Critical. Nova Iorque: Russel SAGE Foundation.
- 30 Edelen MO e Reeve B (2007). Applying Item Response Theory (IRT) modeling to questionnaire development, evaluation and refinement. *Quality of Life Research*, 16:5-18.
- 31 Edelen MO, Thissen D, Teresi JA et al (2006). Identification of Differential Item Functioning Using Item Response Theory and the Likelihood-Based Model Comparison Approach. *Medical Care*, 44:S134-S142.
- 32 Elhan AH, Kutlay S, Küçükdeveci AA et al (2005). Psychometric properties of the Mini-Mental State Examination in patients with acquired brain injury in Turkey. *Journal of Rehabilitation Medicine*, Sep; 37(5):306-11.
- 33 Embretson, SE (2006). The continued search for nonarbitrary metrics in psychology. *American Psychology*, 61(1):50-55.

- 34 Embretson SE e Reise SP (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates Publishers.
- 35 Feyers P (2007). Applying Item Response Theory and computer adaptive testing: the challenges for health outcomes assessment. *Quality of Life Research*, 16 (Suppl 1): 143-55.
- 36 Field A. (2005). *Discovering Statistics using SPSS*. 2nd, SAGE, London
- 37 Fisher WP (1992). Reliability statistics. *Rasch Measurement Transactions*; 6: 238.
- 38 Fischer GH e Molenaar IW (1995). *Rasch Models: foundations, recent developments, and applications*. New York: Springer-Verlag.
- 39 Fitzpatrick R, Norquist JM, Jenkinson C et al (2004). A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. *Quality of Life Research*, 13(2):331-8.
- 40 Fitzpatrick R, Norquist JM Dawson J et al (2003). Rasch scoring of outcomes of total hip replacement. *J Clin Epidemiol.*, Jan;56(1):68-74
- 41 Fliege H, Becker J, Walter OB et al (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of Life Research*, 14(10):277-91.
- 42 Fries JF, Bruce B e Cella D (2005). The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clinical and Experimental Rheumatology*, 23(5 Suppl 39):S53-7.
- 43 Haley SM, Ni P, Hambleton RK et al (2006). Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *Journal of Clinical Epidemiology*, 59(11):1174-82.

- 44 Hahn EA, Cella D, Bode RK et al (2006). Item banks and their potential applications to health status assessment in diverse populations. *Medical Care*, 44 (11 Suppl 3):S189-97.
- 45 Hair Jf, Anderson RE, Tatham R et al (2006). *Análise Multivariada de Dados*. Porto Alegre: ArtMed.
- 46 Hart DL, Cook KF, Mioduski JE et al (2006). Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *Journal of Clinical Epidemiology*, 59(3):290-8.
- 47 Hays RD, Morales RS e Reise SP (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, Sep;38(9 Suppl):II28-42.
- 48 Hopman-Rock M, Van Buuren S, De Kleijn-De et al (2000). Polytomous Rasch analysis as a tool for revision of the severity of disability code of the ICIDH. *Disability and Rehabilitation*. May 20;22(8):363-71.
- 49 Hungi N (2005). Employing Rasch Model to detect biased items. In: S. Alagumalai et al. (eds.), *Applied Rasch Measurement: A Book of Exemplars*, 139–157. Springer. Holanda
- 50 Karbatsos G (2001). The Rasch Model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, 2:398-423.
- 51 Kazdin AE (2006). Arbitrary metrics: implications for identifying evidence-based treatments. *American Psychology*, .61(1):42-9
- 52 Kelderman, H. (1989). Item Bias Detection Using Loglinear IRT. *Psychometrika*, 54(4), 681-697.

- 53 Kelly PA, Kallen MA e Suarez-Almaroz ME (2007). A combined-method psychometric analysis recommended modification of the multidimensional health locus of control scales. *Journal of Clinical Epidemiology*, 60:440-447
- 54 Kline RB (2005). Principles and Practice of Structural Equation Modeling. New York: The Guilford Press.
- 55 Kong KO, Ho HJ, Howe HS et al (2007). Cross-cultural adaptation of the Systemic Lupus Erythematosus Quality of Life Questionnaire into Chinese. *Arthritis and Rheumatism*, Aug, 15; 57(6):980-5.
- 56 Krantz DH (1964). Conjoint Measurement: the Luce-Tuckey axiomatization and some extensions. *Journal of Mathematical Psychology*, 1:82-100.
- 57 Kuzon MW Jr, Urbanchek MG e McCabe S (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, Sep;37(3):265-72.
- 58 Lai JS, Teresi J e Gershon R (2005). Differential item functioning (dif) for small sample sizes. *Evaluation & the Health Professions*, 28(3): 283-294
- 59 Lawton G, Bhakta BB, Chamberlain MA e Tenant A. (2004). The Behçet's Disease Activity Index. *Rheumatology*, 43:73–78
- 60 Linacre JM (2001). Category, Step and Threshold: Definitions & Disorder. *Rasch Measurement Transactions*, 15:1 p.794
- 61 Luce RD (1979). Suppe's contributions to the theory of measurement. In: Bogdan RJ (ed). *Patrick Suppes*. Dordrecht: Reidel.
- 62 Luce RD and Tuckey JW (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1:1-27.
- 63 Marshall M, Lockwood A, Bradley C et al (2000). Unpublished rating scales: A

- major source of bias in randomised controlled trials of treatment for schizophrenia. *British Journal of Psychiatry*, 176:249-52.
- 64 McHorney CA, Haley SM e Ware JE Jr (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, Apr;50(4):451-61.
- 65 McKenna SP, Dowards LC, Meads DM et al (2007). Quality of Life in infants and children with atopic dermatitis: addressing issues of differential item functioning across countries in multinational clinical trials. *Health and Quality of Life Outcomes*, Jul(27):5-45.
- 66 Merbitz C, Morris J e Grip JC (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine Rehabilitation*, Apr;70(4):308-12.
- 67 Michell J (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100:398-407.
- 68 Michell j (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88:355-83.s
- 69 Michell J (1999). An introduction to the logic of psychological measurement. New Jersey: Lawrence Erlbaum Associates Publishers.
- 70 Michell J (2005). Measurement in Psychology. A Critical history of a methodological concept. Cambridge, Reino Unido: Cambridge Press.
- 71 Morales LS, Flowers C, Gutierrez P et al (2006). Item and Scale Differential Functioning of the Mini-Mental State Exam Assessed Using the Differential Item and Test Functioning (DFIT) Framework. *Medical Care*, 44(11 Suppl 3):S143-S151.

- 72 Munzel U e Bandelow B (1998). The use of parametric vs nonparametric tests in the statistical evaluation of rating scales. *Pharmacopsychiatry*, 31 :222–224
- 73 Nijsten T, Unaeze J e Stern RS (2006). Refinement and reduction of the Impact of Psoriasis Questionnaire: Classical Test Theory vs. Rasch analysis. *British Journal of Dermatology*; 154, pp692–700
- 74 Norquist JM, Fitzpatrick R, Dawson R et al (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care*, Jan; 42 (1 Suppl):I25-36.
- 75 Olsen LW (2003). Essays on Georg Rasch and his contributions to statistics. Tese (Doutorado em Economia) – Instituto de Economia, Universidade de Compenhague. Disponível em <http://www.rasch.org/olsen.pdf> (acessado em 17/11/07).
- 76 Osborn PS, Behrens JT, Ryan Jm et al (2000) Applying the Rasch Model for Ordered Categories To Assess the Relationship between Response Choice Content and Category Threshold Disorder. Artigo apresentado no Annual Meeting of the American Educational Research Association (New Orleans, LA, April 2000). Disponível em http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/7c/c6.pdf (acessado em 15/11/2007).
- 77 Pallant J and Tenant A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46:1-18.
- 78 Pallant J, Miller R and Tenant A. (2006). Evaluation of the Edinburgh Post

Natal Depression Scale using Rasch analysis. *BMC Psychiatry*, 6:28-38.

- 79 Prieto L, Alonso J e Lamarca R (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes*, 1:27-40.
- 80 The PROMIS Group. Patient Reported Outcome Measurement Information System (2007). Disponível em <http://www.nihpromis.org>, acessado em 15/11/2007.
- 81 Ramirez M, Teresi JA, Holmes D et al (2007). Differential item functioning (DIF) and the Mini-Mental State Examination (MMSE). Overview, sample, and issues of translation. *Medical Care*, 44(11 Suppl 3):S95-S106.
- 82 Rasch G. (1960). Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press.
- 83 Rasch G (1977). On specific objectivity: an attempt at formalising the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 14:58-94.
- 84 Reeve BB, Hays RD, Bjorner JD et al (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45 (5 Suppl 1): S22-S31.
- 85 Revicki DA e Sloan J (2007). Practical and philosophical issues surrounding a national item bank: if we build it will they come? *Quality of Life Research*, 16 (Suppl 1): 167-74.
- 86 Robitail S, Ravens-Sieberer U, Simeoni MC et al (2007). Testing the structural

- and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Quality of Life Research*, Oct;16(8):1335-45.
- 87 Scott NW, Fayers PM, Aaronson NK et al (2007). The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. *Quality of Life Research*, 16(1):115-29.
- 88 Sirvent JB (2003). Sonidos, fracciones, medias, potencias y funciones exponenciales. *Suma*, Noviembre: 39-44.
- 89 Skevingotn S, Sartorius R, Amir M e The WHOQOL Group (2004). Developing methods for assessing quality of life in different cultural settings: The history of the WHOQOL instruments. *Social Psychiatry and Psychiatric Epidemiology*, 39: 1–8
- 90 Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 2, 199–218
- 91 Smith, E. V. (2002). Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231.
- 92 Song F, Jerosch-Herold C, Holland R et al (2006). Statistical methods for analysing Barthel scores in trials of poststroke interventions: a review and computer simulations. *Clinical Rehabilitation*; 20: 347-356
- 93 Stevens SS (1946). On the theory of the scales of measurement. *Science*, 103:667-80.
- 94 Stevens SS (1968). Measurement, statistics and the schemapiric view. *Science*, 161:849-56.
- 95 Streiner DL e Norman GF (2003). Health MEasurement Scales: a practical

- guide to their development and use. Oxford: Oxford University Press.
- 96 Stucki G, Daltroy L, Katz JN et al (1996). Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology*, 49(7):711-717.
- 97 Suen HK (1990). Principles of test theories. Hillsdale: Erlbaum.
- 98 Suppes P (1954). Some remarks on problems and methods in the philosophy of science. *Philosophy of Science*, 21:242-8.
- 99 Svensson E (2001). Guidelines to statistical evaluation of data from rating scales and questionnaires. *Journal of Rehabilitation Medicine*, 33, 47–48.
- 100 Svensson E (1998). Ordinal invariant measures for individual and group changes in ordered categorical data. *Statistica in Medicina*, Dec 30;17(24):2923-36.
- 101 Tastle WJ e Wierman MJ (2006). An information theoretic measure for the evaluation of ordinal scale data. *Behavioral Research Methods.*, Aug, 38(3):487-94
- 102 Tenant A e Pallant JF (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions*, 2(4): 1082-84.
- 103 Tenant A, Hillman M, Fear J, Pickering A, Chamberlain MA (1996). Are we making the most of the Stanford Health Assessment Questionnaire? *British Journal of Rheumatology*, 5(6):574-8.
- 104 Tenant A, Penta M, Tesio L et al (2006). Assessing and adjusting for cross-cultural validity of impairment and activity limitations scales through differential

- item functioning within the framework of the Rasch model: the PRO-ESOR project. *Medical Care*, Jan, 42(1 Suppl):137-48.
- 105 Tenant A and Pallant JF. (2006). Unidimensionality matters! (A tale of two Smiths?). *Rasch Measurement Transactions*, 20:1048-1051.
- 106 Tenant A, McKenna SP e Hagell P (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value Health*, Sep-Oct;7 Suppl 1:S22-6.
- 107 Teresi JA (2006). Different Approaches to Differential Item Functioning in Health Applications: *Advantages, Disadvantages and Some Neglected Topics*. *Medical Care*, 44(Suppl):S152-170.
- 108 Terrien J (1972). Fundamental Metrology: Present State and Recent Progress in some Important Fields of Fundamental Metrology, and Functions of Consultative Committees of the International Committee of Weights and Measures. *Metrologia*, 8:99-108.
- 109 Tiburcio S (2002). Musica y matematicas. *Elementos Ciencia e Cultura*, 8(44): 21-26.
- 110 Tutz G (2003). Generalized semiparametrically structured ordinal models. *Biometrics*, Jun; 59(2):263-73.
- 111 Van Der Linden W(1998) stochastic order in dichotomous item response models for fixed, adaptive, and multidimensional tests. *Psychometrika*; 63 (3): 211-226
- 112 Van Der Linden W e Hambleton RK (1997). Handbook of Modern Item Response Theory. New York: Springer-Veerlag.

- 113 Walter OB, Becker J, Bjorner JB et al (2007). Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Quality of Life Research*, 16(Suppl 1):143-55.
- 114 Ware JE and Sherbourne CD. (1992). The MOS 36-Item Short Form Health Survey (SF-36): 1. Conceptual framework and item selection. *Medical Care*; 30(6):473-483.
- 115 Ware, J. E., Jr., Bjorner, J. B., & Kosinski, M. (2000). Practical implications of item response theory and computerized adaptive testing: A brief summary of ongoing studies of widely used headache impact scales. *Medical Care*, 38, 1173–1182
- 116 Webster's Encyclopaedic Unabridged Dictionary (1996). New Jersey: Gramercy Books.
- 117 Wiklund I (2004). Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundamental and Clinical Pharmacology*, 18(3):351-63.
- 118 Wilson M (2005). Constructing Measures. An Item Response Modeling Approach. Londres: Lawrence Erlbaum Associates Publishers.
- 119 Wittkowski KM, Lee E, Nussbaum R et al (2004). Combining several ordinal measures in clinical studies. *Statistics in Medicine*, May 30;23(10):1579-92.
- 120 Wong E, Ungvari GS, Leung SK et al (2007). Rating catatonia in patients with chronic schizophrenia: Rasch analysis of the Bush-Francis Catatonia Rating Scale. *International Journal of Methods in Psychiatric Research*, 16(3):161-70.
- 121 Wright and Panchapakesan (1969). A Procedure for Sample-Free Item Analysis. *Educational and Psychological Measurement*.1969; 29: 23-48

- 122 Wright BD, Linacre JM, Smith RM, Heinemann AW, Granger CV. (1997) FIM measurement properties and Rasch model details. *Scandinavian Journal of Rehabilitation Medicine*. Dec;29(4):267-72.
- 123 Wright BD, Linacre JM. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine Rehabilitation*.; 70:857–860.
- 124 Wright, B (1997). S. S. Stevens Revisited. *Rasch Measurement Transactions*, 11:552-3.
- 125 Wright, B.D. (1989). Rasch model from counting right answers. Reprinted in Linacre, J.M. (Ed), *Rasch Measurement Transactions Part 1*. Chicago: MESA Press, 1995.
- 126 Wright BD e Stone M (1999). *Measurement Essentials*. Delaware: Wide Range Inc.
- 127 Zhu W, Updyke WF e Lewandowski C (1997). Post-Hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement*, 1(4):286-304.

4. Questões de Pesquisa, Objetivos e Hipóteses

4.1 Objetivo Geral

Explorar as aplicações do Modelo de Rasch no desenvolvimento e validação de instrumentos psicométricos em saúde mental

4.2 Objetivos Específicos

4.2.1) Artigo 1: O instrumento de qualidade de vida para idosos WHOQOL-OLD foi inicialmente desenvolvido e validado à luz da Teoria de Psicometria Clássica. Investigar se a aplicação de uma teoria moderna robusta (Modelo de Rasch) demanda alterações do formato do instrumento, ou o formato original mostra-se válido também frente a tal Modelo.

4.2.2) Artigo 2: Validar o instrumento de Atitudes Frente ao Envelhecimento para uma amostra de idosos brasileiros, aplicando uma estratégia de combinação de psicometria clássica e moderna (Modelo de Rasch).

4.2.3) Artigo 3: Analisar o comportamento de escalas de resposta tipo Likert de 5 pontos em amostras de idosos com alto nível de educação e em analfabetos.

4.2.4) Artigo 4: Analisar se o instrumento Geriatric Depression Scale, em sua versão em Português, apresenta desempenho psicométrico satisfatório, através de modelo de Rasch.

4.2.5) Artigo Anexo: Estudar se níveis de depressão clínica têm impacto na qualidade de vida de idosos e, se tal impacto também é observado em pacientes com níveis de depressão subclínica.

4.3 Hipóteses

4.3.1 Artigo 1

Hipótese Conceitual: O desempenho psicométrico da escala WHOQOL-OLD, a partir da metodologia de Análise de Rasch, mostrar-se-á adequado, assim como demonstrado pela Teoria de Psicometria Clássica (TPC).

Hipótese Operacional:

H_0 : Desempenho psicométrico de WHOQOL-OLD por TPC = Desempenho por Rasch

H_a : Desempenho psicométrico de WHOQOL-OLD por TPC \neq Desempenho por Rasch

4.3.2 Artigo 2

Hipótese Conceitual: O instrumento Atitudes Frente ao Envelhecimento (AAQ) apresenta propriedades psicométricas adequadas em sua versão em Português, assim como mantém a adequação observada em uma amostra internacional.

Hipótese Operacional:

H_0 : Desempenho do AAQ em Português = Desempenho do AAQ na versão internacional

H_a : QV Desempenho do AAQ em Português \neq Desempenho do AAQ na versão internacional

4.3.3 Artigo 3

Hipótese Conceitual: Sujeitos analfabetos não apresentam habilidade de diferenciar adequadamente entre cinco pontos de uma escala de resposta tipo Likert (EL), quando comparados com sujeitos de alta escolaridade.

Hipótese Operacional:

H_0 : Desempenho da EL em analfabetos = Desempenho da EL em universitários

H_a : Desempenho da EL em analfabetos \neq Desempenho da EL em universitários

4.3.4 Artigo 4

Hipótese Conceitual: A escala Geriatric Depression Scale apresenta propriedades psicométricas adequadas em sua versão em Português, a partir de sua validação através de Teoria de Psicometria Clássica.

Hipótese Operacional:

H_0 : Desempenho da GDS em Português (modelo observado) = Modelo de Rasch esperado

H_a : Desempenho da GDS em Português (modelo observado) \neq Modelo de Rasch esperado

4.3.5 Artigo Anexo

Hipótese Conceitual: Os sintomas depressivos (tanto em intensidade clínica como subclínica) acarretam significativo prejuízo na qualidade de vida (QV) e atitudes frente ao envelhecimento (AFE) de idosos.

Hipótese Operacional:

H₀: QV e AFE em idosos com sintomas depressivos = QV e AFE em idosos assintomáticos

H_a: QV e AFE em idosos com sintomas depressivos ≠ QV e AFE em idosos assintomáticos

5. Proteção dos Direitos Humanos

Este conjunto de estudos foi aprovado pelo Comitê de Ética em Pesquisa do Hospital de Clínicas de Porto Alegre sob o título “Desenvolvimento de uma medida de avaliação de Qualidade de Vida para idosos e sua relação com envelhecimento saudável” (protocolo de aprovação número 01-374).

Assim sendo, o projeto está em conformidade com as Diretrizes e Normas Internacionais e Nacionais de pesquisa, especialmente quanto à Resolução 196/96 e complementares do Conselho Nacional de Pesquisa.

Cabe ressaltar que as coletas de dados realizadas em centros internacionais respeitaram a legislação local, assim como foram devidamente apreciadas e aprovadas por seus respectivos Comitês de Ética em Pesquisa.

ARTIGO 1

**The Brazilian WHOQOL-OLD Module version: A Rasch analysis of
a new measure**

Artigo aceito para publicação na Revista de Saúde Pública

**The Brazilian WHOQOL-OLD Module version: A Rasch analysis of a new
measure**

**Versão em Português do Módulo WHOQOL-OLD: Uma análise de Rasch de
um novo instrumento**

Eduardo Chachamovich^{1,2}, Marcelo P. Fleck¹, Clarissa Trentini¹, Mick Power²

1 – Programa de Pós-Graduação em Psiquiatria, UFRGS

Rua Ramiro Barcelos, 2350.

Porto Alegre, RS

Brazil

2 – Clinical and Health Psychology, University of Edinburgh

Medical School

Teviot Place

EH8 9AG, Edinburgh

Trabalho realizado no Programa de Pós-Graduação em Psiquiatria, Universidade
Federal do Rio Grande do Sul.

Autor para correspondência: Eduardo Chachamovich

Rua Florêncio Ygartua, 391/308

CEP 90430-010

Porto Alegre – RS

Email: echacha.ez@terra.com.br

Subvenção: O autor Eduardo Chachamovich foi financiado para bolsa de Estágio de Doutorado no Exterior pela CAPES, processo número 3604-06/3

Running title: The WHOQOL-OLD Module : A Rasch analysis

RESUMO:

Introdução: O mundo passa por uma profunda mudança demográfica, uma vez que os idosos vivem mais e mais saudavelmente. Esta mudança na pirâmide etária demanda pesquisas específicas sobre o processo de envelhecimento. O Grupo de Qualidade de Vida da Organização Mundial de Saúde desenvolveu recentemente o Módulo WHOQOL-OLD com o objetivo de permitir comparações transculturais. A validação da versão em português do Módulo WHOQOL-OLD envolveu a abordagem psicométrica clássica e indicou propriedades psicométricas adequadas. A teoria de Rasch representa uma abordagem psicométrica moderna para o desenvolvimento e validação de instrumentos. Tem sido reconhecida como uma ferramenta potente para examinar a performance de instrumentos detalhadamente. Ademais, é útil em fornecer soluções a instrumentos com performance inadequada.

Objetivo: Analisar a Versão Brasileira do Módulo WHOQOL-OLD através de análise de Rasch e explorar alterações potenciais do instrumento objetivando aumentar a adequação psicométrica.

Resultados: A amostra é composta de 424 sujeitos com idade igual ou superior a 60 anos. Dois domínios dos seis apresentaram interação item-total insuficiente. Remodelar a escala de resposta e deletar itens com pior performance acarretou melhora da escala. Todos os domínios apresentaram unidimensionalidade e independência local.

Conclusões: a análise dos domínios e itens individualmente foi capaz de indicar que o domínio Intimidade teve boa performance, ao contrário dos resultados

gerados pela abordagem psicométrica clássica. Ademais, os achados confirmaram que o domínio Funcionamento dos Sentidos não parece ser adequado em seu formato atual. As alterações na escala de resposta e a exclusão de itens problemáticos determinaram melhora da performance da escala.

Descritores: qualidade de vida, idosos, envelhecimento, WHOQOL-OLD, Rasch

ABSTRACT

Introduction: The world is experiencing a profound demographic shift as older people are living longer and healthier. This shift in the age pyramid demands further specific research on the aging process. The World Health Organization Quality of Life Group has recently developed the WHOQOL-OLD module with the aim to permit cross-cultural comparisons. The validation of the Brazilian version of the WHOQOL-OLD module involved classic psychometric approach and indicated suitable psychometric properties. The Rasch theory is a modern psychometric approach to the development and validation of instruments. It has emerged as a powerful tool for examining the performance of instrument in depth. In addition, it is also helpful for providing potential solutions for misperforming instruments.

Objective: to analyze the Brazilian Version of the WHOQOL-OLD module using Rasch analysis and explore potential alterations of the instrument in order to increase its psychometric adequacy.

Results: The sample is composed of 424 subjects aged 60 or older. Two out of six domains showed inadequate item-trait interaction. Rescoring the response scale and deleting the most misperforming items led to scale improvement. Unidimensionality and local independence were present in all domains.

Conclusion: The examination of the domains and the items individually was capable of showing that the Intimacy domain does perform well, in contrast to the results of the classical approach. In addition, it confirmed that the Sensory Abilities domain does not seem to be suitable in its current format. Finally, alterations in the response scale and deletion of problematic items improved the scale performance.

Keywords: quality of life, elderly, ageing, WHOQOL-OLD, Rasch

INTRODUCTION

The world is experiencing a profound and irreversible demographic shift as older people are living longer and healthier than ever before (UN, 2003). The most dramatic increases in proportions of older people are evident in the oldest section of society (people over 80 years old) with an almost fivefold increase from 69 million in 2000 to 377 million in 2050 (UN, 2003). The World Health Organization has described this demographic shift as a major societal achievement, and a challenge (WHO, 2001). The increase in longevity is being experienced in the developed and the developing world alike, but where the developed world grew rich before it grew old, the developing world is growing old before it has grown rich (WHO, 2002).

This shift in the age pyramid due to greater numbers of elderly people demands further specific research on the aging process. One important field to be assessed is quality of life. Although several studies have reported findings on this issue, systematic reviews have pointed out that the instruments most frequently used in these investigations are not sufficiently comprehensive and/or are not validated for application in older adult populations (Haywood et al, 2005; Chachamovich et al, 2006).

The World Health Organization Quality of Life Group has recently developed the WHOQOL-OLD module (Power et al, 2005). Through a simultaneous transcultural methodology, this instrument is designed to be suitable for cross-cultural comparisons. In addition, it was developed to assess quality of life specifically for the elderly, thus ensuring that important areas concerning old age

are covered by the instrument. Its comprehensiveness is sustained by an initial intense qualitative phase (Fleck et al, 2003; Hawthorne, 2006). The WHOQOL-OLD module represents an additional tool, alongside the WHOQOL-100 or WHOQOL-BREF, and it is a useful alternative in the investigation of quality of life in older adults, including relevant aspects not covered by the instruments originally designed for non-elderly populations.

The validation of the Brazilian version of the WHOQOL-OLD module is reported in detail elsewhere (Fleck et al, 2006). Briefly, it involved classic psychometric approach to analyze internal consistency, discriminant validity, criterion validity, concurrent validity and test-retest reliability. The findings indicated suitable psychometric properties for this version.

The Rasch measurement theory is a modern psychometric approach to the development and validation of instruments. It has emerged as a powerful tool for examining the performance of instrument in depth, allowing both the instrument as whole and individual items to be assessed. In addition, the Rasch model is also helpful for providing potential solutions for misperforming instruments. It is suggested that combining both traditional and modern psychometric approaches is a fruitful strategy to enhance power of validation processes (Tammaru et al, 2006). Furthermore, the use of the Rasch measurement model for development and application of quality of life instruments has been increasingly stressed (Tenant et al, 2004; Swaine-Verdier et al, 2004; Power et al, 2005).

The present study aims to explore the Brazilian Version of the WHOQOL-OLD module using a modern psychometric approach (Rasch analysis). In addition,

potential alterations of the instrument in order to increase its psychometric adequacy will be tested whenever necessary.

METHOD

Subjects and procedures

The data collected for the original classic validation (Fleck et al, 2006) was also analyzed for the present paper. A minimal sample of 300 subjects stratified by gender (50% women and 50% men), age (60-69 years of age, 70-79 years of age and over 80) and perceived health status (50% considering themselves healthy and 50% unhealthy) should be selected at a university hospital, nursing homes, and in the community according to the WHOQOL-OLD project. Convenience sampling was used. The stratification process provided minimum subsamples that allowed for the assessment of the instrument under different conditions.

The inclusion criteria were age 60 or above and clinical ability to understand and respond to the instruments administered. The subjects were required to answer the question "In general, do you consider yourself healthy or unhealthy?", and were later stratified as healthy or unhealthy exclusively according to their subjective perception, regardless of their actual objective health status. This methodology is based on the theoretical background for the quality of life instruments developed by the WHO, once the quality of life construct is seen as multidimensional and basically subjective (The WHOQOL Group, 1998).

All respondents were informed about the objectives of the study and the confidentiality of the data obtained. The subjects received and signed the Informed Consent approved by the Research and Ethics Committee of the university hospital where the study was carried out.

Instruments

The subjects completed a socio-demographic data form, the WHOQOL-OLD module and the Geriatric Depression Scale 15-item version (Sheik and Yesavage, 1986). The WHOQOL-BREF instrument was also part of the assessment, and its psychometric performance is reported elsewhere (Chachamovich et al, 2006).

The socio-demographic data form included questions about gender, age, educational level, marital status, subjective perception of health status, and consumption of alcohol, tobacco and illegal substances. The data obtained from this questionnaire was utilized for demographic description, as well as for DIF analysis.

The WHOQOL-OLD is a 24-item self-report instrument. It is divided into six domains (Sensory Abilities, Autonomy, Past-Present-Future Activities, Social Participation, Death and Dying and Intimacy). Each domain provides an individual score. In addition, an overall score is calculated from the set of the 24 items. Answers are based on a 5-point Likert response scale (Power et al, 2005). It is validated for Brazilian Portuguese, and this version presents good classic psychometric performance (Fleck et al, 2006).

Rasch Analysis

Data was examined by way of the Rasch model using the RUMM 2020 software (Andrich et al., 2003). Linacre states that the ideal sample size varies according to the scale targeting. For a well-targeted scale (40-60% endorsement rates on dichotomous items), a sample size of 108 would have a 99% confidence of person estimation of ± 0.5 logits. For non well-targeted scales, though, a minimal sample size for satisfactory estimations would be 243 subjects (Linacre, 1994).

The Rasch model is understood as a template which puts into operation the axioms for additive conjoint measurement (Pallant et al, 2006). This theory presents a set of methods to determine whether a variable has an additive structure and, then, is amenable to be measured on an interval scale (Rasch, 1960). Originally developed to be applied in dichotomous scales, the Rasch model is also applicable in polytomous data (Andrich, 1978).

Basically, the Rasch model assumes that the probability of a given subject endorsing an item is a function of the relative distance between the item location and the person location on a linear common scale (Pallant and Tenant, 2007). In the case of a scale to measure depression, for example, the probability that a person is endorsing an item is a logistic function of the difference between the subject's ability (level of depression) and the level of depression expressed by the item. The following equation illustrates this statement

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \theta_n - b_i$$

where \ln is the normal log, P is the probability of a person n to endorse the item, θ is the person's level of ability and b is the level of ability expressed by the item. If the data fits the Rasch model, then both the person's ability and item difficulty will be placed in a common metric scale (log-units scale or logit), which permit a linear transformation of the raw scale; thus, when the data fits the model, and the assumptions of local independence are met, the scale is then suitable for valid parametric approaches (Pallant et al, 2006). Since Rasch analysis is strongly dependent on unidimensionality, each of the six WHOQOL-OLD domains was tested individually as separated scales (Pallant and Pallant, 2006).

Apart from unidimensionality, local independence is also considered a Rasch assumption. Items are required not to have dependence on each other, so that the probability of endorsing one item is not associated to any other in the scale. Local independence should be examined for each scale to be analyzed by the Rasch model.

If Rasch assumptions are satisfied, and the scale fits the expected model, then it is also guaranteed that the performance of the instrument is stable and is not dependent on the sample being assessed, or certain characteristics such as gender or age, which is called specific objectivity (Tenant et al, 2004).

Overall fit statistics were examined first. An item-trait interaction was analysed using a Chi-Square test, which indicates the invariance property if the p-

value is not significant (thus indicating similarity between expected and observed models). The standardized distributions of items and persons were examined by way of a diagram.

Furthermore, individual item statistics were analysed for residuals and Chi-Square statistics. Again, if a determined item fits the model, low residual (± 2.5) and non-significant Chi-Square statistics are expected. Bonferroni correction was applied to control multiple tests effect. Threshold disorders were also examined by threshold maps and category probability curves for each individual item.

An estimate of internal consistency was also obtained through the Person Separation Index (PSI), which is comparable to the Cronbach's Alpha coefficient. Items were examined for differential item functioning (DIF). The presence of DIF indicates that a subgroup (e.g., males or young adults) has a consistent way of responding to an item, despite having the same amount of the latent trait. Both uniform DIF (when the difference is constant through the whole range of the item curve) and non-uniform DIF (when the difference occurs only at a certain level of attribute) were checked.

Finally, modifications were tested when fit statistics indicated misfit. Item rescaling and deletion were carried out in order to achieve the best item structure possible.

RESULTS

Demographics. The sample is composed of 424 subjects. Table 1 describes sample characteristics. The Geriatric Depression Scale means and SD indicate that the sample is predominantly non-depressed. In addition, around two thirds of

the subjects perceived themselves as being healthy, despite the objective health condition. Subjective perception is known to be related to depression levels. Thus, the high rate of 'healthy' subjects may be considered an indirect effect of low depression levels in the sample.

TABLE 1

Rasch Analysis Results

The verification for missing values shows that only items 1 and 3 had extremely low missing value rates (between 0.2% and 0.4%). The distributions of responses across the five points did not present major problems. These findings corroborate the high responsiveness of the WHOQOL-OLD in a Brazilian sample. It is probable that the close assistance that research staff offered to subjects during data collection is somehow related to the unexpected low number of missing values. Table 2 presents the item contents, missing values, median and distributions.

TABLE 2

The item-trait interaction was analyzed for the six domains individually through chi-square statistics. This test aims at checking whether the observed model (i.e., the data collected) fits the expected model (based on a probabilistic adaptation of Guttman scale) (Andrich, 1988). Thus, as Kline states, it is primarily

a test of “badness-of-fit”, since statistical positive results (p-values above the critical one, after Bonferroni correction) indicate that the observed model is different from the expected (Kline, 2005). The Death and Dying domain presented an inadequate result (Domain $\chi^2 = 51.72$, $p=.00012$). The Sensory Abilities domain also showed high chi-square results (Domain $\chi^2 = 101.10$ and $p=.0000$).

Local dependence was examined for the six domains and for the set of 24 items together. A correlation of the residuals for all items was carried out. Coefficients equal or higher than 0.3 are considered indicators of local dependence. No dependence was found for any domain or for the overall scale.

Items 4,5,9 and 20 showed reversed threshold. Thresholds indicate the point where there is exactly the probability of .50 that a subject will respond to the item between a certain response category and the adjacent one. Threshold disorders, thus, suggest that the response scale is not efficient to discriminate between two ability levels, so that subjects with more ability could respond in the same category as another with lower ability. In other words, the response scale would not be working adequately to order subjects with distinct levels of ability. These items were examined and rescored according to the point of the disorder in the response scale. For the items 4,5 and 9, response categories two and three were merged into one. For the item 20, categories three and four were collapsed (values for the original instrument).

Figure 1 illustrates the category probability curves of the item 5 in its 5-point original form and after rescoring. It is possible to observe that the original form presents reversed thresholds (i.e., category number 2 is not endorsed at any

point). After rescaling, the categories are well distributed. It is also important to notice that the RUMM2020 software used in this analysis (Andrich et al, 2003) automatically renames the categories in order to assign the value 0 for the first category. In the instrument, however, the categories range from 1 to 5.

FIGURE 1

The distributions of persons and item thresholds are illustrated in figure 2. The persons' locations are placed on the top half of the graph. The average mean person location value was 0.719 (SD 0.744). This is slightly above the average of scale items (which would be zero logits). Threshold distribution is located on the bottom half of the graph. The peak of information of the scale (if taken as a 24-item set) is located between 0 and -1 logits. However, the thresholds adequately cover all the range of ability, which ensures that the scale is capable of providing information for all levels.

FIGURE 2

Differential item functioning was assessed by gender (male and female) and age (60 to 79 years old and 80 or older). Item bias indicates that the item performance is not homogeneous and, thus, presents distinct performance on

different subjects when controlling for the level of underlying construct measured by the test (Crane et al, 2007). As a result, scores obtained from an item with DIF are not comparable across populations. Items were analyzed for uniform and non-uniform DIF. Briefly, the first is related to a constant difference of functioning through the entire spectrum of the construct, while the second indicates that the DIF is present only in a determined part of the curve (Crane et al, 2006). Uniform DIF items can be either excluded from the scale or, alternatively, they can be used to create two different scales (and then the item would have distinct weights in each).

Item 3 (Sensory Abilities domain) showed uniform DIF for age. No DIF was found for the other items.

Scale Modifications

The first step in the scale modification was rescoreing response categories. Besides solving the threshold disorders, the item-trait interaction showed improvement for the Sensory Abilities domain (original $\chi^2 = 142.44$ and after rescoreing $\chi^2 = 93.32$). This improvement was not sufficient to adjust this domain to the expected model.

Item 3 showed differential functioning, as well as misfit of chi-square test and residuals. These three statistics suggest that item 3 is not performing according to the expected Rasch model. Thus, item 3 was deleted and the domain

was, then, re-examined. The item-trait interaction showed improvement (χ^2 changed from 93.32 to 59.28). However, the model after deleting is still misfitting.

The Death and Dying domain also showed item-trait interaction misfit in its original format ($\chi^2=60.03$). Rescoring item 20 determined improvement of the model ($\chi^2=51.72$). At this stage, values were still non-significant, indicating persistent misfit. Deletion of item 18 (which presented high chi-square results) resulted in an adjusted structure.

Table 3 describes the fit statistics for the refined WHOQOL-OLD version.

TABLE 3

Discussion

The present paper aims to describe a Rasch analysis of a new quality of life measure specifically designed for older adults. The WHOQOL-OLD module was developed under a simultaneous transcultural methodology, which is able to include different cultural contexts since the first steps of the instrument construction (Guillemin, 1995). This is understood as an important characteristic of the WHOQOL-OLD (Power et al, 2005).

In addition to the theoretical design, it is also crucial that the validation process of a new international measure is carried out adequately. This ensures that the original strengths of the instrument are kept in the new version in a different language. The validation of a scale or instrument is a longitudinal process and should ideally involve tests in distinct contexts.

The combination of different psychometric approaches for the validation or development of a new measure is supported in the literature. Particularly, it has been argued that the Rasch measurement model is able to add important input, since it puts into operation the axioms for additive conjoint measurement (Pallant et al, 2006). Using both traditional and the Rasch analyses seems to be a useful strategy and provide relevant insight regarding scale performance (Power et al, 2005; Tammaru et al, 2006).

The present findings are in line with the results previously reported through classical psychometric theory (Fleck et al, 2006). The Sensory Abilities domain showed inadequate performance in multiple linear regression analyses in the previous study. The Rasch analysis corroborated the domain misfitting. The Intimacy domain, however, showed problems in the classical psychometric approach (multiple linear regression), but not in Rasch analysis. This discrepancy indicates that the domain itself functions well as a set, and the items show satisfactory performance. It is suggested the previous findings are due to the limitations of the multiple linear regression, particularly the choice of a suitable dependent variable.

Rescoring and item deletion determined insufficient improvement for the Sensory Abilities domain. Interestingly, item rescoring and deletion significantly

improved the performance of the Death and Dying domain. After alterations, the model statistics fit the Rasch model.

These potential alterations should not promote crucial modifications in the scale format, since they can be made during the statistical analysis phase and not necessarily in the data collection stage. Replications of these findings in different samples are needed to confirm the results.

In summary, the Rasch analysis added important information about the psychometric performance of the Brazilian version of the WHOQOL-OLD module. The examination of the domains and the items individually was capable of showing that the Intimacy domain does perform well, in contrast to the results of the classical approach. In addition, it confirmed that the Sensory Abilities domain does not seem to be suitable in its current format. Finally, alterations in the response scale and deletion of problematic items improved the scale performance.

References

1. Andrich D, Lyne A, Sheridan B and Luo G(2003). RUMM 2020. Perth: RUMM Laboratory.
2. Andrich D. Rating formulation for ordered response categories. *Psychometrika*, 1978, 43:561-573.
3. Andrich D. Rasch models for measurement. Sage University Paper, London, 1988.
4. Chachamovich E, Trentini C and Fleck MP. Assessment of the psychometric performance of the WHOQOL-BREF instrument in a sample of Brazilian older adults. *Int Psychoger*, 2006, Jul, 27:1-12.
5. Crane PK, Gibbons LE, Jolley L et al. Differential item functioning related to education and age in Italian version of the Mini-Mental State Examination. *Int Psychoger*, 2006, Sep, 18(3):505-15.
6. Crane PK, Gibbons LE, Narasimhalu K et al. Rapid detection of differential item functioning in assessments of health-related quality of life: The Functional Assessment of Cancer therapy. *Qual Life Res*, 2007, 16:101-114.

7. Fleck MP, Chachamovich E and Trentini C. Development and validation of the Portuguese version of the WHOQOL-OLD module. *Rev Saude Publica*. 2006 Oct;40(5):785-91
8. Fleck, M.P., Chachamovich, E., Trentini, C.M. WHOQOL-OLD Project: method and focus group results in Brazil. *Rev. Saúde Pública*, 2003, 37(6): 793-799
9. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Reumathol*, 1995, 24(2):61-3.
10. Hawthorne G, Davidson N, Quinn K et al. Issues in conducting cross-cultural research: implementation of an agreed international protocol designed by the WHOQOL Group for the conduct of focus groups eliciting the quality of life of older adults.. *Qual Life Res*, 2006, Sep, 15(7):1257-70.
11. Haywood, K.L., Garrat, A.M., Fitzpatrick, R. Quality of life in older people: A structured review of generic self-assessed health instruments. *Qual Life Res*, 2005, 14:1651-1668.
12. Kline RB. Principles and practice of structural equation modelling. 2005, Guilford Pres, 2nd edition.

13. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions*, 1994, 7:28.
14. Pallant J and Tenant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*, 2007, 46:1-18.
15. Pallant J, Miller R and Tenant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry*, 2006, 6:28-38.
16. Power, M., Quinn, K., Schmidt, S., WHOQOL-OLD Group. Development of the WHOQOL-Old module. *Qual Life Res*, 2005, 14:2197-2214
17. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press, 1960.
18. Sheik JI and Yesavage JA (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin Gerontol*, 37:819-820.
19. Swaine-Verdier A, Doward LC, Hagell P et al. Adapting Quality of Life Instruments. *Value in Health*, 2004, 7 (Suppl 1):S27-S30.

20. Tammaru M, McKenna SP, Meads DM et al. Adaptation of the rheumatoid arthritis quality of life scale for Estonia. *Rheumatol Int*, 2006, 26:655-662.
21. Tenant A, McKenna SP and Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instrument. *Value in Health*, 2004, 7(Suppl 1):S22-S26.
22. Tenant A and Pallant JF. Unidimensionality matters! (A tale of two Smiths?). *Rasch Measurement Transactions*, 2006, 20:1048-1051.
23. The WHOQOL group. The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties, 1998. *Soc Sci Med*; 1988, 46:1569-85
24. United Nations. World Population Prospects: The 2002 revision. New York: United Nations Population Division, 2003.
25. WHO. Active Ageing: A Policy Framework. Geneva: World Health Organisation, 2002.

Table 1 – Sample demographics (n=424)

	N (%) or M (SD)
Age	
60-69 y.o.	173 (40.9)
70-79 y.o.	153 (36.2)
80y.o.	97 (22.9)
Gender	
Male	152 (35.8)
Female	272 (64.2)
Perceived Health Status	
Healthy	286 (67.5)
Unhealthy	138 (32.5)
Marital Status	
Single	29 (6.8)
Married	212 (50.0)
Separated	30 (7.1)
Widowed	128 (30.2)
Educational Level	
Illiterate	7 (1.7)
Basic Level	165 (38.9)
High School	110 (25.9)
College	90 (21.2)
Depression Level	
GDS 15 (Means, SD)	3.99 (2.91)

Table 2 – WHOQOL-OLD items descriptions

Item content	Mean	SD	MV(%)	Distribution					Skew	Kurt
				1	2	3	4	5		
Sensory Abilities										
1 Impairments to senses affect daily life	3.89	1.1	0.2	2.4	11.8	20.8	24.1	40.9	-.663	-.641
2 Loss of sens abilities affect participation in activities	4.05	1.0	0.0	1.4	10.6	17.0	23.3	47.6	-.853	-.400
3 Problems with sens functioning affect ability to interact	4.06	1.2	0.2	6.4	5.9	14.4	22.0	51.3	-1.18	-.392
4 Rate sensory functioning	3.78	.85	0.0	.7	6.6	25.5	48.1	19.1	-.486	.061
Autonomy										
5 Freedom to make own decisions	3.90	.96	0.2	2.4	9.0	10.6	52.2	25.8	-1.04	.862
6 Feel in control of your future	3.34	1.0	0.0	6.4	16.5	26.4	38.4	12.3	-.424	-.524
7 People around you are respectful of your freedom	3.80	.89	0.0	1.7	7.5	19.1	52.6	19.1	-.804	.655
8 Able to do things you'd like	3.50	1.0	0.0	2.8	13.4	34.9	29.0	19.8	-.195	-.593
Past, Present and Future Activities										
9 Satisfied with opportunities to continue achieving	3.37	1.0	0.0	4.5	16.5	30.2	34.9	13.9	-.304	-.528
10 Received the recognition you deserve in life	3.63	.95	0.0	3.1	8.5	26.4	46.2	15.8	-.664	.306
11 Satisfied with what you've achieved in life	3.96	.85	0.0	.7	6.4	15.1	52.1	25.7	.852	.654
12 Happy with things to look forward to	3.71	.77	0.0	1.4	3.3	30.4	52.8	12.0	.774	1.09
Social Participation										
13 Have enough to do each day	3.53	.95	0.0	2.6	10.6	32.5	39.4	14.9	.957	-.130
14 Satisfied with the way you use your time	3.72	.89	0.0	1.2	11.1	17.7	55.0	15.1	.894	.266
15 Satisfied with level of activity	3.62	.98	0.0	2.4	13.7	19.8	48.1	16.0	.987	-.162
16 Satisfied with opportunity to participate in community	3.49	.96	0.0	1.9	12.7	34.7	35.6	15.1	.960	-.409
Death and Dying										
17 Concerned about the way you will die	3.71	1.2	0.0	3.5	17.9	18.4	24.3	35.8	1.22	-.989
18 Afraid of not being able to control death	3.81	1.2	0.0	3.1	17.7	16.0	21.9	41.3	1.23	-.964
19 Scared of dying	3.86	1.2	0.0	3.3	14.4	17.7	22.6	42.0	1.20	-.746
20 Fear pain before death	2.87	1.2	0.2	12.5	36.6	19.1	15.1	16.5	1.29	-1.04
Intimacy										
21 Feel a sense of companionship in life	3.60	1.0	0.4	7.1	7.6	20.6	47.4	17.3	1.08	.309
22 Experience love in your life	3.88	.93	0.4	3.8	5.0	13.0	56.2	22.0	.937	1.78
23 Opportunities to love	3.62	1.1	0.4	6.6	8.8	20.9	43.1	20.6	1.10	.056
24 Opportunities to be loved	3.55	.94	0.4	5.0	11.4	25.4	40.8	17.5	1.06	-.173

MV = missing values

Table 3 – Fit statistics for the refined WHOQOL-OLD Module

Items / Domains	Domain χ^2 Fit	P value	PSI	Item χ^2 Fit	Item Residual	Location	DIF Gender*	DIF Age*
<i>Sensory Abilities</i>	59.28	.0000*	.737					
Impairments to senses affect daily life				21.615*	-1.355	0.023		
Rate sensory functioning				26.712*	-0.342	-0.11		
Problems with sensory functions affect ability to interact				10.954*	0.684	0.087		
<i>Autonomy</i>	34.11	.082	.713					
Freedom to make own decisions				11.06	-0.107	-0.385		
Feel in control of your future				11.443	0.242	0.584		
Able to do things you'd like to				7.159	2.101	0.078		
People around you are respectful of your freedom				4.478	0.237	-0.277		
<i>Past, Present and Future Activities</i>	27.92	.262	.788					
Happy with things to look forward to				13.138	-0.167	-0.027		
Satisfied with opportunities to continue achieving				4.943	-1.906	0.53		
Received the recognition you deserve in life				3.5	0.909	0.169		
Satisfied with what you've achieved in life				6.35	1.115	-0.672		
<i>Social Participation</i>	28.69	.095	.812					
Satisfied with the way you use your time				7.709	-2.094	-0.292		
Satisfied with level of activity				10.976	-2.308	0.008		
Have enough to do each day				6.672	2.565	0.166		
Satisfied with opportunity to participate in community				3.277	1.068	0.118		
<i>Death and Dying</i>	30.90	.017	.800					
Concerned about the way you will die				15.92*	-.93	-0.351		
Scared of dying				11.53	-0.75	-0.507		
Fear pain before death				3.44	2.22	0.858		
<i>Intimacy</i>	14.85	.535	.867					
Feel a sense of companionship in life				1.995	0.886	0.241		
Experience love in your life				8.183	-2.101	-0.537		
Opportunities to love				3.44	-2.37	0.116		
Opportunities to be loved				1.232	0.383	0.18		
Overall Score (24 items)	563.30	.0000*	.889					

* Significant at a p value <.05 corrected for Bonferroni Multiple Comparisons

Figure 1 – Item 5 Category Probability Curves in original response scale and after rescoring

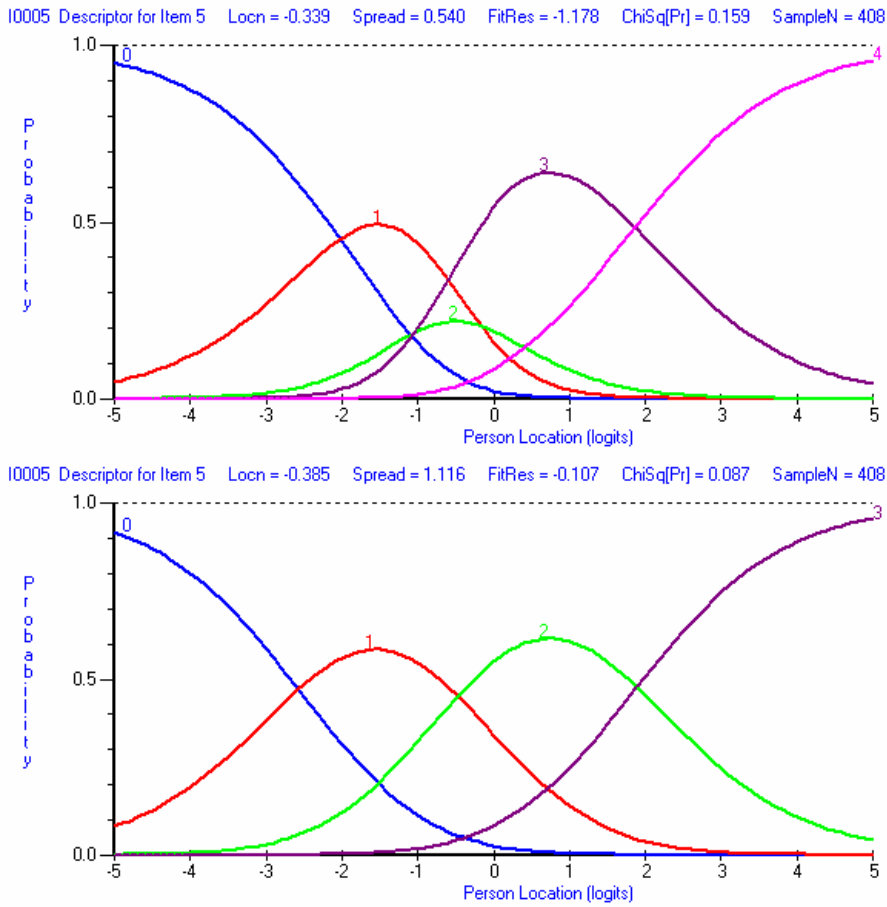
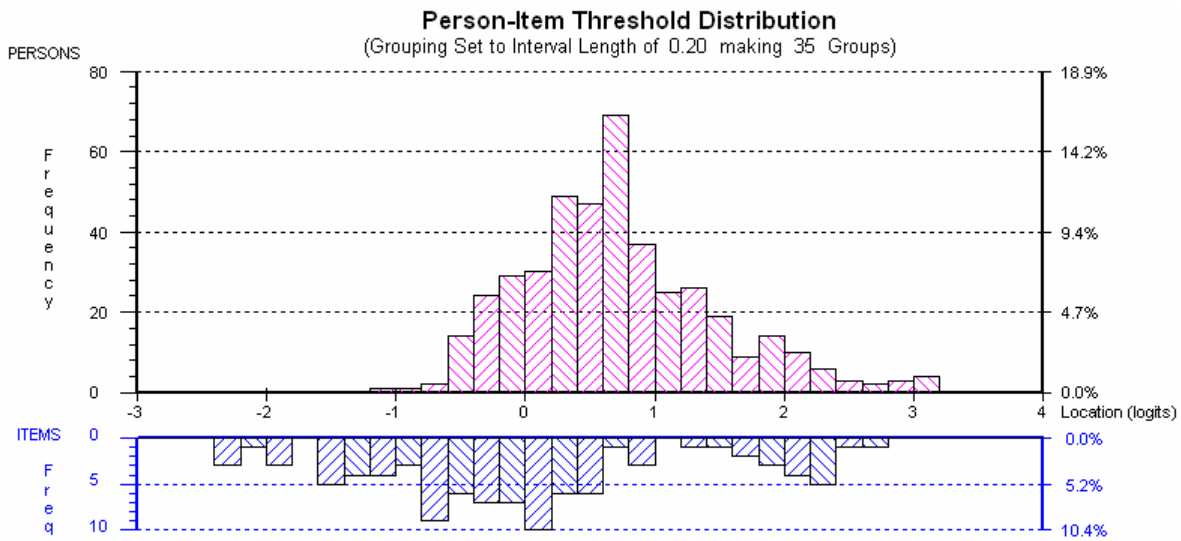


Figure 2 – Person item distribution graph for the WHOQOL-OLD scale



ARTIGO 2

Development and validation of the Brazilian version of the Attitudes to Aging Questionnaire (AAQ): an example of merging classical psychometric theory and the Rasch measurement model

Publicado na revista Health and Quality of Life Measurement

Development and validation of the Brazilian version of the Attitudes to Aging Questionnaire (AAQ): an example of merging classical psychometric theory and the Rasch measurement model

Chachamovich E^{1,2ξ}, Fleck MP^{1*}, Trentini CM^{1*}, Laidlaw K^{2*}, Power MJ^{2*}

¹ Post-Graduate Program of Psychiatry, Universidade Federal do Rio Grande do Sul, Brazil

² Section of Clinical and Health Psychology, University of Edinburgh, UK

* These authors contributed equally to this work

ξ Corresponding author

Address for correspondence: Eduardo Chachamovich
Rua Florencio Ygartua, 391/308
Porto Alegre – RS
CEP 90430-010
Brasil

Email: echacha.ez@terra.com.br
Phone : +55 51 91230015
Fax : +55 51 32644152

Email addresses:

EC: echacha.ez@terra.com.br
MPF: mfleck.voy@terra.com.br
TCM: clarissatrentini@terra.com.br
KL: klaidlaw@ed.ac.uk
MJP: mjpower@ed.ac.uk

ABSTRACT

Background: Aging has determined a demographic shift in the world, which is considered a major societal achievement, and a challenge. Aging is primarily a subjective experience, shaped by factors such as gender and culture. There is a lack of instruments to assess attitudes to aging adequately. In addition, there is no instrument developed or validated in developing region contexts, so that the particularities of ageing in these areas are not included in the measures available. This paper aims to develop and validate a reliable attitude to aging instrument by combining classical psychometric approach and Rasch analysis.

Methods: Pilot study and field trial are described in details. Statistical analysis included classic psychometric theory (EFA and CFA) and Rasch measurement model. The latter was applied to examine unidimensionality, response scale and item fit.

Results: Sample was composed of 424 Brazilian old adults, which was compared to an international sample (n=5238). The final instrument shows excellent psychometric performance (discriminant validity, confirmatory factor analysis and Rasch fit statistics). Rasch analysis indicated that modifications in the response scale and item deletions improved the initial solution derived from the classic approach.

Conclusions: The combination of classic and modern psychometric theories in a complementary way is fruitful for development and validation of instruments. The construction of a reliable Brazilian Attitudes to Aging Questionnaire is important for assessing cultural specificities of aging in a transcultural perspective and can be applied in international cross-cultural investigations running less risk of cultural bias.

Keywords: aging, attitude, psychometrics, AAQ, Rasch

BACKGROUND

The world is experiencing a profound and irreversible demographic shift as older people are living longer and healthier than ever before [1,2]. The world's older adult population is estimated to show a threefold increase over the next fifty years, from 606 million people today to 2 billion in 2050 [2]. In 2002, older people constituted 7 per cent of the world's population and this figure is expected to rise to 17 per cent globally by 2050 [3]. The most dramatic increases in proportions of older people are evident in the oldest old section of society (people aged 80 years plus) with an almost fivefold increase from 69 million in 2000 to 377 million in 2050 [4].

The World Health Organisation has described this demographic shift as a major societal achievement, and a challenge [5]. The increase in longevity is being experienced in the developed and the developing world alike, but where the developed world grew rich before it grew old, the developing world is growing old before it has grown rich [5]. While older people are living longer they are generally remaining healthier with an increase in percentage of life lived with good health. Nonetheless older people are still seen as net burdens on society rather than net contributors to it [5,6].

Quantifying the raise of proportion of old adults in the world population is relevant but insufficient. It is also important to study the quality of this increase. The experience of ageing is primarily subjective and depends on several factors, such as gender, physical condition, environment, behavioural and social determinants, psychological strategies and culture [5,7,8,9,10]. Culture is considered particularly relevant since it shapes the way in which one ages due to the influence it has on how the elderly are seen by a determined context [5]. Moreover, the cultural aspects could be understood as a pathway through which the external aspects would impact on ageing experiences.

Authors state that the vast majority of research and discussion is done by young adults, whereas older adults would be the most indicated to propose adequate ways of doing it [11,12]. Bowling and

Diepe argue that lay viewers are important for testing the validity of existing models and measures, since most of the discussion tends to reflect only the academic point of view [13]. Even though investigating the ageing process has been a topic of increased interest, there is a remarkable lack of well-designed and tested instruments to assess it. The few developed so far are either not specific to cover older adult's experiences or have been exclusively carried out in developed countries [14]. As far as we are aware, there is no instrument developed or validated in developing region contexts, so that the particularities of ageing in these areas are not included in the measures available.

To address this issue, the WHOQOL Group has developed the AAQ instrument under a simultaneous methodology [15], which ensured the participation of different centres throughout the world (described in details in Laidlaw *et al*, 2007) [14]. Briefly, the development process included centres from distinct cultural contexts in qualitative item generation, piloting and field testing. The applied methodology followed the one established by the World Health Organization Quality of Life Group [16,17] for the development and adaptation of quality of life measures and was used for the development of the WHOQOL-OLD module [18,19].

Regarding development of new measures or validation of existing ones, new approaches have been added to the traditional ones in order to expand the scale's properties beyond reliability and validity [20]. The Rasch model has been adopted since it permits that data collected may be compared to an expected model and allows testing other important scale features, such as reversed response thresholds and differential item functioning.

The present paper aims to illustrate the potential combination of classical psychometric theory and Rasch Analysis in the validation of the AAQ instrument in a Brazilian sample of older adults.

METHODS

PILOT STUDY

The pilot study followed the methodology applied by the WHOQOL Group in developing quality of life measures [16,17]. This includes translation and back-translation of the items and instructions by distinct professionals, as well as semantic and formal examination by the coordinator centre. Convenience sampling was used. The main purpose of this stage was to collect data about the item performance in order to produce a reduced version after refinement. The combination of classical and modern (item response theory) statistical analyses was used at this point. A set of 44 items were tested in an opportunistic sample of 143 subjects (age range 60-99, 59% female, 55% living alone, and 59% considered themselves subjectively healthy). Patients with dementia, other significant cognitive impairments and/or terminal illness were excluded. Data collected at this stage were sent to the coordinator centre to be merged with other centres' information.

Statistical analyses were carried out to check the items regarding missing values, item response frequency distributions, item and subscale correlations and internal reliability. No missing values were found in any of the 44 items in the Brazilian sample. The analysis of the pooled international data indicated the need of item refinement, which resulted in a 38-item version to be tested in the field trial (see Laidlaw *et al* (2007) for more details on this refinement stage) [14].

FIELD TRIAL

The Brazilian Field Trial was carried out with a non-probabilistic opportunistic sample of 424 older adults recruited from a university hospital, community houses and nursing homes, elderly community groups, and their own homes. Subjects were invited to take part of the study and were asked to indicate other potential participants (snowball strategy). Sampling was used according to previous stratification determined by subjective perception of health status (50% healthy ones and 50% unhealthy ones), gender (50% female) and age (60-69 years of age, 70-79 years of age and over 79 years of age).

Subjective perception of health status was assessed by the question “In general, you consider yourself healthy or unhealthy?”, regardless of the objective health condition. Exclusion criteria followed the ones used in the pilot study [14]. The purpose of stratification was to ensure a minimal representation in each subgroup to make further analyses possible.

This version comprised the 33 items from the Pilot Study plus 5 items added by the Coordinator Centre (Edinburgh) in order to cover areas not sufficiently investigated by the original format. These 5 items were translated and back-translated and re-examined by the coordinator centre. In addition, subjects completed a socio-demographic form and the Geriatric Depression Scale 15-item version [21].

Statistical analysis

The combination of classical and modern psychometric approaches was applied. The descriptive data analysis was used to determine item response frequency distributions, missing values analysis, item and subscales correlations and internal reliability analyses. Exploratory and Confirmatory Factor analysis were performed to assess whether the Brazilian data fit the international pooled model. Finally, an IRT approach, in particular, that of the Rasch model as implemented in the RUMM 2020 program [22], was used to examine the performance of items in the Brazilian dataset.

RESULTS

Demographics. Table 1 describes the socio-demographic characteristics of both the Brazilian and the international samples. Note that the international sample is composed of the data collected in all centers apart from Brazil. Chi-Square and Independent T-tests were carried out to check statistical

differences across both samples. Following the detection of differences in gender and educational level distributions, as well as in the mean depression level, an Independent T-test was then run to compare means of the three original AAQ factor scores (as described in Laidlaw *et al*, 2007) [14] between the two samples. Briefly, the factor scores were calculated by summing the items included in each factor. Results indicate statistical differences in all three factor scores, as well as in the overall score.

TABLE 1

An Ancova analysis was then carried out to assess the extent to which the interaction among depression, gender and educational level was implied in determining differences in the scores (overall and each factor). Comparisons between both samples were run to rule out the possibility that differences in posterior factor analyses are due to distinct sample characteristics. Table 2 illustrates the Ancova findings, indicating that the statistical difference in the distribution of these variables between the two samples does not interfere significantly with the score variations [23].

TABLE 2

Descriptives. Summary descriptives statistics for item analyses are shown in Table 3. There is low frequency of missing values across the items. Comparison of the missing frequencies with the international dataset showed a lower frequency in the Brazilian sample.

TABLE 3

Exploratory Factor Analysis. Data were initially examined through Exploratory Factor Analysis (Principal Component Analysis with Varimax Rotation). Extraction strategy included selecting factors with eigenvalues higher than 1 (and confronted to Monte Carlo Parallel Analysis to control for spurious findings) and scree plot observation [24,25,26]. The three-factor solution (indicated both by the Kaiser Rule plus Parallel Analysis and Scree Plot) accounted for 34.45% of the total variance, whereas in the international sample the same structure was responsible for 32.74%.

Figure 1 shows the Scree Plot for both the Brazilian and International Samples, indicating remarkable similarities between both.

FIGURE 1

EFA findings were compared to the international ones. There is a great similarity of the item loadings when comparing to the EFA run in the international dataset. Out of 38 items, only five (items 4, 5, 9, 15 and 31) loaded onto different factors across both datasets. It is important to notice that items 4 and 31 were not retained in the final AAQ version since they lowered CFA results in further international analyses.

The item reliability was analyzed through Cronbach's alpha coefficients for the three subscales suggested by the EFA. The Brazilian dataset showed coefficients of .863 for the Subscale I (and .845 for the International dataset), .804 for the Subscale II (.822 for the International sample) and .671 for the Subscale III (.701 for the International subscale).

The Item Total Correlation Analysis was then carried out in distinct steps. Firstly, the Brazilian dataset was analyzed to verify correlations below a critical cut-point ($r=0.400$). Secondly, the International dataset underwent the same analysis. Thirdly, both findings were compared to verify potential discrepancies. Six items in the Brazilian dataset showed insufficient correlations (items 1,5,6,11,18 and 19). All these six items proved to show low coefficients in the International dataset too. Out of these, only item 18 remained in the final international AAQ version.

The Multi-trait Analysis Program (MAP) [27] was also used to assess scale fit and internal reliability of the three-factor model. Although six items loaded highly on other factors besides the predicted one (9, 13, 21, 24, 33 and 34, $r \geq .409 < .528$), no items presented higher correlations with an unpredicted factor than with the predicted one. Furthermore, the directions presented by the MAP analysis (correlation coefficients) were in accordance with the EFA loadings.

Confirmatory Factor Analysis. CFA was carried out using AMOS 6.0 software [28]. First, the 38 items three-correlated-factor solution was tested, showing insufficient results ($\chi^2 = 1516.605$ $p < .001$, $df = 662$, $CFI = 0.792$, $RMSEA = 0.055$). In order to verify the impact of the correlation among factors, the uncorrelated solution was then tested, showing further decrease in model fit ($\chi^2 = 1943.632$ $p < .001$, $df = 665$, $CFI = 0.688$, $RMSEA = 0.067$).

Following the steps adopted by the international development of AAQ [14], the 31-item three-factor solution was then assessed in order to verify potential improvement in model fit. Similarly to the international findings, this structure showed insufficient improvement ($\chi^2 = 1005.621$ $p < .001$, $df = 431$, $CFI = 0.820$, $RMSEA = 0.056$). Again, allowing interfactor correlation determines great model fit improvement.

The final 24-item version was also tested in the Brazilian dataset, according to the structure illustrated in Figure 2.

FIGURE 2

Remarkable improvements in model fit were shown ($\chi^2 = 645.190$ $p = .061$, $df = 249$, $CFI = .835$, $RMSEA = .061$). The comparison of these indexes to the international ones indicate that the performance of the Brazilian final version is similar (international findings present $CFI = .842$ and $RMSEA = .056$)

Discriminant validity. To assess the discriminant validity, a correlation between each domain score and the depression levels was performed. It was predicted that depression levels would be negatively correlated to the three factors, and that the physical factor should present a lower coefficient than the two psychological factors. In fact, the correlation results showed coefficients of $r = -.595$ with psychosocial loss, $r = -.591$ with psychological growth and $r = -.356$ with physical change.

Item Response Theory. Responses were tested according to the Rasch model for polytomous scales [29]. Basically, the responses patterns observed in data collected are tested against an expected probabilistic form of the Guttman Scale [30]. Different fit statistics are applied to determine whether the observed data fits the expected model or not [31]. According to Rasch measurement theory, a scale should have the same performance, independently of the sample being assessed (e.g., age or gender) [20,21]. Reverse thresholds, an overall Chi-Square test (indicating whether the observed data differs from the expected model), item Chi-Square fit and Item fit-residuals were tested. In addition to these fit indexes, the item bias DIF (differential item functioning) was verified, since it can determine decrease in model fit, as well as measurement inappropriateness. The Person Separation Index (PSI) was calculated for each factor as

an indicator of internal consistency reliability. In fact, the PSI gives information comparable to the Cronbach's Alpha from classic psychometric theory.

Table 4 presents the Rasch findings for the 24-item version in its original form. At this stage, the 5-point Likert response scale was maintained in its original form. As mentioned above, the Chi-Square (both for the model and for items separately) has the purpose of assessing whether the data collected fits the expected theoretical model. Thus, p values lower than 0.05 (corrected for Bonferroni Multiple Comparisons) indicate that the first is significantly different from the second, rejecting the desired similarity [32]. Item residuals (a sum of item and individual person deviations) also permit the assessment of item fit, and values from -2.5 to +2.5 show adequate fit.

TABLE 4

Results described in Table 4 show that 6 items (9, 14, 15, 19, 21 and 22) presented high residuals and/or item χ^2 scores significantly different from the expected. The model fit for the three subscales also indicated misfitting. Furthermore, 15 out of 24 items presented threshold disorders, which suggests that the response scale is not adequate and therefore contribute to the misfittings found both in model and item levels.

Thus, rescaling items was carried out in order to improve the model. Firstly, the category probability curves were checked for each item. This approach allows the investigator to verify what response categories present disorders and, thus, what specific categories should be collapsed to improve the scale. Factors I and II demanded that categories two and three were merged, whereas factor III needed categories 3 and 4 collapsed together.

Analysis using the new 4-point scale showed that Factors I and III had remarkable improvement, with no model or item misfittings. On the other hand, Factor II presented a slight increased fit, but still insufficient (Model $\chi^2 = 87.127$, DF 48, $P=0.0004$, $PSI=.752$). The second step was then deleting the items responsible for the remaining misfitting, namely items 19 and 22. The final model, then, proved adequate fit. No reversed threshold or DIF remained after rescaling and item deletion (Factor II). Person Separation Indexes showed adequate scores for group comparisons (i.e., $PSI > .70$). Table V presents the indexes for the final model.

Local independence of items and unidimensionality (two Rasch assumptions) were assessed for the three final factors through two statistical tests. Item residuals correlations were firstly analysed to check the potential presence of local dependence (i.e., two items highly correlated in the final model, so that the response to one would be determined by the other). No correlations above 0.300 were found, which indicates local independence. Secondly, the pattern of residuals was analysed thorough PCA of the residuals. The first PCA factor was divided into two subsets (defining the most positive and negative loadings on the first residual component). These two subsets were then separately fitted into Rasch Model and the person estimates were obtained. An Independent T-test was then carried out to detect potential differences between the two subsets, which would indicate the presence of multidimensionality in the model [20]. No significant differences were found for the three factors of the scale (Factor 1, $p=0.051$, Factor 2 $p=0.654$, Factor 3 $p=0.090$).

TABLE 5

Discussion

The present paper had two complementary aims. First, it had the goal of presenting a validated Brazilian version of the Attitudes to Aging Scale. This version will permit that aging experiences may be assessed in a distinct and poorly investigated population. Furthermore, since aging is a widespread phenomenon and is highly dependent on socio-cultural aspects, it is extremely important that new measures of this construct can be successfully applied in different contexts. This would permit that adequate cross-cultural investigations on attitudes to aging may be carried out, including a valid and reliable instrument.

Secondly, this article aims to present a comprehensive approach in validating new measures, which include both classical psychometric theory and modern methodologies together in a complementary way. While the traditional approach provides relevant information regarding discriminant validity, missing values distributions and factor analyses loading, Rasch analysis represents a powerful tool in assessing item bias, threshold disorders and model fit [20].

The Attitudes to Aging Questionnaire is a unique measure of perception regarding aging, since it was developed through a well-established international methodology and based since its principle in focus groups run with older adults [15,16,17,33]. Furthermore, it relies on the assumption that the subjective perception of the aging process is the ultimate construct to be measured, other than objective indicators of physical activity or psychological distress.

Regarding the psychometric performance, the Brazilian version demonstrates good performance on both classical and Rasch approaches. Despite the insufficient goodness-of-fit indexes in CFA ($CFI < .90$), suitable discriminant validity, and excellent fit indicators from Rasch analysis suggested that the Brazilian version has satisfactory performance and, thus, can be applied in different studies reliably.

Another relevant issue regarding the findings of the AAQ validation is the construct similarity between the international sample and the Brazilian one. The three factors proposed by the international analysis seem to be replicated in the Brazilian dataset. Indeed, Psychosocial Loss, Physical Change and Psychological Growth represented the theoretical ground upon which items were grouped during the

factor analysis phase. It could indicate that the perception of aging did not differ significantly between the two samples and raises the question of whether these similarities remain or not in other different cultures. The demonstration of cultural invariance of the core attitudes to aging could lead to the possibility of reliable comparisons, which is needed by both researchers and policy makers.

It is suggested, however, that rescaling and two item deletions could increase Brazilian scale fit and performance. These potential alterations should not promote crucial modifications in the scale format, since they can be made during the statistical analysis phase and not necessarily in the data collection stage. Since this is the first psychometric analysis of the Brazilian AAQ version, authors encourage the scale users to verify whether the 22-item version maintains its superiority over the original 24-item format in distinct samples, and then explicitly decide for one format.

CONCLUSIONS

The described findings support the hypothesis that the development of a new international instrument according to a simultaneous methodology, which includes an intense qualitative initial phase, is adequate to generate reliable cross-cultural measures. In conclusion, the Brazilian version of the AAQ instrument is a reliable, valid and consistent tool to assess attitudes to aging and can be applied in international cross-cultural investigations running less risk of cultural bias.

COMPETING INTERESTS

Authors do not have any conflict of interest.

AUTHORS' CONTRIBUTIONS

EC participated in the study design, data collection, statistical analysis and drafted the manuscript; MPF participated in the study design, statistical analysis and helped to draft the manuscript; CMT participated in the study design and data collection; KL helped to draft the manuscript and took part in the theoretical discussion; MJP participated in the study design, statistical analysis and helped to draft the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This paper was partially supported by CAPES, scholarship number PDEE 3604-06/3

References:

1. Kinsella, K & Velkoff, V. A. **US Census Bureau, Series P95/01-1, An Aging World: 2001.** Washington D.C.: US Government Printing Office; 2001.
2. United Nations. **World Population Ageing: 1950-2050.** Department of Economic and Social Affairs, Population Division. New York: United Nations Publications; 2001.
3. US Census Bureau. **International Population Reports WP/02, Global Population Profile: 2002.** U.S. Government Printing Office, *Washington, DC*; 2004.
4. United Nations. **World Population Prospects: The 2002 revision.** *United Nations Population Division*; New York; 2003.
5. WHO. **Active Ageing: A Policy Framework: World Health Organisation .**Geneva; 2002.
6. WHO. **Ageing: Exploding the Myths.: World Health Organisation..** Geneva; 1999.
7. Baltes and Smith. **New frontiers in the future of aging: from successful aging to the young old to the dilemmas of the fourth age.** *Gerontology* 2003, Mar-Apr;49(2):123-35.
8. Levy Br, Slade MD, Kunkel SR, Kasl, SV. **Longevity increased by positive self-perceptions of aging.** *J Pers Soc Psychol* 2002., 83(2): 261-270
9. Knight BG. *Psychotherapy with Older adults*, 3rd Edition. Thousand Oaks: Sage Publications 2004.
10. Ebner, NC., Freund, AM. and Baltes, PB. **Developmental changes in personal goal orientation from young to late adulthood: From striving for gains to maintenance and prevention of losses.** *Psychology and Aging* 2006, 21: 664-678.
11. Duhl LJ. **Aging by one who is aging.** *J Epidemiol Community Health* 2005, 59(10):816-7.
12. Boduroglu A, Yoon C, Luo T and Park DC. **Age-related stereotypes: A comparison of American and Chinese cultures.** *Gerontology* 2006; 52:324-333.
13. Bowling A and Dieppe P. **What is successful ageing and who should define it?** *BMJ* 2005, 331:1458-1551.
14. Laidlaw, K., Power, M. J., Schmidt, S. and the WHOQOL Group **The attitudes to ageing questionnaire (AAQ): Development and psychometric properties.** *Int J Geriatr Psychiatry* 2007, 22: 367-379.
15. Bullinger M, Power M, Aaronson NK, Cella DF, Anderson RT: **Creating and evaluating cross-cultural instruments.** In: *Quality of life and pharmacoeconomics in clinical trials.* Edited by B. Spilker. Hagerstown, MD. Lippincott-Raven 1996, 659-668.

16. The WHOQOL Group. **The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties.** *Soc Sci Med* 1998, 46:1569-85
17. The WHOQOL Group. Development of The World Health Organization WHOQOL-BREF Quality of Life Assessment. *Psychol Med* 1998, 28:551-558.
18. Power MJ, Quinn K, Schmidt S *et al.* **Development of the WHOQOL-old module.** *Qual Life Res* 2005, 14(10):2197-214.
19. Fleck MP, Chachamovich E, and Trentini C. **Development and validation of the Portuguese version of the WHOQOL-OLD module.** *Rev Saude Publica* 2006, 40(5):785-91.
20. Pallant J and Tenant A. **An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS).** *Br J Clin Psychol*, 2007, 46:1-18.
21. Sheik JI and Yesavage JA. **Geriatric Depression Scale (GDS): recent evidence and development of a shorter version.** *Clin Gerontol* 1986, 37:819-820.
22. Andrich D, Lyne A, Sheridan B and Luo G. *RUMM 2020*. Perth: RUMM Laboratory; 2003.
23. Field A. **Discovering Statistics using SPSS.** 2nd, SAGE, London, 2005.
24. Kauffman JD and Dunlap WP. **Determining the number of factors to retain: a Widows-based FORTRAN-ISL program for parallel analysis.** *Behav Res Methods Instrum Comput* 2000, 32(3):389-95
25. Zwick WR and Velicer WF. **Comparison of five rules for determining the number of components to retain.** *Psychol Bulletin* 1986, 99:432-442.
26. O'Connor BP. **SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP tests.** *Behav Res Methods Instrum Comput* 2000,32(3):396-402
27. Hays RD, Hayashi T, Carson S *et al.*: **User's guide for the Multitrait Analysis Program (MAP).** Santa Monica, CA. The Rand Corporation, N-2786-RC; 1988.
28. Arbuckle JA: **Amos 6.0 User's Guide.** Amos Development Corporation. Spring House, PA, USA, 2005.
29. Andrich D. **Rating formulation for ordered response categories.** *Psychometrika* 1978, 43:561-573.
30. Gutman L: **The basis of scalogram analysis.** In: *Measurement and prediction*. Edited by Stouffer SA. Princeton, NJ: Princeton University Press, 1950.
31. Smith EV. **Detecting and evaluation of the impact of multidimensionality using item fit statistics and principal component analysis of residuals.** *Journal of Appl Meas* 2002, 2:199-218.
32. Bland JM and Altman DG. **Multiple significant tests: The Bonferroni Method.** *BMJ* 1995, 310:170.

33. Guillemin F. **Cross-cultural adaptation and validation of health status measures.** *Scand J Reumathol* 1995, 24(2):61-3.

Figure 1- Scree Plots illustrating EFA Eigenvalues for Brazilian and International Samples

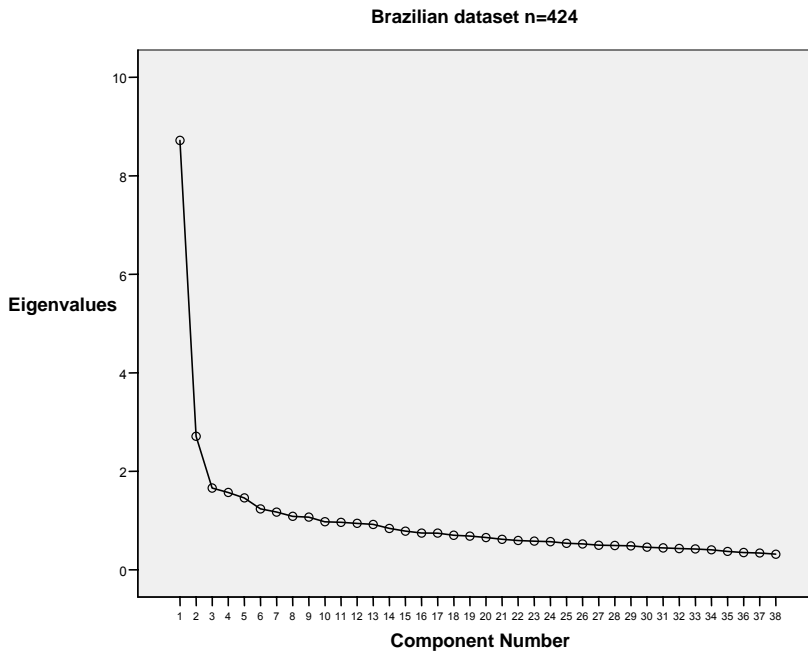
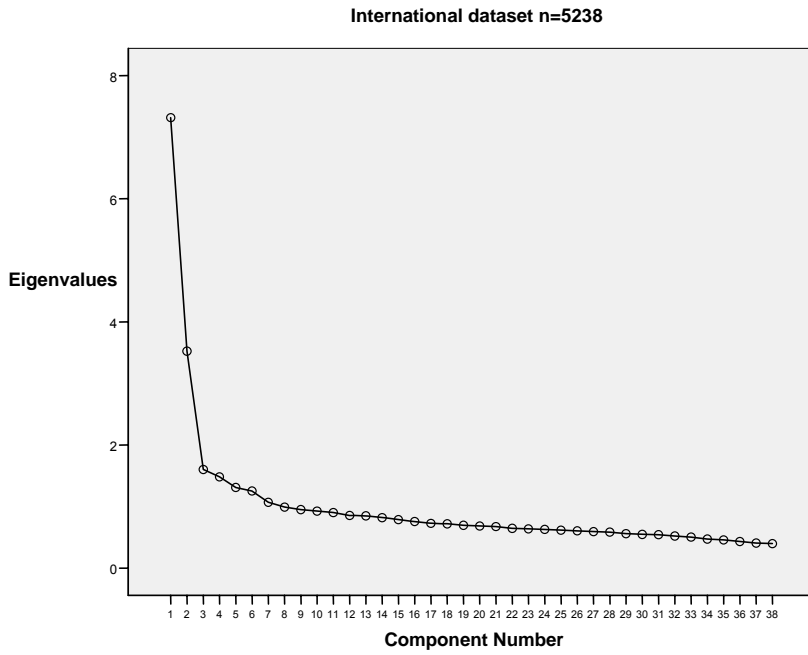


Figure 2. Final 24-version model tested

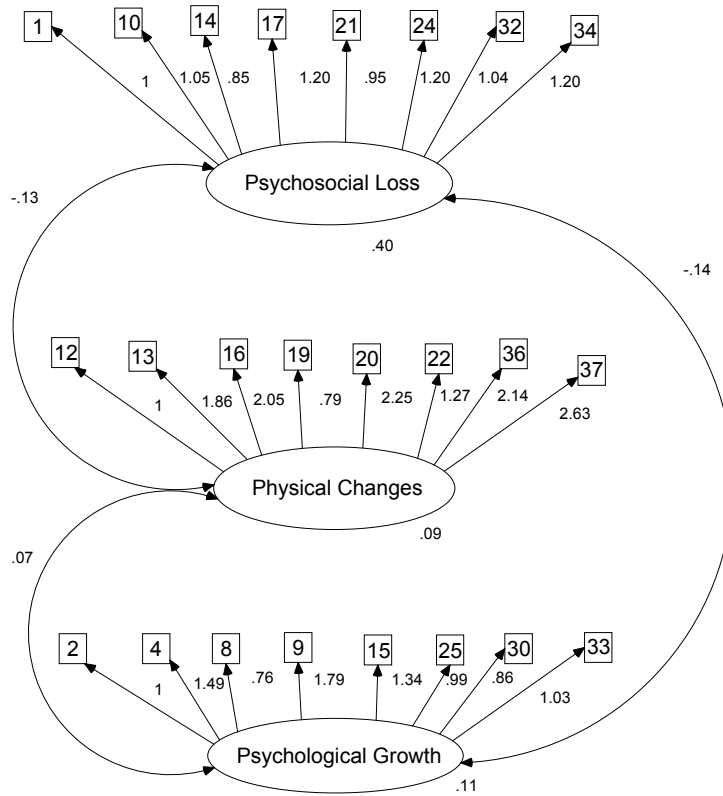


Table 1 . Socio-demographic characteristics of Brazilian and International Samples

	Brazilian sample n=424	International sample n=5238	P
	N (%) or M (SD)	N (%) or M (SD)	
Age			0.64^a
60-69 years old	173 (40.9)	1983 (39.1)	
70-79 years old	153 (36.2)	1948 (38.4)	
80 or + years old	97 (22.9)	1141 (22.5)	
Gender			0.013^b
Male	152 (35.8)	2191 (42.1)	
Female	272 (64.2)	3014 (57.9)	
Perceived Health Status			0.215^b
Healthy	286 (67.5)	3573 (70.8)	
Unhealthy	138 (32.5)	1476 (29.2)	
Marital Status			0.275^a
Single	29 (6.8)	275 (5.5)	
Married	212 (50.0)	2688 (54)	
Separated	30 (7.1)	397 (8)	
Widowed	128 (30.2)	1371 (27.5)	
Educational Level			0.000^a
Illiterate	7 (1.7)	138 (2.7)	
Basic Level	165 (38.9)	1441 (28.3)	
High School	110 (25.9)	1956 (38.4)	
College	90 (21.2)	1449 (28.5)	
Depression Level			0.041^b
GDS 15	3.99 (2.91)	3.68 (2.69)	

^a Chi-Square test; ^b independent t test

Table 2 - Ancova analyses including Educational level, gender and depression between Brazilian and International Samples

Interaction	Means Br	Means Int	F	P	Partial Eta Sq.
Total score					
Gender (m/f)	132.8 / 137.3	129.9 / 128.9	1.231	.267	.000
Ed Level (high/low)	139.3 / 134.5	132.1 / 128.3	18.96	.000	.004
Depression ($\leq 5 / > 5$)	141.2 / 119.4	134.4 / 110.8	2914.5	.000	.430
Gender x Ed Level x Depression	-	-	.084	.773	.000
Factor I score					
Gender	49.4 / 51.1	49.7 / 48.5	13.5	.000	.003
Ed Level	51.8 / 50.5	50.7 / 48.4	37.3	.000	.007
Depression	53.1 / 42.8	51.4 / 40.0	2233.7	.000	.352
Gender x Ed Level x Depression	-	-	.001	.971	.000
Factor II score					
Gender	50.3 / 52.7	49.9 / 49.8	.073	.787	.000
Ed Level	54.1 / 51.1	51.2 / 49.4	14.59	.000	.003
Depression	54.0 / 45.3	51.9 / 42.3	1746.4	.000	.301
Gender x Ed Level x Depression	-	-	1.25	.263	.000
Factor III score					
Gender	33.0 / 33.4	30.2 / 30.3	1.80	.179	.000
Ed Level	33.3 / 33.3	30.2 / 30.9	2.29	.130	.000
Depression	34.0 / 31.2	31.0 / 28.7	304.9	.000	.067
Gender x Ed Level x Depression	-	-	.321	.571	.000

Table 3 – Descriptive analysis of the set of 38 items in the Brazilian sample (n=424)

Item content	Mean	SD	MV(%)	Distribution					Skew	Kurt
				1	2	3	4	5		
1 People as old as they feel	3.42	1.18	0	7.3	19.3	13.7	42.9	16.7	-.52	-.76
2 Better able to cope with life	3.81	.781	0	.9	6.4	16.7	62.3	16.7	.781	1.411
3 Old age time of illness	2.24	1.015	0	25	42.2	17.5	14.4	.9	.554	-.549
4 Privilege to grow old	3.96	.93	0	1.9	6.6	14.6	47.6	29.2	-.96	.82
5 Interested in new technology	3.0	1.02	0	6.8	27.1	30.7	30.2	5.2	-.087	-.748
6 Interested in love	3.64	.881	0	2.4	8	25.2	52.4	12	-.766	.666
7 Old age is a time of loneliness	2.27	1.029	0	23.3	44.1	16.3	14.6	1.7	1.029	-.409
8 Wisdom comes with age	3.76	.872	0	1.4	8.7	18.2	55.9	15.8	.872	.664
9 Pleasant things about growing older	3.79	.826	0	1.2	7.8	16.5	60.1	14.4	.826	1.082
10 Old age depressing time of life	2.38	.997	0	19.1	41.5	22.2	16.5	.7	.997	-.752
11 Capacities and abilities decline with age	3.54	.870	.2	3.1	11.6	18.4	62.4	4.5	-1.145	.832
12 Important to take exercise at any age	4.26	.666	0	.7	1.4	4	59	34.9	.666	4.101
13 Growing older easier than I thought	3.41	.981	0	5.9	9.7	30.2	45.8	8.5	.981	.261
14 More difficult to talk about feelings	2.44	1.118	0	25.9	26.4	26.9	19.1	1.7	1.118	-1.073
15 More accepting of myself	3.10	1.097	0	10.1	18.4	29.2	35.6	6.6	1.097	-.674
16 I don't feel old	3.40	1.132	0	8.3	12.3	25.2	39.4	14.9	1.132	-.389
17 Old age mainly as a time of loss	2.17	1.137	0	38.4	23.3	22.2	14.6	1.4	1.137	-.970
18 Personal beliefs mean more as I grow older	3.61	1.18	0	9.5	8.5	16	44.8	21.5	-.868	-.051
19 My identity is not defined by my age	3.29	1.133	.2	11.6	9.9	25	44.3	9	1.133	-.333
20 More energy than I expected for my age	3.32	1.063	.2	6.9	16.1	23.3	44.7	8.7	1.063	-.408
21 Loss physical independence as I get older	2.80	1.156	0	18.2	20.3	28.5	29	3.8	1.156	-1.039
22 Physical health problems don't hold me back	3.25	1.176	.2	11.1	15.1	22.2	40.4	11.1	1.176	-.686
23 Unhappy with changes in physical appearance	2.16	1.128	.2	38.5	23.9	21.7	14.7	1.2	.496	-.979
24 More difficult to make new friends	2.08	1.162	0	44.8	19.6	18.6	15.8	.9	1.162	-1.030
25 Pass on benefits of experience	3.94	.821	.5	1.4	4.3	15.4	56.6	22.3	.821	1.618
26 Fear losing financial independence	2.36	1.287	.2	38.1	17	19.9	21	4	-.358	-1.239
27 Time to do things that really interest me	3.43	1.00	.5	5.9	11.1	26.1	47.7	9.5	-.741	.109
28 Want continue doing work long as possible	3.58	1.23	.2	10.2	9.5	16.8	39.2	24.3	-.760	-.372
29 Worried I'll become a financial burden to family	2.23	1.28	.2	40.9	21.5	16.5	15.6	5.4	.636	-.855
30 Believe my life has made a difference	3.73	.847	.2	2.4	5.4	22.2	56.5	13.5	.847	1.369
31 Just as meaning now as always	3.73	.931	.5	2.4	9.7	16.8	54.5	16.6	-.882	.602
32 Don't feel involved in society	2.55	1.184	.5	25.9	21.5	25.5	24.3	2.4	1.184	-1.229
33 Want to give a good example	4.07	.735	.2	1.4	1.9	9.7	62.6	24.3	.735	3.619
34 I feel excluded because of my age	2.17	1.143	.2	39.2	20.8	25	13	1.9	1.143	-.928
35 Future fills me with dread	2.12	1.15	.5	41	23	21.8	11.1	3.1	.673	-.597
36 Health is better than expected for my age	3.38	1.122	.2	8.7	13	22	44.4	11.8	1.122	-.361
37 Keep myself fit and active by exercising	3.02	1.284	.5	17.1	17.8	23.7	29.1	12.3	1.284	-1.077
38 Important relationships become more satisfying	3.26	1.03	.2	7.8	12.3	34.5	36.9	8.5	-.499	-.195

Items in bold were retained in the international final version

Table 4 – Rasch Analysis of the original 24-item final version including the 5-point Likert response scale

Content		Model χ^2 Fit (df)	P value	Item χ^2 Fit	Item Residual	Rev Threshold	DIF Analyses	
							Gender	Age (60- 79 / 80+)
Subscale I		77.06 (40)	.00003					
PSI = .869	Item 7			3.08	1.01	✓		
	Item 10			12.77	-0.06			
	Item 14			15.27	3.11			
	Item 17			5.52	-0.60			
	Item 21			21.12	3.49			
	Item 24			11.74	-1.25	✓		
	Item 32			10.70	1.61			
	Item 34			6.38	-1.07	✓		
Subscale II		109.4 (48)	.00001					
PSI = .807	Item 12			10.57	-0.41	✓		Uniform
	Item 13			6.57	.58	✓		
	Item 16			4.65	.02	✓		
	Item 19			42.61	4.96	✓		Uniform
	Item 20			11.79	-1.04			
	Item 22			17.47	3.76	✓		
	Item 36			10.40	.66			
	Item 37			5.34	.32	✓		
Subscale III		59.06 (48)	.131					
PSI = .745	Item 2			1.94	.54	✓		
	Item 4			10.11	-0.31	✓		
	Item 8			3.17	1.24	✓		
	Item 9			19.17	-2.05	✓		
	Item 15			9.01	3.43			
	Item 25			1.34	.37			
	Item 30			6.73	1.58	✓		
	Item 33			7.55	-1.73	✓		

In bold, item-residuals > 2.5 or item χ^2 fit with $p < .05$ corrected for Bonferroni Multiple Comparisons

Table 5 - Final 22-item version, including the rescored 4-point response scale

		Model χ^2 Fit (df)	P value*	Item χ^2 Fit*	Item Residual*	Rev Threshold	Gender	DIF Analyses * Age (60-79 / 80+)	Depression
Subscale I		66.36 (40)	.006						
PSI = .815	Item 7			2.945	-0.276				
	Item 10			9.337	-0.592				
	Item 14			5.263	1.409				
	Item 17			5.339	-1.734				
	Item 21			17.108	2.359				
	Item 24			12.571	-2.492				
	Item 32			6.098	1.00				
	Item 34			7.701	-1.507				
Subscale II		65.56 (42)	.011						
PSI = .750	Item 12			4.019	0.434				
	Item 13			3.446	0.7				
	Item 16			3.515	1.239				
	Item 20			9.203	-0.935				
	Item 36			2.897	-0.439				
	Item 37			9.07	-0.842				
Subscale III		59.38 (48)	.125						
PSI = .710	Item 2			1.622	0.362				
	Item 4			9.552	-0.534				
	Item 8			10.845	0.783				
	Item 9			16.299	-1.409				
	Item 15			5.281	1.273				
	Item 25			1.732	-0.242				
	Item 30			6.889	1.175				
	Item 33			7.163	-1.995				

* all p non-significant for 0.05 after Bonferroni correction

ARTIGO 3

A multi-point Likert Scale proved not to be suitable for illiterate subjects

Aceito para publicação na revista Journal of Clinical Epidemiology

A multi-point Likert Scale proved not to be suitable for illiterate subjectsEduardo Chachamovich^{1,2}, Marcelo P. Fleck¹, Mick Power²

1- Post-Graduate Program on Psychiatry, University Federal of Rio Grande do Sul

2 – Clinical and Health Psychology, Medical School, University of Edinburgh

Corresponding author: Eduardo Chachamovich
Address: Rua Florencio Ygartua, 391/308
Porto Alegre – RS
CEP 90430-010
Brazil

Phone: +55 51 91230015
Fax: +55 51 32644152
Email: echacha.ez@terra.com.br

A multi-point Likert Scale proved not to be suitable for non-readers

Abstract

Objective: to assess the effect of illiteracy on a 5-point Likert scale, using the WHOQOL-BREF response scale. It is hypothesized that illiteracy is related to loss of discriminant power among the five categories response scale.

Study design and setting: In a cross-sectional design, non-readers (n=141), and subjects educated at a graduate level (n=907) were assessed. Statistical analyses combined classic and modern psychometric approaches (Confirmatory Factor Analysis and Rasch Analysis).

Results: There is a remarkable difference in the psychometric performance of items across the two subgroups. Fit indexes proved to be closer to the ideal for the graduate group, but not for the non-reader group. Reducing the number of response categories improved the model exclusively for the illiterate sample. Non-readers do not interpret the scale as a true five category scale, but exhibit a response pattern indicative that only three categories are in fact considered.

Conclusion: These results confirm the hypothesis that a multiple-category response scale is not suitable for non-readers. They suggest that a simpler response scale should be adopted to achieve a more reliable and satisfactory psychometric performance in this population. This effect seems to be stronger when more abstract and subjective constructs are involved.

Keywords: illiteracy, psychometrics, Likert, Rasch, quality of life, scale

Running title: Illiteracy and Likert scales

Word count: 3886

A multi-point Likert Scale proved not to be suitable for non-readers

INTRODUCTION

Illiteracy has been considered a major sociological issue in the last two decades [1]. Reports indicate that as much as 20% of the British population is illiterate [2]. In the USA, the National Adult Literacy Survey indicated that 23% of the population had reading and writing performances compatible to illiteracy levels [3]. Furthermore, findings presented by Unesco point out that about a third of the world's total population is illiterate [4]. Literacy is not restricted only to the ability to read and write, but it is in fact connected to several areas of brain functioning [5,6,7,8].

Several studies have focused the repercussions of illiteracy on functional neuroimaging tests, as well as on morphological exams [6,9,10,11]. Briefly, it has been shown that the inter-hemispheric connections are smaller in non-readers when compared to educated ones. In addition, reduced blood flow in the corpus callosum during some verbal tasks is also reported.

Besides neuroanatomical alterations, there is a growing body of evidence regarding the effects of literacy in neuropsychological test performance [1,5,6,7,12,13,14,15]. It is widely demonstrated that illiteracy has a direct effect on both verbal (naming, verbal list, verbal abstraction, phonologic related words and repetition of pseudowords) and non-verbal tasks (visual recognition memory, figure recall and visual retention). These findings corroborate the reports which indicate that illiteracy is not restricted to verbal functioning pathways [16].

Regarding the health care literature, recent studies indicate that literacy has been emphasized as a major predictor of health status, health-related knowledge and health-related behaviours [15,17,18]. Illiteracy also represents a crucial predictor of cognitive decline [19]. Furthermore, understanding and following medical advice and prescription are significantly affected by lack of literacy [20].

In the mental health field, illiteracy raises a particular concern, since the vast majority of studies in this field use scales and instruments [21]. There are still vast problems in applying instruments and neuropsychological tests for this particular population [22]. It is postulated that hierarchical response scales (such as Likert scales) could not be suitable for non-readers, once they may not be as able to discriminate between several categories as highly educated individuals [21,23].

The effect that low literacy has on comprehension of the response alternatives was examined on the Beck Depression Inventory (BDI) [21]. In a limited sample, authors have reported that low literacy was associated with comprehension problems. The BDI is composed of items with four response options, each one represented by a statement. To the best of our knowledge, there are no reports about the impact of illiteracy on the performance of multiple-point response scales.

Item Response Theory (IRT) is a modern statistical approach and represents a powerful tool to examine instruments and items in detail [24]. The Rasch model, in particular, possesses features capable of assessing individual item statistics and response category performance in depth [25,26,27]. The application of Rasch analysis to assess Likert response scale properties in health measurement have been reported extensively [28,29,30].

This paper aims to assess the effect of illiteracy on the WHOQOL-BREF response scale using Rasch analysis. This instrument is a well-known questionnaire

for the assessment of quality of life developed by The WHOQOL Group [31] and is suitable for clinical and non-clinical samples. The response scale is composed of fixed five-point Likert scales. Since there is evidence in the literature that indicates that Likert scales do not function adequately for people with intellectual disability [32,33]. In addition, Williams and Swanson (2001) reported that low-level readers presented problems with a five-point Likert scale [34]. Thus, we hypothesized that illiteracy could be related to loss of discriminant power among the five categories response scale.

SUBJECTS AND METHOD

SUBJECTS

The data were drawn from a larger study on quality of life in older adults (WHOQOL-OLD). This project was carried out by the World Health Organization Quality of Life Group (WHOQOL) and involved 20 countries around the world. An opportunistic sample of 5566 subjects was assessed using instruments to measure quality of life, depression and socio-demographic data [35].

The complete sample involved subjects from different levels of education, ranging from illiteracy to post-graduate. Between them, 81.17% of the subjects were divided into primary school, high school and college. Out of the total sample, 1048 subjects (18.83%) were selected to compose two subgroups according to the education level. The first comprised non-readers exclusively (n=141), and the second is composed of subjects educated at a graduate level (n=907). Illiteracy was not formally assessed, and subjects were considered to be non-readers by stating that they

were not able to write or read in daily life. Research assistants would invite subjects to try to read the initial sentences of the socio-demographic form to confirm the inability to read. Graduate and illiterate subjects were selected since they represent extreme groups. All the human data included in this study was obtained in compliance with regulations of the Hospital de Clinical de Porto Alegre (Ethics Committee approval number 01.374)

PROCEDURES

The assessment was run in a cross-sectional design. The subjects were required to complete a socio-demographic form and the WHOQOL-BREF instrument. The latter is a 26-item self-report instrument developed by the WHO to assess generic quality of life. Due to its simultaneous development strategy [36,37], it allows for cross-cultural comparisons and is validated for both healthy and unhealthy subjects [31,38,39,40]. In addition, it proved to be suitable for older adult populations [41,42,43].

Basically, it is composed of 26 items, divided into four domains: physical, psychological, social and environmental. The first two items (*How would you rate your quality of life* and *How satisfied are you with your health*) are not included in any domain, since they account for an overall score. The response to each item is chosen from a 5-point Likert scale [31,44]. The WHOQOL-BREF instrument has been applied in many studies recently, both in clinical and community settings [45,46]. However, there are very few reports on the response scale performance up to date.

The WHOQOL-BREF was originally designed to be a self-administered instrument. The subjects with high education were required to complete the instrument without any interference from the research assistant. For non-readers, the instrument was read loud. The application was carried out on an individual basis (to guarantee confidentiality and no interference) and the research assistants were told not to change phrasing or give any further explanation about items in order to guarantee comparability of the findings.

STATISTICAL ANALYSIS

Statistical analyses combined classic and modern approach (Item Response Theory). First, Confirmatory Factor Analysis (CFA) was carried out to assess whether the data collected would fit the proposed four-factor WHOQOL-BREF model. The CFA analyses were run in AMOS software [47], which uses graphical interface to determine model characteristics. Figure I exhibits the final model analysed. The goodness-of-fit of the model was verified. According to Kline (2005), fit statistics should always include the Chi-square statistic, a measure of the overall explained variance, a measure of fit that adjusts for parsimonious models, and an index of standardized residuals [48].

The Chi-square statistic reports the likelihood that the data would fit the *a priori* hypothesized model (the reason why it is also called the likelihood ratio chi-square). A significant Chi-square indicates a difference between the model and the data. Kline actually describes it as a measure of “badness-of-fit”, since the higher its value, the worse the model is. Even though it should be presented, Kline states that

there are some problems in relying solely on chi-square to assess model adequacy, since it is affected by the sample size and large correlations among variables [48].

The GFI and Tucker-Lewis parameters were also measured. The first indicates the overall explained variance. The second is a measure of the extent to which the models represent the data. For both, scores above .90 are considered superior indices.

The Root Mean Square Error of Approximation (RMSEA) is a parsimony-adjusted index which includes a built-in correction for model complexity. Thus, if two models with similar overall explanatory power are given, the simpler one would be favored [48]. A rule of thumb would be that scores below .05 indicate close approximate fit, values between .06 and .08 suggest reasonable error of approximation, and values above .10 indicate poor fit.

The IRT analysis was carried out as a second step. It consisted of verifying discrepancies in the psychometric performance of the WHOQOL-BREF items between the two subgroups (illiterate and graduate). Furthermore, changes in the response scale were analyzed to check whether they may represent improvement or not.

Data were examined for the Rasch model using the RUMM 2020 software [49]. Basically, the Rasch model assumes that the probability of a given subject endorsing an item is a function of the relative distance between the item location and the person location on a linear common scale [24]. In the case of a scale to measure depression, the probability that a person is endorsing an item is a logistic function of the difference between the subject's ability (level of depression) and the level of depression indicated by the item. The following equation illustrates this statement

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \theta_n - b_i$$

where \ln is the normal log, P is the probability that a person n endorses the item, θ is the person's level of depression and b is the level of depression expressed by the item.

Three overall fit statistics were first examined. An item-trait interaction was analysed by a Chi-Square test, which indicates the invariance property if the p-value is not significant (thus indicating similarity between expected and observed models). The standardized distributions of items and persons were examined, and, if both fit the model, values of zero for means and 1 for standard deviations are expected. The Person Separation Index (similar to Cronbach's Alpha in the classical psychometric theory) was also examined.

Furthermore, individual item and person statistics were analysed for residuals and Chi-Square statistics. Again, if an item fits the model, low residual (± 2.5) and non-significant Chi-Square statistics are expected. Bonferroni correction was applied to control for multiple tests effect [50].

Finally, items were examined for differential item functioning (DIF). The presence of DIF indicates that a subgroup (e.g., males or young adults) has a consistently different manner of responding to an item, despite having the same amount of the latent trait. Both uniform DIF (when the difference is constant through the whole range of the item curve) and non-uniform DIF (when the difference occurs only in a certain level of attribute) were checked.

RESULTS

DESCRIPTIVES

Table 1 shows the sample characteristics (n=1048). Health Status perception is based solely on the subjective perception of one's health, regardless of the objective condition. Subjects were invited to respond to the question "Would you consider yourself healthy or unhealthy?" and would be categorized according to the answer.

Subjects were selected only in industrialized cities in 20 countries worldwide. The non-reader sample was assessed in 8 different countries (Montevideo (Uruguay), Bath and Edinburgh (United Kingdom), Barcelona (Spain), Copenhagen (Denmark), Seattle (United States), Porto Alegre (Brazil), Izmir (Turkey) and Umea (Sweden)). Graduate subjects were interviewed in all 20 centres.

INSERT TABLE I HERE

CONFIRMATORY FACTOR ANALYSIS

Confirmatory Factor Analysis (CFA) was carried out with AMOS 6.0 software [47]. Two models were tested. First, a four uncorrelated factors solution was assessed. CFA results showed low model fit scores (Chi-Sq = 3596.43, df = 252, $p < 0.001$, CFI = .710, GFI = .765, TLI = .683, RMSEA = .113). A second model was tested, allowing for correlation among the four factors. The latter showed marked improvement when compared to the first one (Chi-Sq = 1670.84, df = 246, $p < .0001$, CFI = .877, GFI = .879, TLI = .862, RMSEA = .074). Fit indexes proved to be reasonable and close to the target cut-points.

Figure 1 illustrates the final adopted model. Note that items 1 and 2 are not included.

INSERT FIGURE 1 HERE

ITEM RESPONSE THEORY

Items were analyzed divided into the four domains. Item fit statistics are thus reported for each domain as well. Threshold problems were examined through the Threshold Map and Category Probability Curve (both are different ways of investigating response inconsistency in RUMM 2020).

Reversed Thresholds

Reversed thresholds are tested and informed by the Threshold Map. Then, the careful examination of the Categories Probability Curves indicates whether all the five response categories are used or whether one (or more) are not in fact being considered by the respondents.

Figure 2 illustrates the Threshold Map for the Physical Domain.

INSERT FIGURE 2 HERE

RUMM 2020 exhibits graphically the items with threshold problems (i.e., items 3, 4 and 10 in the physical domain for non-readers, but not for high educated ones). In addition, it is also possible to verify the response distribution and check potential differences between both subsamples. By checking the items free of reversed thresholds, it can be seen that the distribution for the graduate sample appears more homogeneous than the one for the illiterate sample. It is particularly marked for the items 10, 16 and 18 in the Figure 2.

The Psychological Domain has no reversed threshold for either sample. Some items (e.g., items 7, 11 and 26) show that response categories 2, 3 and 4 are too close in the non-reader sample (but are homogeneously distributed in the graduate group). Although not presenting reversed thresholds, this distribution pattern indicates that the three intermediate response categories do not function as separated entities (i.e., do not have the same separation power across the two subsamples).

The Social Domain showed one reversed threshold in the two subsamples (item 21 in the non-reader group and item 22 in the graduate group). Again, the items with no reversed thresholds have a more homogeneous distribution in the high educated sample.

The Environmental Domain shows a different pattern. Items 8 and 9 have reversed thresholds only in the high educated sample. The distribution of the response categories for the other six items presents a more homogeneous pattern for the non-reader group.

Category Probability Curves

Besides verifying reversed thresholds, the RUMM 2020 software offers a graphical representation of the category probability curve for each item. Beyond the simple notation for threshold disorders, this feature allows the examination of which categories are problematic and the distribution pattern across the ability span. The ideal distribution should cover the span and be homogeneously placed across it, in order to be informative. When threshold disorders occur, the graphical disposition helps to indicate what strategies could be applied to improve the response scale performance.

The examination of the 24 items in both non-reader and graduate samples shows marked differences. The category distributions are homogeneously placed in the highly educated subjects. However, non-readers exhibit a pattern of either collapsed or grouped categories. The graphical representation suggests that only three effective categories are considered, even though five response options are presented. Results indicate that the two anchor categories (i.e., options 1 and 5) are effective. The three intermediate options, though, seem to be merged into one. Figure 3 illustrates the category probability curves for the item 26 in its original format and after rescoring.

INSERT FIGURE 3 HERE

Threshold disorders determine model and item misfits, since they invert the expected pattern [25]. Collapsing categories could bring loss of information and, thus, decreased model usefulness [25,51]. Rescoring is one possible strategy to address

response category problems [26]. By doing so, two or more categories with non-discriminant thresholds are merged into one, which may correct threshold order and improve the categories distribution. Figure 3 (in the bottom) illustrates item 26 after being rescored from five into three categories. The effect that rescoreing has in promoting homogeneity in categories distribution is observable.

Item-Trait Interaction

Item-trait interaction statistics indicate whether the observed data fit the expected model or not. Fitting the expected model would mean that subjects with more ability in fact have higher probability of affirming items than the others with low ability. In addition, it would indicate that more difficult items require high ability to be endorsed [24,52] according to the Guttman pattern [53]. As mentioned above, the closer to non-significance the chi-square value is, the more logical the model is.

Since category probability curves showed disorders for the non-reader subsample, further analysis were carried out for both the 5-point original response scale and the 3-point scale derived from the rescoreing process. Table 2 presents the item-trait chi-square, item fit and person fit. In addition, Person Separation Indexes are described, indicating the amount of information provided by the model.

INSERT TABLE 2 HERE

Results show that the model Chi-square and probability values shifted remarkably towards the expected model in the non-reader sample (with exception of

the Social Domain), when comparing the original and the rescored versions. As rescoring determined an approximation to the expected model, it indicates that data are more likely to follow a logic probabilistic distribution when rescored.

On the other hand, Chi-square and probability values showed no alterations when items were rescored in the graduate sample. In fact, values for the Social and Environmental Domains presented a shift towards non-significance, indicating worse performance due to information loss.

Individual Item Fit Statistics

Differential Item functioning (DIF) was first examined across the two subsamples (n=1048). Items were analyzed by domains. Out of the 24 items, 11 items showed uniform DIF and 3 items presented non-uniform DIF according to educational level. Five items were located in the Physical domain, five were in the Psychological domain and two were in the Social domain. Interestingly, the Environmental domain (not widely affected by DIF) showed the best threshold performance for the non-reader group. Regarding the gender effect, only 3 out of 24 items showed DIF.

Individual item fit statistics were analyzed through chi-square results and residuals. The former indicates the level of misfit between the item data and the expected item curve. Thus, the higher the value is, the worse performance of the item. Residuals, on the other hand, are related to item deviation and, thus, should ideally be close to zero. Values between ± 2.5 suggest adequate fit statistics.

Intra-group DIF effects were also examined for gender and subjective perception of health status. Again, it is expected that an item with good psychometric

performance show no differential functioning regardless of any other individual characteristic than the latent trait of interest. Table 3 describes the item fit statistics.

INSERT TABLE 3 HERE

Results from item analyses illustrate that, in general, problematic chi-square and residuals values are shown in the non-reader group, particularly in the Physical and Psychological domains. DIF is also related to the Physical domain in the illiterate sample. Although not marked, there is a tendency for the misfit to locate in non-readers, when compared to the graduate sample.

After rescoring, item fits were reassessed to check potential improvements or impairments. Threshold disorders did not remain after rescoring. In fact, chi-square values presented a marked decrease for the non-reader sample. Only two items showed significant misfit after rescoring (and 8 in the original 5-category response scale). For the set of 24 items, there was a reduction in chi-square values, indicating that rescoring improved the item performance and determined an approximation between the observed and the expected item curves. DIF also showed reduction for the illiterate group.

In contrast to the findings described for the illiterate sample, chi-square values tended to be higher in the graduate sample. Some items with good previous performance became misfitting (for example, items 10, 18, 20 and 22). After rescoring, ten items showed significant high chi-square values (and only 6 in the

original 5-point response scale). DIF was not altered by rescoring in this subsample. Table 4 describes item fit statistics and DIF results after rescoring.

In general, rescoring produced improvements in item fit statistics for the non-reader group, but not for the graduate group. In the latter, merging response categories led to decreased model fit and impairment of item performance.

INSERT TABLE 4 HERE

DISCUSSION

This study aimed to test whether illiteracy has an effect on the performance of a multiple-point Likert scale. It was hypothesized that, due to illiteracy's comprehensive repercussions on the human brain and cognitive ability, the multiple-point response scale might not be suitable for non-readers. It was also predicted that the multiple-point scale could in reality determine loss of information and consequent impaired psychometric performance for non-readers. To investigate this research question, two subsamples (of non-readers and graduates) were required to complete the WHOQOL-BREF instrument. Statistical analyses followed the Rasch model, a modern and powerful tool for examining scale and item properties [54].

The assessment of literacy skills represents a crucial point in the present study. No formal assessment was carried out (for example, the application of a specific instrument), and objective indexes (such as years of education) do not reflect literacy ability adequately [3]. Thus, authors have decided only to include in the study subjects with high probability of representing non-readers and readers. By this means,

we attempted to avoid the possible bias of considering lack of formal schooling as an indicative of illiteracy, as well as considering few years of formal schooling as an indicative of ability to read.

The study sample was comprised of older adults selected from a large international study on quality of life [35]. Elderly have had fewer opportunities to receive formal education and, thus, represent a significant portion of the illiterate population [4,7].

Results indicated that there is a remarkable difference in the psychometric performance of the same instrument across the two subgroups. While fit indexes (both model and item) proved to be closer to the ideal for the graduate group, they fall away from the limits for the non-reader group. Item biases (DIF) were also more frequent in the latter than in the former. The effects of illiteracy seem to be more noticeable in Physical, Psychological and Social domains, but less in the Environmental domain. Findings in Table 3 show that item fit statistics for the Environmental domain are not affected by educational level, and the same satisfactory pattern can be checked in the threshold distribution. In general, items included in this domain perform well in both groups.

A close inspection of these items suggests that they are heavily related to objective and real themes (such as transport, conditions of living place and access to health services). On the other hand, the Physical, Psychological and Social domains include items approaching more abstract constructs (such as self satisfaction, feeling life is meaningful, the extent to which one enjoys life, satisfaction with ability to perform daily activities). This finding is in line with the ones reported by Luria [55], Scrinber and Cole [56] and Manly et al [7], who indicated that illiteracy has a stronger effect on more subjective and abstract tasks than in objective and concrete ones.

Recoding the response scales from five to three points was then tested to check the effect it has on psychometric performance. The option of recoding into three categories was drawn from the category probability curves, which indicated that three response options in fact covered the ability span adequately for the non-reader group. It was expected that model and item fit statistics would improve if the response scale was not adequate. In contrast to this, fit indexes were predicted to worsen if the original response scale was more suitable than the new recoded one [25]. Results showed that the three categories format produced improved psychometric performance for the non-reader group (both in model and item fit indexes) in all domains apart from the Environmental one. On the other hand, collapsing categories caused impairments for the graduate sample.

These results confirm the hypothesis that a multiple-category response scale is not suitable for non-readers. Furthermore, they suggest that a simpler response scale should be adopted to achieve a more reliable and satisfactory psychometric performance in this population. In fact, this effect seems to be stronger when more abstract, complex, and subjective constructs are involved.

A potential limitation of the present study is the cohort effect. It is possible that results are affected by some age-specific factors, since the sample is composed exclusively of older adults. Further investigations including younger subjects could address this issue. The lack of specific assessment of literacy determined that comparisons including groups with intermediate levels of literacy abilities could not be run. Since this group represents the majority of adults [3], this analysis could have even larger policy and clinical implications.

In summary, our findings suggest that a five-point Likert response scale is not suitable for non-readers. In addition, they show that illiterate subjects do not interpret

a five-point scale as a true five category scale, but instead exhibit a response pattern that indicate that only three categories are in fact considered. In conclusion, instruments and scales should be adequately examined and refined in order to assess non-readers properly.

Acknowledgements: Authors acknowledge the CAPES for funding the first author's PhD scholarship.

REFERENCES

1. Matute E, Leal F, Zarabozo D et al. Does literacy have an effect on stick construction tasks? *Journal of the International Neuropsychological Society*, 2000, 6:668-672.
2. Payne G. Re-counting 'illiteracy': literacy skills in the sociology of social inequality. *The British Journal of Sociology*, 2006, 57(2):219-240.
3. Kutner M, Greenberg E, Jin Y et al. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. National Center for Education Statistics, US Department of Education. Washington DC. US Government Print Office; 2006.
4. UNESCO Department of Statistics. Education for All, Global Monitoring Report. [Internet]. Available <http://unesco.unesco.org/images/0014/001416/141639e.pdf> [assessed on April 13th 2007]; 2006.
5. Ostrosky-Solis F, Ardila A, Rosselli M et al. Neuropsychological Test Performance in Illiterate Subjects. *Archives of Clinical Neuropsychology*, 1998, 13(7):645-660.
6. Castro-Caldas A, Petersson KM, Reis A et al. The illiterate brain: learning to read and write during childhood influences the functional organization of adult brain. *Brain*, 1998, 121:1053-1063.
7. Manly JJ, Jacobs D, Sano M et al. Effect of illiteracy on neuropsychological test performance in nondemented, education-matched elders. *Journal of the International Neuropsychological Society*, 1999, 5:191-202.
8. Dellatolas G, Braga LW, Souza LN et al. Cognitive consequences of early phase of literacy. *Journal of the International Neuropsychological Society*, 2002, 9:771-782.
9. Petersson KM, Reis A and Ingvar M. Cognitive processing in literate and illiterate subjects: A review of some recent behavioural and functional neuroimaging data. *Scandinavian Journal of Psychology*, 2001, 42:251-267.
10. Petersson KM, Reis A, Askelof S et al. Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of Cognitive Neuroscience*, 2000, 12:364-382.

11. Reis A and Castro-Caldas A. Illiteracy: a cause for biased cognitive development. *Journal of the International Neuropsychological Society*, 1997, Sep, 3(5):444-450.
12. Byrd DA, Jacobs DM, Hilton HJ et al. Sources of errors on visuoperceptual tasks: Role of education, literacy, and search strategy. *Brain and Cognition*, 2005, 58:251-257.
13. Ardilla A. Cognitive evaluation in illiterate persons. *Revista de Neurologia*, 2000, Mar 1-15; 30(5):465-8.
14. Bramao I, Mendonca A, Faisca L et al. The impact of reading and writing skills on a visuo-motor integration task: A comparison between illiterate and literate subjects. *Journal of the International Neuropsychological Society*, 2007, 13:359-364.
15. Sentell T and Halpin HA. Importance of adult literacy in understanding health disparities. *Journal of General Internal Medicine*, 2006, 21:862-866.
16. Rosselli M and Adrila A. The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, 2003, 52:326-333.
17. Baker DW, Parker RM, Williams MV et al . The relationship of patient reading ability to self-reported health and use of health services. *American Journal of Public Health*, 1997, Jun, 87(6):1027-30.
18. Lindau ST, Tomori C, Lyons T et al. The association of health literacy with cervical cancer prevention knowledge and health behaviours in a multiethnic cohort of women. *American Journal of Obstetrics and Gynaecology*, 2002, 186:938-943.
19. Manly JJ, Schupf N, Tang M et al. Cognitive Decline and Literacy Among Ethnically Diverse Elderly. *Journal Geriatric Psychiatry and Neurology*, 2005, 18:213-217.
20. Buchbinder R, Hall S and Youd JM. Functional health literacy of patients with rheumatoid arthritis attending a community-based rheumatology practice. *Journal of Rheumatology*, 2006, May, 33(5):879-886.
21. Sentell T and Ratcliff-Baird B. Literacy and Comprehension of Beck Depression Inventory response alternatives. *Community Mental Health Journal*, 2003, 39(4):323-331.
22. Deloche G, Souza L, Braga LW et al. A calculation and number processing battery for clinical application in illiterates and semi-literates. *Cortex*, 1999, 35:503-521.

23. Mayeaux EJ Jr, Davis TC, Jackson RH et al. Literacy and self-reported educational levels in relation to Mini-mental State Examination scores. *Family Medicine*, 1995, Nov-Dec, 27(10):658-662.
24. Pallant J and Tenant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 2007, 46:1-18.
25. Andrich D. A rating formulation for ordered response categories. *Psychometrika*, 1978, 43(4):561-573.
26. Pallant J, Miller R and Tenant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry*, 2006, 6:28-38.
27. Garamendi E, Pesudovs K, Stevens M et al . The refractive status and vision profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Research*, 2006, 46:1375-1383.
28. McHorney C, Haley S and Ware J. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, 1997, 50(4):451-461.
29. Haley SM, McHorney CA, Ware J. Evaluation of the MOS SF-36 Functioning Scale: I. Unidimensionality and reproducibility of the Rasch scale. *Journal of Clinical Epidemiology*, 1994, 47:671-684.
30. Stucki G, Daltroy L, Katz JN et al. Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not be equal to the sum of the parts. *Journal of Clinical Epidemiology*, 1996, 49:711-717.
31. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment . *Psychological Medicine*; 1998, May, 28(3):551-8.
32. Hartley SL and MacLean Jr WE. A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research*, 2006, 50(11):813-827.
33. Verri A, Cummins RA, Petito F et al. An Italian-Australian comparison of quality of life among people with intellectual disability living in the community. *Journal of Intellectual Disability Research*, 1999, 43(6):513-522.

34. Williams SA and Swanson MS . The effect of reading ability and response formats on patients' abilities to respond to a patient satisfaction scale. *The Journal of Continuing Education in Nursing*, 2001, 32(2):60-67.
35. Power MJ, Quinn K, Schmidt S et al. Development of the WHOQOL-OLD module. *Quality of Life Research*, 2005, Dec;14(10):2197-214.
36. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scandinavian Journal of Rheumatology*, 1995, 24(2):61-3.
37. Bullinger M, Power M, Aaronson NK, Cella DF, Anderson RT. Creating and evaluating cross-cultural instruments. In: B. Spilker, Ed. *Quality of life and pharmacoeconomics in clinical trials*. Hagerstown, MD: Lippincott-Raven; 1996: 659-668.
38. Skevington SM, Lofty M, Connel KA et al. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research*, 2004a, Mar, 13(2):299-310.
39. Skevington SM, Sartorius N and Amir M. Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL instrument. *Social Psychiatry and Psychiatric Epidemiology*, 2004b, Jan, 39(1):1-8.
40. Saxena S, Carlson D, Billington R et al. The WHO quality of life assessment (WHOQOL-Bref): the importance of its items for cross-cultural research. *Quality of Life Research*, 2001, 10(8):711-721.
41. Chachamovich E, Trentini C and Fleck MP. Assessment of the psychometric performance of the WHOQOL-BREF instrument in a sample of Brazilian older adults. *International Psychogeriatrics*, 2006, Jul, 27:1-12.
42. Naumann VJ and Byrne GJ. WHOQOL-BREF as a measure of quality of life in older patients with depression. *International Psychogeriatrics*, 2004, Jun, 16(2):159-73.
43. Hwang HF, Liang WM, Chiu YN et al. Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age Ageing*, 2003, Nov, 32(6):593-600.
44. Likert R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932, 140:1-55.

45. Lima AFB and Fleck MP. Subsyndromal depression: An impact on quality of life? *Journal of Affective Disorders*, 2006, doi:10.1016/j.jad.2006.10.010
46. Izutsu T, Tsutsumi A, Islam AM et al. Mental health, quality of life, and nutritional status of adolescents in Dhaka, Bangladesh: comparison between an urban slum and a non-slum area. *Social Science and Medicine*, 2006, Sep, 63(6):1477-88.
47. Arbuckle JA. Amos 6.0 User's Guide. Amos Development Corporation. Spring House, PA, USA, 2005.
48. Kline RB. Principles and practice of structural equation modelling. New York: Guilford Pres, 2005.
49. Andrich D, Lyne A, Sheridan B and Luo G. RUMM 2020. Perth: RUMM Laboratory, 2003.
50. Bland JM and Altman DG. Multiple significant tests: The Bonferroni Method. *British Medical Journal*, 1995, 310:170.
51. Andersen E . Sufficient statistics and latent trait models. *Psychometrika*, 1977, 42(1):69-81.
52. Andrich D. Rasch Models for measurement. London: Sage University Paper; 1988.
53. Gutman L The basis of scalogram analysis. In: Stouffer Sa ,Ed. Measurement and prediction. Princeton, NJ: Princeton University Press; 1950.
54. Stout W. Psychometrics: from practice to theory and back. *Psychometrika*, 2002, Dec, 67(4):485-518.
55. Luria AR. Cognitive development, its cultural and social foundations. Cambridge, MA: Harvard University Press; 1976.
56. Scribner S and Cole M. The psychology of literacy. Cambridge, MA: Harvard University Press; 1981.

Table 1 - Sample characteristics (n=1048)

	Illiterate (N = 141)	Graduate (N = 907)
Age		
Mean and SD	73.29 (7.14)	70.37 (7.60)
Gender		
Male	51 (36.2)	479 (52.8)
Female	90 (63.8)	428 (47.2)
Marital Status		
Single	3 (2.1)	55 (6.1)
Married	53 (37.6)	419 (68.6)
Separated	6 (4.3)	82 (9.1)
Widowed	79 (56.0)	146 (16.2)
Health Status Perception		
Healthy	66 (46.8)	714 (79.5)
Unhealthy	75 (53.2)	184 (20.5)

Table 2 – Item-trait interaction statistics for illiterate and graduate samples

	Illiterate 5 points	Graduate 5 points	Illiterate 3 points	Graduate 3 points
Physical				
Chi-square	202.62	176.45	11.34	175.05
Probability	0.00000	0.00000	0.658	0.00000
PSI	0.713	0.839	0.359	0.688
Item fit (M,SD)	0.981 (2.16)	0.420 (4.53)	-1.17 (0.68)	-3.41 (2.32)
Person Fit (M, SD)	-0.361 (1.73)	-0.340 (1.19)	-1.34 (1.85)	-0.86 (1.37)
Psychological				
Chi-square	149.18	75.59	24.83	78.73
Probability	0.00000	0.0012	0.015	0.00000
PSI	0.704	0.813	0.280	0.448
Item fit (M,SD)	0.617 (2.64)	-0.043 (2.82)	-2.12 (1.13)	-4.75 (1.67)
Person Fit (M, SD)	-0.330 (1.43)	-0.550 (1.55)	-1.38 (1.43)	-1.21 (1.48)
Social				
Chi-square	17.01	41.92	21.90	143.94
Probability	0.149	0.0003	0.001	0.00000
PSI	0.447	0.679	0.371	0.092
Item fit (M,SD)	0.680 (1.14)	-0.229 (1.04)	-1.09 (0.08)	-5.31 (0.17)
Person Fit (M, SD)	-0.489 (1.19)	-0.576 (1.07)	-0.57 (0.66)	-0.61 (0.56)
Environmental				
Chi-square	53.54	68.27	27.15	106.81
Probability	0.074	0.035	0.297	0.00002
PSI	0.869	0.876	0.541	0.721
Item fit (M,SD)	0.282 (1.44)	0.394 (1.86)	-2.47 (0.67)	-3.80 (0.68)
Person Fit (M, SD)	-0.378 (1.34)	-0.420 (1.41)	-1.49 (1.56)	-1.05 (1.65)

Table 3 – Description of item fit statistics in the two subsamples (5-point response scale)

Items	Illiterate (n=141)					Graduate (n=907)				
	χ^2 fit	Residual	Rev Thres	DIF Gender*	DIF Health*	χ^2 fit	Residual	Rev Thres	DIF Gender*	DIF Health*
Physical										
3	55.65	3.40	✓	0.470	0.000	5.749	2.11		0.611	0.726
4	54.18	3.75	✓	0.940	0.000	9.966	3.89		0.743	0.411
10	15.33	-0.37		0.825	0.003	7.304	-1.99		0.039	0.806
15	22.86	-1.45	✓	0.810	0.007	17.17	-1.33		0.000	0.002
16	4.93	2.48		0.765	0.069	90.79	8.07		0.007	0.000
17	25.71	-0.24		0.878	0.000	36.15	-5.25		0.191	0.003
18	23.90	-0.71		0.539	0.000	9.322	-2.56		0.060	0.669
Psychological										
5	20.01	-1.52		0.252	0.002	19.11	-3.05		0.967	0.636
6	12.37	-0.77		0.003	0.705	13.66	-3.32		0.007	0.554
7	16.00	1.33		0.808	0.007	10.17	1.93		0.592	0.002
11	4.32	-0.17		0.046	0.067	9.23	1.77		0.232	0.835
19	13.91	-0.80		0.248	0.014	6.72	-0.97		0.134	0.190
26	82.55	5.64		0.065	0.005	16.67	3.39		0.000	0.634
Social										
20	3.92	-0.14		0.176	0.023	8.51	-1.39		0.878	0.033
21	7.62	1.99	✓	0.362	0.520	21.78	0.63		0.003	0.838
22	5.46	0.19		0.607	0.266	11.61	0.07	✓	0.000	0.048
Environmental										
8	3.93	0.92		0.780	0.028	8.37	-1.71	✓	0.007	0.612
9	7.35	-1.35		0.853	0.022	6.56	-1.02	✓	0.343	0.601
12	7.45	-0.68		0.066	0.821	4.99	1.34		0.017	0.395
13	8.92	1.87		0.356	0.119	4.39	-0.79		0.968	0.476
14	15.41	2.69		0.587	0.022	25.24	3.37		0.032	0.373
23	4.47	-0.81		0.225	0.219	12.44	-1.37		0.006	0.915
24	3.03	-0.72		0.416	0.464	0.34	1.30		0.615	0.236
25	2.94	0.35		0.814	0.180	5.91	2.03		0.251	0.528

* p values, bolded values indicate $p < 0.05$ (Bonferroni corrected)

Table 4 - Description of item fit statistics in the two subsamples (rescored into 3-point response scale)

Items	Illiterate 3 Point Scale					Graduate 3 Point Scale				
	χ^2 fit	Residual	Rev Thres	DIF Gender	DIF Health	χ^2 fit	Residual	Rev Thres	DIF Gender	DIF Health
Physical										
3	4.79	-0.27		0.632	0.000	16.34	-1.62		0.506	0.312
4	3.79	-0.59		0.474	0.000	9.59	-1.38		0.273	0.554
10	3.28	-0.95		0.843	0.001	19.79	-4.33		0.676	0.628
15	2.84	-1.90		0.791	0.011	30.82	-5.31		0.149	0.003
16	5.20	-0.96		0.470	0.270	36.65	-0.34		0.440	0.000
17	7.68	-1.37		0.449	0.382	48.23	-6.67		0.069	0.106
18	4.19	-2.19		0.739	0.081	25.22	-4.19		0.207	0.056
Psychological										
5	4.87	-3.03		0.318	0.480	12.12	-5.70		0.967	0.636
6	2.50	-2.97		0.392	0.311	30.10	-7.01		0.007	0.554
7	3.90	-1.99		0.474	0.621	13.29	-2.58		0.592	0.002
11	1.94	-2.92		0.125	0.049	3.49	-5.09		0.232	0.835
19	3.86	-1.73		0.694	0.000	12.24	-5.11		0.134	0.190
26	12.73	-0.10				12.31	-3.00		0.000	0.634
Social										
20	3.96	-1.56		0.112	0.198	62.19	-5.19		0.124	0.763
21	13.09	-1.12		0.480	0.002	23.10	-5.24		0.067	0.004
22	10.18	-1.00		0.224	0.204	58.65	-5.51		0.134	0.001
Environmental										
8	3.49	-2.18		0.508	0.634	6.65	-4.08		0.230	0.172
9	2.19	-2.87		0.296	0.706	6.61	-4.39		0.002	0.205
12	10.28	-2.80		0.232	0.567	17.20	-4.00		0.010	0.181
13	1.75	-1.88		0.737	0.432	11.54	-2.76		0.130	0.529
14	6.69	-1.26		0.494	0.019	13.57	-2.79		0.730	0.896
23	5.02	-3.04		0.463	0.048	26.92	-4.54		0.010	0.570
24	0.61	-3.31		0.147	0.985	13.72	-3.66		0.710	0.070
25	3.00	-2.38		0.538	0.448	10.56	-4.18		0.049	0.817

* p values, bolded values indicate $p < 0.05$ (Bonferroni corrected)

Figure 1 – CFA of the final model

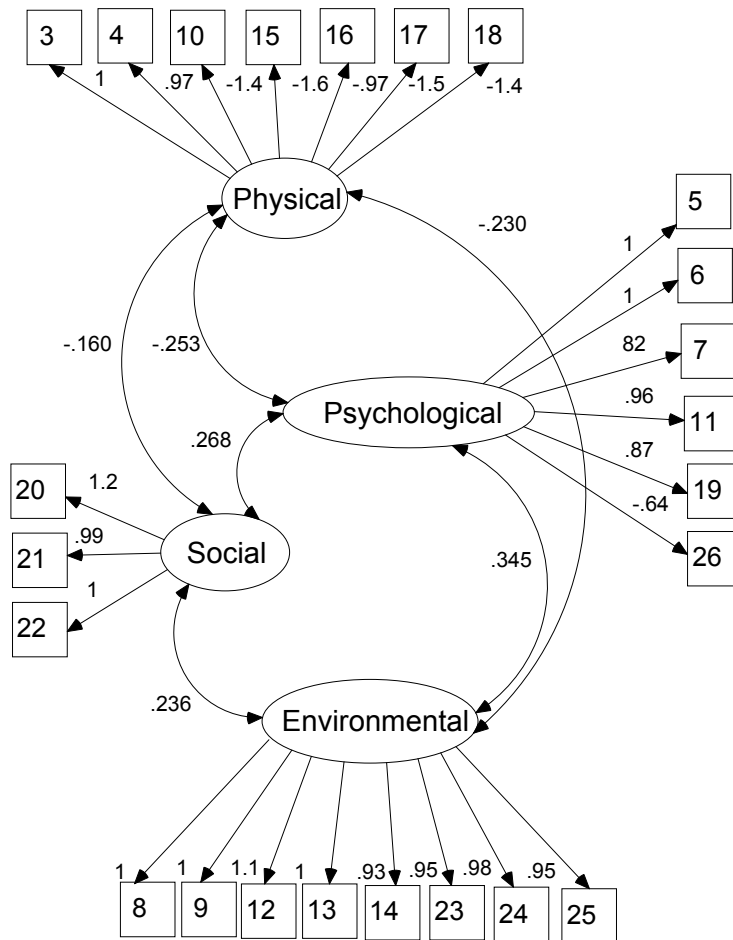


Figure 2 – Threshold Map for the Physical Domain (illiterate subsample on the top and graduate subsample at the bottom)

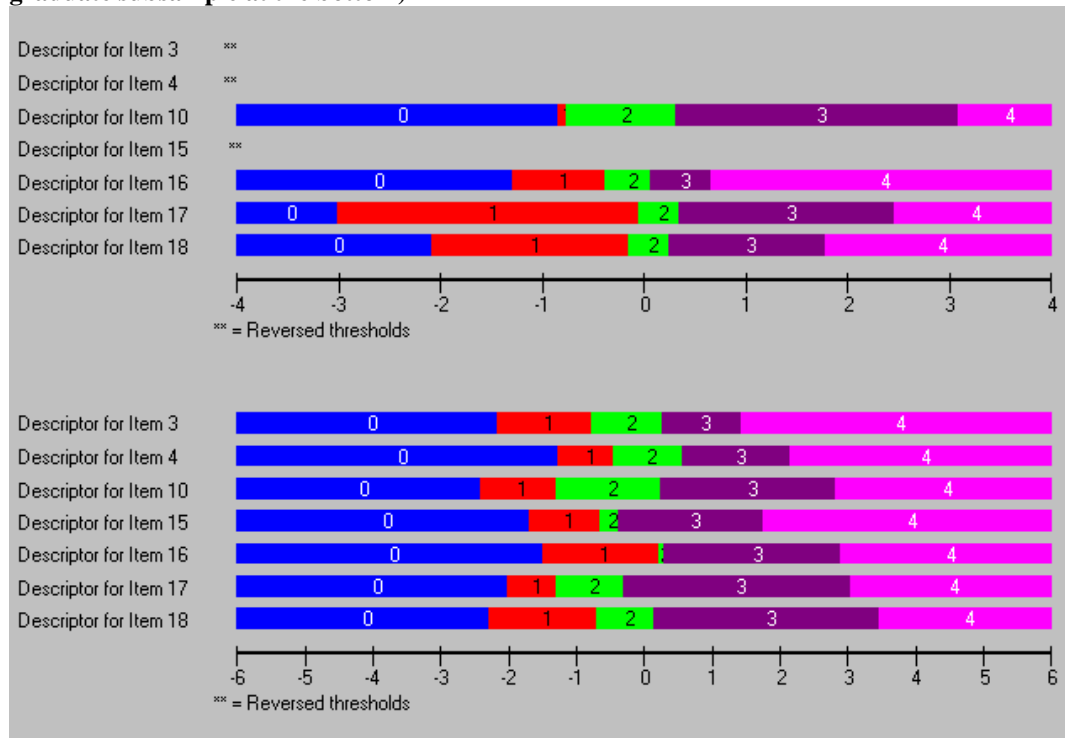
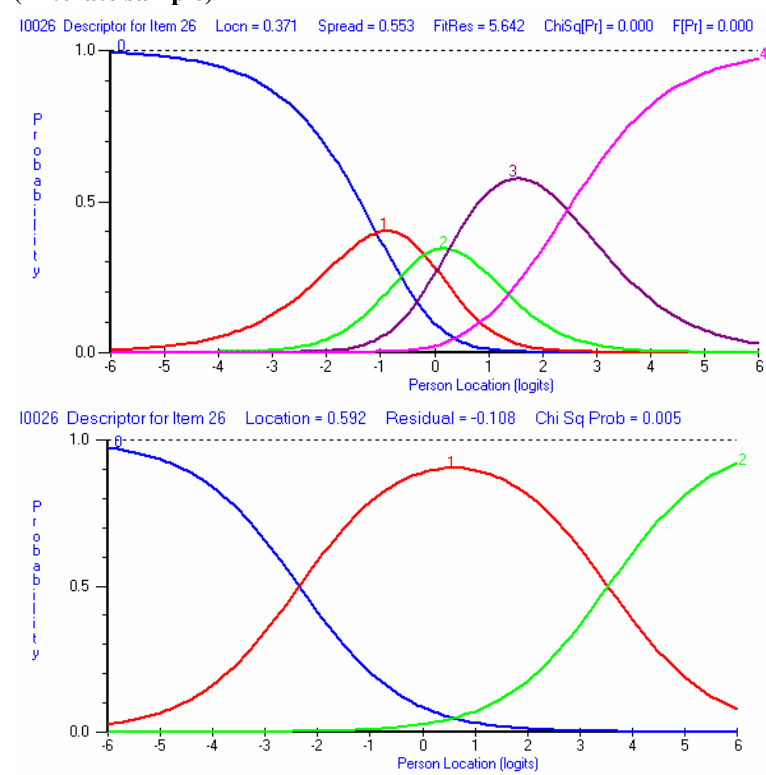


Figure 3 – Category Probability Curves for item 26 in its original 5-point scale and rescored (illiterate sample)





New Investigator/Student Award

presented to

Eduardo Chachamovich, PhD student

for an outstanding paper entitled "What impact does illiteracy have on a multi-point Likert Scale?"
presented by the

International Society for Quality of Life Research

October 13, 2007

Madeleine King, PhD, ISOQOL President

ARTIGO 4

The cultural influence in measuring depression of older adults: Is GDS-15 a suitable instrument in Brazil?

Submetido para publicacao à revista International Psychogeriatrics

The cultural influence in measuring depression of older adults: Is GDS-15 a suitable instrument in Brazil?

Prof. Eduardo Chachamovich^{1,2}, Prof. Marcelo P. Fleck², Prof. Mick Power¹

1- Clinical and Health Psychology, University of Edinburgh
Medical School, Teviot Place, EDINBURGH EH8 9AG.

2- Department of Psychiatry and Legal Medicine, Universidade Federal do Rio Grande do Sul and Hospital de Clinicas de Porto Alegre
Rua Ramiro Barcelos, 2350, Porto Alegre, Brazil. CEP 90035-903

Corresponding Author: Eduardo Chachamovich

Address: Rua Florencio Ygartua 391/308

Porto Alegre – RS

CEP 90430-010

Brazil

Email: echacha.ez@terra.com.br

Phone: +55 51 91230015

Fax: +55 51 32235409

ABSTRACT

Introduction: Depressive symptoms are the most prevalent mental health condition in older adults. Since it cannot be measured directly, the use of instruments is mandatory. The 15-item GDS is one of the most widely used scales to measure depression in elderly. It is recognized that the cultural context is a major determinant of the instrument's psychometric performance. The present study aims to explore whether the 15-item GDS is a suitable instrument in a Brazilian sample. In addition, it explores the potential improvement in the psychometric performance by item refinement.

Methods: 424 elderly adults selected through convenience sampling completed the 15-item GDS. Data were analyzed by the Rasch Measurement Model. The Rasch analysis is a powerful modern approach to explore psychometric performance of instruments in health sciences. It examines both the scale and the individual item performance in depth.

Results: The 15-item GDS proved not to be suitable in a Brazilian sample. Item misfit and DIF were responsible for considerable misperformance. Scale reduction led to a 10-item structure. This refined format presented adequate psychometric performance and no DIF.

Conclusion: The present study offers an alternative and more adequate version of the GDS to be applied in Brazilian subjects. Further investigations are needed to develop a set of cultural-invariant items, which could then be applied in transcultural investigations free of bias.

KEYWORDS: measurement, scale, elderly, Rasch, psychometrics

RUNNING TITLE: Is GDS-15 a suitable instrument in Brazil?

INTRODUCTION

Depressive symptoms are the most prevalent mental health condition in older adults (Osborn et al, 2003). Depression is associated with mortality, morbidity (Street et al, 2007), functioning impairment and reduced well-being (Beekman et al, 2002). In later life, depression has been also linked to a decrease in quality of life (Blazer et al, 2005).

Since depression cannot be measured directly, the use of questionnaires is mandatory in order to assess the severity of the symptomatology. Instruments frequently used in clinical practice often have diagnostic cut-points and, thus, play an important role in the clinicians' armamentarium. The Geriatric Depression Scale (Yesavage et al, 1983) was one such instrument developed to measure depression in older adults. At the time it was developed, existing scales were not specific to this population, which caused problems in applying them in the elderly population (Yesavage et al, 1983).

The first version of the Geriatric Depression Scale (GDS) consisted of 30 dichotomous items and has been widely used in investigations in different conditions, such as nursing environments (Jongenelis et al, 2005), Parkinson disease (McDonald et al, 2006), dementia (Chahine et al, 2006), and suicide (Heisel et al, 2005).

One continual criticism about the original GDS-30 scale has concerned its length, which may be burdensome to some older adults (Jongenelis et al, 2005, Almeida et al, 1999; de Craen et al, 2003). Consequently, a brief 15-item version was developed to address this issue (Sheik and Yesavage, 1986). This abbreviated version has been recommended for use within the over-75 years old health check by the Royal college of General Practitioners (D'Ath et al, 1994) and is now one of the most widely used depression scales in older populations (Almeida et al, 1999; Friedman et al, 2005). Comprising 15 yes/no questions derived from the original 30-item version, it has been used in large epidemiological surveys (Osborn et al, 2002; de Craen et al, 2003), as well as in several distinct clinical conditions. In addition, the GDS-15 has been frequently used as a tool for diagnostic procedures, computing a total score by adding positive responses, and several papers have consequently explored diagnostic

cut-points (Osborn et al, 2002; D'Ath et al, 1994; de Craen et al, 2003; Korner et al, 2007).

Although the GDS-15 has been widely applied in a series of research contexts, and investigated its diagnostic properties, few studies have explored its factorial composition and its psychometric performance using classic psychometric approaches (Chau et al, 2006; Friedman et al, 2005; Incalzi et al, 2003; Malakouti et al, 2006). Briefly, the findings indicate good internal consistency and the presence of two or three-factors extracted by principal components analysis.

It is recognized that the cultural context plays an important role in determining the performance of psychological instruments (Saxena et al, 2001). The adaptation and validation processes may lead to inadequacy of original scales and questionnaires in some cases. In addition, the cultural particularities may in fact demand adjustments to guarantee that the original instrument is suitable and performs well in a determined cultural context (McKenna et al, 2007). Not only the translation and back-translation processes have to be carefully run, but also the semantic equivalence should be guaranteed between the original version and the translated one (Schmidt and Bullinger, 2003). No investigation was carried out in Latin American countries to check whether the 15-item GDS structure remains stable or not.

The aim of this current study is to assess the psychometric performance of the GDS-15 in a sample of Brazilian older adults through Rasch analysis. The analysis focuses upon whether the Brazilian version of the GDS-15 functions as a unidimensional Rasch scale, and the potential bias of items by age, gender and educational level. It objectives ultimately to verify if the original set of items retains its psychometric properties in a sample of Brazilian subjects.

METHODS

Participants

A total of 424 subjects were interviewed, as part of a larger study developed by the WHOQOL Group (Power et al, 2005; Chachamovich et al, 2007). Recruitment was carried out through opportunistic sampling.

Procedure

Face-to-face interviews either in individual or group settings were carried out with adults aged 60 or more. Subjects were invited to complete a socio-demographic form and the GDS-15. Recruitment was done in nursing homes, older adults community groups and hospitals (both outpatients and inpatients units).

Rasch analysis

Data were tested against Rasch model expectations using the RUMM 2020 software (Andrich et al, 2003). The Rasch model can be seen as a template which puts into operation the axioms for additive conjoint measurement (Pallant et al, 2006). This theory presents a set of methods to determine whether a variable has an additive structure and consequently provides measurement on an interval scale.

Basically, the Rasch model assumes that the probability of a given subject endorsing an item is a function of the relative distance between the item location and the person location on a linear common scale (Pallant and Tenant, 2007). In the case of a scale to measure depression, the probability that a person is endorsing an item is a logistic function of the difference between the subject's level of depression and the level of depression expressed by the item. This can be expressed in the logit form:

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \theta_n - b_i$$

where \ln is the normal log, P is the probability of a person n to endorse the item, θ is the person's level of depression and b is the level of depression expressed by the item. If the data fit the Rasch model, then both the person's estimate of depression and items level of difficulty will be placed on the same metric scale (log-units scale or

logits), which permit a linear transformation of the raw ordinal scale; thus, when the data fit the model, and the assumptions of local independence are met, the scale is then suitable for both use in both its ordinal form as well as, through the transformation, for valid parametric approaches (Pallant et al, 2006). Only if the data fit the Rasch model (i.e., if the scale is intervalar), the additivity property of the items is then assured. In other words, only if the scale is truly intervalar the items can be summed free of bias (Bond and Fox, 2007).

To check whether or not the data meet the model expectations, three overall fit statistics are examined. An item-trait interaction was analysed using a Chi-Square test, which indicates the invariance property across different levels of the trait (thus indicating similarity between expected and observed models). Standardized distributions of items and persons were also examined, and, where data meet model expectations values of zero for means and 1 for standard deviations are expected.

Furthermore, individual item and person statistics are analysed for residuals and Chi-Square statistics. Again, if a determined item fits the model, low residual (± 2.5) and non-significant Chi-Square statistics are expected. Bonferroni correction was applied to control for multiple tests effects.

An estimate of internal consistency was also obtained through the Person Separation Index (PSI), which is comparable to the Cronbach's Alpha coefficient.

It is also possible to test that the scale is invariant across groups within the Rasch analysis framework (Tenant et al 2004). This comes from assessing the Differential Item Functioning (DIF) of items by group. The presence of DIF indicates that a subgroup (e.g., males) has a consistently different way of responding to an item (than females), despite having the same amount of the latent trait (depression). Both uniform DIF (when the difference is constant through the whole range of the item curve) and non-uniform DIF (when the difference occurs only at a certain level of attribute) were checked. Items were thus examined for DIF, since it represents a serious performance problem.

The assumption of local independence is tested by an examination of the person-item residual correlation. Correlation values higher than +0.3 indicate lack of independence. Unidimensionality was assessed by the principal component analysis of the residuals. The first PCA factor was divided into two subsets (one with items with the most positive loadings and the other including items with the most negative loadings). Person estimates were then obtained from these two subsets of items and an independent t-test was used to detect significant differences, which would indicate lack of unidimensionality.

Linacre (1994) states that the ideal sample size varies according to the scale targeting. For a well targeted scale (40-60% endorsement rates on a dichotomous items), a sample size of 108 would have a 99% confidence of person estimation of ± 0.5 logits. For not well-targeted scales, a satisfactory sample size would be 243 (Linacre, 1994).

RESULTS

The mean age of subjects was 72.3 years (SD 8.9) and 64.2 % were female (Table 1).

INSERT TABLE 1 ABOUT HERE

The overall fit of the GDS-15 to the Rasch model, as assessed by the item-trait interaction total chi-square, showed a highly significant result ($\chi^2 = 215.96$, $df=90$, $p<0.0001$), indicating misfit between the data and the model expectations. The item fit residual mean was -0.695 with a SD of 2.290 and the person's residual mean was -0.169 with a SD of 0.833. Thus the item fit appears poor, while the persons' statistics indicate no serious misfit. The Person Separation Index was 0.785, suggesting good person separation reliability for group application.

Given the apparent misfit, a more detailed examination of item fit was undertaken (Table II). Several items were found to demonstrate either significantly high residuals, or significant chi-square probabilities, or both.

INSERT TABLE 2 ABOUT HERE

Regarding DIF analyses, items 7 and 9 showed uniform DIF for gender. Item 13 showed uniform DIF for marital status (married *vs* non-married). No items showed DIF for age (among subjects aged from 60-69, 70-79 and 80-plus) or educational level (basic *vs* advanced).

The distributions of items and persons were placed in a common metric scale. The average mean person location value was -1.444 (SD 1.39), which indicates that the subjects were below the average of the scale.

SCALE REDUCTION

Since the 15-item scale proved to be inadequate and showed high misfit statistics, the next step would be trying improvement strategies and rechecking fit statistics successively. Items were deleted based on elevated chi-square values and/or high residuals, indicating poor fit statistics. DIF was also checked in each step and items with DIF were deleted, if it was the case.

After each item exclusion, item-trait interaction was examined (chi-square and p value), as well as item and person means and standard deviations. Item residuals and chi-square values were also verified.

Deleting item 9 led to a minor improvement of the scale, but still item-trait interaction remained far from ideal ($\chi^2 = 167.31$, $p < .0001$). The deletion step was repeated for items 10, 7 and 15. Tests of local dependence and unidimensionality were carried out. Item-trait interaction, person fit, PSI and DIF were monitored in each step. Results showed local independence and unidimensionality were present with item deletion.

This 11-item version showed satisfactory item-trait interaction statistics ($\chi^2 = 63.22$, $p = 0.030$ (non-significant after Bonferroni correction)). However, item 5 showed uniform DIF for age, which cause important misperformance when the scale is to be used to subjects in a wide range of ages.

Thus, the item 5 was deleted and the 10-item solution was then tested. Scale remained unidimensional and with no local dependence. The item-trait interaction presented adequate results ($\chi^2 = 58.11$, $p = 0.031$). The Person Separation Index was

0.768, and no item showed misfit or DIF problems. The final 10-item version is described in Table III bellow.

INSERT TABLE 3 ABOUT HERE

Item fit statistics showed remarkable improvement. The abbreviated model fits the expected Rasch model. Item mean was -0.279 (SD 1.43) and person mean was -0.135 (SD 0.735). Person Separation Index (.768) indicates good reliability power for group basis, but not for an individual level.

Figure 1 illustrates the distribution of persons and items in a common metric scale, when the 10-item solution was analysed. The information curve shows that the area where the scale provides more information is not coincident with the person distribution. This finding indicates that the 10-item GDS Scale is not suitable for assessing subjects with mild levels of depression or no depressive symptoms at all.

INSERT FIGURE 1 ABOUT HERE

DISCUSSION

This paper aimed to verifying the psychometric properties of the 15-item version of the Geriatric Depression Scale in a Brazilian sample. It also aimed to exploring alternative formats of the scale if the original proved not to be psychometrically adequate.

The application of Rasch model allows the examination of scales in a detailed level, from the model as a whole to individual item fit. Basically, besides testing unidimensionality, the Rasch model tests whether the scale presents characteristics that permit parametric analyses. Among others properties involved, items could be summed free of bias if they represent the construct in the same way (i.e., if they have the same weight). If not, the instrument should be considered as ordinal in nature, and consequently must be analysed through non-parametric statistics (Tenant et al, 2004).

In fact, the use of Rasch analyses to the development, validation and refinement of instruments has markedly increased. It represents a powerful tool in assessing the psychometric performance of different versions of the same scale, and also derives a set of invariant items. These sets are then suitable for being applied in a series of distinct researches free of sampling bias.

Results of the Rasch analysis on the 15-item version of GDS indicate that there is a marked misfit between the observed and the expected model. Together with the individual item misfit, these findings show that the GDS-15 should not be considered an interval scale and, therefore, is not adequate for parametric calculations. In addition, Rasch analysis was able to demonstrate unidimensionality. In fact, Rasch methodology is considered the most powerful methodology for establishing unidimensionality (Doward et al, 2004).

Scale reduction was then carried out in order to search for an improved psychometric solution for the instrument in a Brazilian sample. Item deletion was carried out according to misfit indexes and DIF problems. This strategy is composed of several steps, which includes deleting one item per turn and reanalysing all fit statistics to check potential improvements. The 10-item version derived from the refinement shows improvements regarding item χ^2 and residuals. No DIF problems were found in the reduced version. In fact, the overall statistics demonstrates that the refined scale fits Rasch model, indicating it is thus suitable for parametric tests. Person Separation Index remained stable through the refinement process. When further item deletion was carried out (in order to derive even shorter versions of the scale), the PSI decreased to levels below .70 (data not shown).

These findings indicate that the 15-item structure, originally extracted from the extensive 30-item instrument based on diagnostic accuracy, was influenced by cultural bias when tested in a Brazilian sample. For this sample, refinement led to a more adequate scale, presenting suitable psychometric performance. It is stated that the sequential approach of instrument development is more prone to cultural bias. In this process, the instrument is originated in a certain country and is then translated to a target idiom (Schmidt and Bullinger, 2003). It is suggested that alternative approaches (such as the simultaneous) are more adequate to derive cross-culturally valid instruments. This new 10-item set could represent a concise and strong scale to be tested in other cultures as an alternative to the 15-item GDS instrument.

In summary, the 15-item Geriatric Depression Scale proved not to fit the Rasch Model in its original format and, thus, not to be suitable for total score calculations in a Brazilian sample. Item deletion led to a 10-item solution, which has a strong scale structure and is suitable for all sorts of parametric statistics with no loss of performance. The present study offers an alternative and more adequate version of the GDS to be applied specifically in Brazilian subjects. Further investigations are needed to develop a set of cultural-invariant items, which could then be applied in transcultural investigations free of bias.

CONFLICT OF INTEREST DECLARATION: None

DESCRIPTION OF AUTHORS' ROLES: All authors were involved in designing the study and writing the paper. E. Chachamovich also supervised the data collection and carried out the statistical analysis. M. Power collaborated in the statistical analysis.

REFERENCES

1. Almeida O and Almeida S. (1999). Confiabilidade da versao brasileira da escala de depressao em geriatria (GDS) versao reduzida. *Arquivos de Neuropsiquiatria*, 57(2-B):421-426.
2. Andrich D, Lyne A, Sheridan B and Luo G(2003). RUMM 2020. Perth: RUMM Laboratory.
3. Beekman AT, Pennix BW, Deeg DJ et al (2002). The impact of depression on the well-being, disability and use of services in older adults: a longitudinal perspective. *Acta Psychiatrica Scandinavica*, Jan, 105(1):20-7.
4. Blazer DG and Hybels C (2005). Origins of depression in later life. *Psychological Medicine*, 35:1-12.
5. Bond TG and Fox CM. (2007). Applying the Rasch Model: Fundamnetal Measurement in Human Sciences. 2nd Edition, Lawrence Erlbaum Associates, USA.
6. Chachamovich E, Trentini CM and Fleck MP (2007). Assessment of the psychometric performance of the WHOQOL-BREF instrument in a sample of Brazilian older adults. *International Psychogeriatrics*; 19,4:635-646.
7. Chau J, Martin CR, Thompson DR, Chang AM, Woo J. (2006). Factor structure of the Chinese version of the Geriatric Depression Scale. *Psychology, Health and Medicine*, Feb; 11(1):48-59.
8. de Craen AJM, Heeren TJ, Gussekloo J (2003). Accuracy of the 15-item Geriatric Depression Scale (GDS-15) in a community sample of oldest old. *International Journal of Geriatric Psychiatry*,18:63-66.
9. D'Ath P, Katona P, Mullan E et al (1994). Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS) and the development of shorter versions. *Family Practice*, 11(3):260-266.
10. Doward LC, Meads D, Thorsen H. (2004). Requirements for Quality of Life instruments in Clinical Research. *Value in Health*, 7(S1):S13-S16.

11. Friedman B, Heisel M and Delavan RL (2005). Psychometric Properties of the 15-item Geriatric Depression scale in functionality impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatric Society*, 53:1570-1576.
12. Heisel MJ, Flett GL, Duberstein PR, Lyness JM, (2005). Does the geriatric depression scale (GDS) distinguish between older adults with high versus low levels of suicidal ideation? *American Journal of Geriatric Psychiatry*; Oct;13(10):876-83
13. Incalzi RA, Cesari M, Pedoni C et al (2003). Construct validity of the 15-item geriatric depression scale in older medical inpatients. *Journal of Geriatric Psychiatry and Neurology*., Mar;16(1):23-8
14. Jongenelis K, Pot AM, Eises AMH, Gerritsen DL et al. (2005). Diagnostic accuracy of the original 30-Item and shortened versions of the Geriatric Depression Scale in nursing home patients. *International Journal of Geriatric Psychiatry*, 20:1067-1074.
15. Korner A, Lauritzen L, Abelskov K, Gulmann NC, Brodersen AM, Wedervang-Jensen T, Marie Kjeldgaard K. (2007). Rating scales for depression in the elderly: external and internal validity. *Journal of Clinical Psychiatry*, Mar;68(3):384-9.
16. Linacre JM (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7:28.
17. Malakouti SK, Fatollahi P, Mirabzadeh A et al (2006). Reliability, validity and factor structure of the GDS-15 in Iranian elderly. *International Journal of Geriatric Psychiatry*, 21:288-293.
18. McKenna SP, Doward LC, Meads DM, Tenant A, Lawton G and Grueger J. (2007). Quality of life in infants and children with atopic dermatitis: Addressing issues of differential item functioning across cultures in multinational clinical trials. *Health and Quality of Life Outcomes*, 5:45, doi 10.1186/1477-7525-5-45.
19. McDonald WM, Holtzheimer PE, Haber M et al (2006). Validity of the 30-item Geriatric Depression Scale in patients with Parkinson's Disease. *Movement Disorders*, 21(10):1618-1622.
20. Osborn DPJ, Fletcher AE, Smeeth L et al. (2003). Factors associated with depression in a representative sample of 14217 people aged 75 and over in the United Kingdom: Results from the MRC trial of assessment and management of older people in the community. *International Journal of Geriatric Psychiatry*, 18:623-630.

21. Osborn DPJ, Fletcher A, Smeeth L et al. (2002). Geriatric Depression Scale Scores in a representative sample of 14545 people aged 75 and over in the United Kingdom: results from the MRC Trial of Assessment and Management of Older People in the Community. *International Journal of Geriatric Psychiatry*, 17:375-382.
22. Pallant J and Tenant A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46:1-18.
23. Pallant J, Miller R and Tenant A. (2006). Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry*,6:28-38.
24. Power MJ, Quinn K, Schmidt S et al. (2005). Development of the WHOQOL-old module. *Quality of Life Research.*, Dec;14(10):2197-214.
25. Saxena S, Carlson D, Billington R and Orley J. (2001). The WHO quality of life assessment instrument (WHOQOL-BREF): The importance of its items for cross-cultural research. *Quality of Life Research*, 10:711-721.
26. Schmidt S and Bullinger M. (2003). Current issues in cross-cultural quality of life instrument development. *Archives of Physical Medicine Rehabilitation*; 84, Suppl :S29-34.
27. Sheik JI and Yesavage JA (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontology*, 37:819-820.
28. Street H, O'connor M and Robinson H (2007). Depression in older adults: exploring the relationship between goal setting and physical health. *International Journal of Geriatric Psychiatry*, Mar, 29.
29. Tenant A, McKenna S and Hagell P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health*, 7(S1):S22-S26.
30. Yesavage JA, Brink TL, Rose T et al (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatry Research*, 17(1):37-49.

Table I - Demographics

Age	N (%) or M (SD)
60-69 years old	173 (40.9)
70-79 years old	153 (36.2)
80 plus years old	97 (22.9)
Gender	
Male	152 (35.8)
Female	272 (64.2)
Marital Status	
Single	29 (6.8)
Married	212 (50.0)
Separated	30 (7.1)
Widowed	128 (30.2)
Educational Level	
Basic Level ^a	172 (40.9)
Advanced Level ^b	252 (59.1)

a=primary school, b=high school or university

Table II – Fit of 15-item GDS to Rasch model

Item	Location	SE	Fit Residual	DF	Chi-Sq	DF	Prob
Satisfied with life	0.904	0.157	-2.219	329	8.688	6	0.191916
Dropped activities	-1.047	0.121	-0.788	329	8.341	6	0.214161
Feel life is empty	-0.161	0.13	-2.907	329	14.477	6	0.024743
Often get bored	-0.472	0.125	-0.504	329	15.816	6	0.014779
Are in good spirits	0.184	0.136	-1.809	329	12.572	6	0.050357
Afraid something bad	0.527	0.145	-1.804	329	10.858	6	0.09286
Feel happy	0.298	0.139	-3.631	329	22.782	6	0.000874
Often feel helpless	0.569	0.146	-1.878	329	5.014	6	0.542055
Prefer stay home	-1.455	0.121	3.935	329	32.587	6	0.000013
Feel memory problems	-0.419	0.126	3.883	329	34.28	6	0.000006
Wonderful be alive	1.453	0.181	0.633	329	13.338	6	0.037973
Fell worthless	0.726	0.151	-2.758	329	11.294	6	0.079703
Full of energy	-0.928	0.121	-1.687	329	13.597	6	0.034472
Situation is hopeless	-0.171	0.129	-0.356	329	5.132	6	0.526959
Most people are better	-0.008	0.132	1.471	329	7.194	6	0.303293

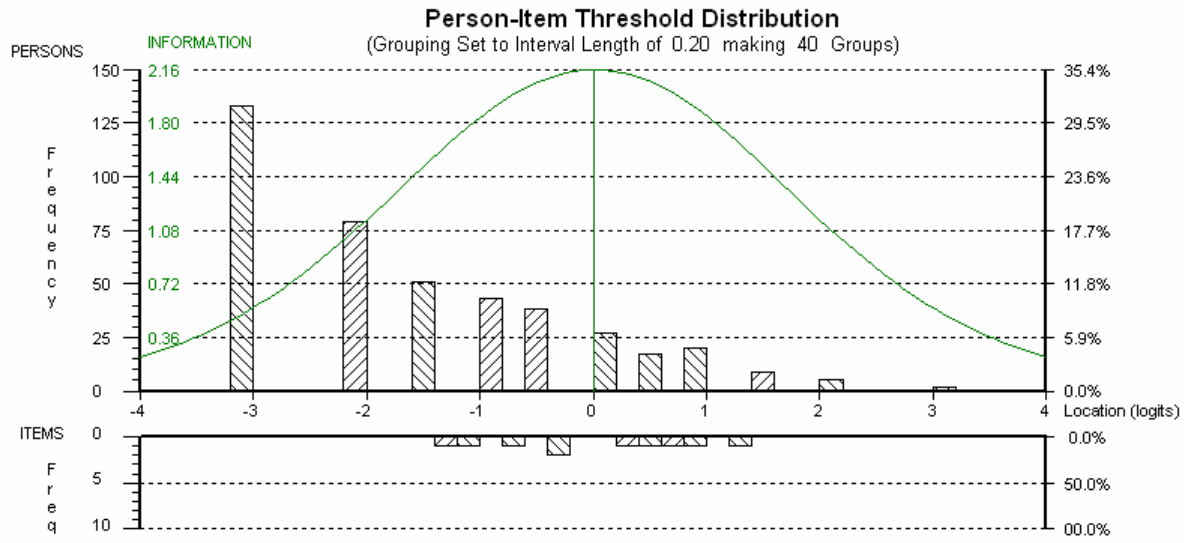
Misfitting values are in bold (after Bonferroni correction)

Table III – Fit of the 10-item GDS to Rasch model

Item	Location	SE	Fit Residual	DF	Chi-Sq	DF	Prob
Satisfied with life	0.832	0.163	-1.88	259	10.772	4	0.029251
Dropped activities	-1.266	0.132	-0.138	259	3.236	4	0.519169
Feel life is empty	-0.314	0.136	-2.204	259	12.258	4	0.015535
Often get bored	-0.625	0.133	1.703	259	5.995	4	0.199537
Afraid something bad	0.397	0.15	-0.612	259	2.922	4	0.571043
Often feel helpless	0.439	0.151	-0.708	259	3.424	4	0.489468
Wonderful be alive	1.329	0.183	0.917	259	8.892	4	0.063852
Fell worthless	0.617	0.156	-2.004	259	7.795	4	0.09937
Full of energy	-1.086	0.132	1.161	259	0.449	4	0.978229
Situation is hopeless	-0.323	0.136	0.979	259	2.374	4	0.667329

No misfitting values

Figure 1- Targeting of the 10-item GDS



10. Considerações Finais

A implementação de novos modelos de testes estatísticos na área da saúde acarretou uma rica e profícua discussão acerca do processo de mensuração de constructos complexos e abstratos. Tal discussão teórica se destaca na área da saúde mental, caracterizada fundamentalmente por lidar com tais constructos.

Mais além da revisão teórica pura ou de debates filosóficos acerca da investigação de fenômenos mentais, o estudo em profundidade dos novos modelos estatísticos acarretou marcadas e crescentes mudanças nas clássicas estratégias de medida de processos mentais. Entre os modelos chamados modernos em psicometria, o Modelo de Rasch representa uma ferramenta útil e robusta, uma vez que possibilita que os instrumentos sejam avaliados mais detalhadamente em relação ao seu desempenho.

O desenvolvimento de novos instrumentos de medida em psiquiatria e psicologia, como de resto em várias áreas do conhecimento médico, tem incluído as inovações da Teoria de Resposta ao Item. Validações de instrumentos já utilizados para outras culturas também se beneficiam de modelos modernos que assegurem sua validade. A busca de evidências na literatura indexada sobre psicometria ilustra o crescente uso de tais estratégias estatísticas.

A presente Tese de Doutorado busca demonstrar a aplicação do Modelo de Rasch no desenvolvimento de um instrumento novo, com o objetivo de mensurar a percepção subjetiva do processo de envelhecimento.

O instrumento AAQ, originalmente construído em inglês a partir de uma coleção transcultural de dados, foi então validado em sua versão em Português. Ademais, os dados gerados pela versão em Português foram comparados aos internacionais, com o intuito de assegurar o desempenho satisfatório do novo instrumento.

Em relação a instrumentos já existentes e há muito utilizados em investigações de populações idosas (tais como o Geriatric Depression Scale), pretendeu-se examinar em mais detalhes tais instrumentos tanto em nível nacional como em nível internacional. Deste modo, alterações são sugeridas para que os dados gerados a partir destes instrumentos apresentem maior credibilidade.

Fatores clínicos foram examinados em populações idosas. Particularmente, depressão clínica e subclínica foram avaliadas em relação ao impacto que determinam na qualidade de vida e na percepção de envelhecimento em amostras internacionais. O reflexo do analfabetismo no desempenho de escalas de Likert com varias categorias de resposta foi também examinado, demonstrando que idosos com baixa escolaridade tendem a não considerar tantas categorias de resposta como os de alto nível de escolaridade.

Por fim, a presente Tese de Doutorado teve como intuito colaborar com o refinamento de medidas em psiquiatria e psicologia, de modo a possibilitar que estratégias modernas de psicometria sejam incorporadas à pesquisa. O refinamento de medidas, especialmente com a obtenção de medidas de estrutura intervalar, acarretarão resultados mais fidedignos de importantes avaliações na área da saúde.

11. Anexos

11.1 Artigo Anexo

**Impact of major depression and subsyndromal symptoms on Quality of
Life and Attitudes to Aging in an international sample of older adults**

Aceito para publicação em The Gerontologist (The Gerontological Society of
America)

**Impact of major depression and subsyndromal symptoms on Quality of Life and
Attitudes to Aging in an international sample of older adults**

Authors: Eduardo Chachamovich¹, Marcelo Fleck¹, Ken Laidlaw², Mick Power²

Affiliations : 1- Department of Psychiatry, University of Rio Grande do Sul, Brazil
2- Clinical and Health Psychology, University of Edinburgh

Correspondence to: Eduardo Chachamovich
Rua Florencio Ygartua 391/308
Porto Alegre – RS
CEP 90430-010
Brazil
Phone: +55 51 91230015
Fax : +55 51 32644152
Email: echacha.ez@terra.com.br

Authors' contact details:

a) Marcelo P Fleck
Rua Ramiro Barcelos 2350
Porto Alegre - RS
CEP 90035-903
Brazil
Phone: +55 51 33308965
Email: mfleck.voy@terra.com.br

b) Ken Laidlaw
Medical School
Teviot Place
Edinburgh EH8 9AG.
Tel: 0131 651 3943
FAX: 0131 651 3971
Email: k.laidlaw@ed.ac.uk

c) Mick Power
Medical School
Teviot Place
Edinburgh EH8 9AG.
Tel: 0131 651 3943
FAX: 0131 651 3971
Email: mjpower@ed.ac.uk

Impact of major depression and subsyndromal symptoms on Quality of Life and Attitudes to Aging in an international sample of older adults

ABSTRACT

Purpose of the study: The impact of major depression in quality of life (QOL) and aging experiences in older adults has been reported. Studies have demonstrated that the clinical diagnosis of major depression is the strongest predictor for QOL. It is postulated that some findings are biased because of the use of inadequate instruments. Although subsyndromal depression is more prevalent than major depression, there are not any reports on its impact on QOL and/or attitudes to aging. The present study aims at assessing the association of major and subsyndromal depression on QOL and attitudes to aging in a large international sample.

Design and method: 4316 subjects were assessed in 20 countries from the five continents. Study design was cross-sectional. The WHOQOL-OLD, WHOQOL-BREF and AAQ were used. Statistical analyses involved hierarchical multiple regression, as well as comparison of means.

Results: Even relatively minor levels of depression are associated with significant decrease in all QOL domains and with negative attitudes to aging pattern (overall WHOQOL-OLD R^2 change=.421). QOL and attitudes to aging scores are lower as depression intensity is increased, even in subsyndromal levels (overall WHOQOL-OLD mean scores of 95.7 vs 86.4, $p<0.001$). This phenomenon happens not only for the clinical depressed group, but also for subsyndromic subjects.

Implications: Present findings suggest that classifying a subject as non-depressed is not sufficient and still is not informative about his/her QOL and attitudes to aging status.

Keywords: well-being, aging, subclinical,

INTRODUCTION

Depression has been recognized as a major mental health problem in older age. It is also the most prevalent mental condition in the elderly (Chan, Chien, Thompson, Chiu, & Lam, 2006). It is currently projected that depression will be the second leading cause of disability world wide in 2020 (Demyttenaere et al., 2004). Furthermore, it carries a poor prognosis, being associated with increased mortality, morbidity and use of health facilities (Covinsky et al., 1999; Dozeman et al., 2007; Rovner, 1993; Street, O'Connor, & Robinson, 2007).

The impact of depression in quantitative outcomes (such as mortality) has been extensively demonstrated (Covinsky et al., 1999). In addition, the role that major depression plays in the quality of life and aging experiences in older adults has been reported (Chan et al., 2006; Low & Molzahn, 2007). Several studies have demonstrated that the clinical diagnosis of major depression is the most influential predictor for impairments in quality of life (Chan et al., 2006; Netuveli, Wiggins, Hildon, Montgomery, & Blane, 2006; Sobocki et al., 2007; Stafford, Berk, Reddy, & Jackson, 2007), even when confounding factors (such as age, gender, living arrangements or physical conditions) are controlled.

It is postulated that some findings may be biased due to the inadequacy of the instruments applied in these investigations. The majority of instruments have been used without adequate validation (Haywood, Garratt, & Fitzpatrick, 2005), do not take into consideration several aspects of life that older adults consider fundamental (Pearlman & Uhlmann, 1988), and/or are not suitable for older adults since they have been developed primarily for young adult populations (Brazier, Walters, Nicholl, &

Kohler, 1996; Chachamovich, Trentini, & Fleck, 2007; Power, Quinn, & Schmidt, 2005).

The prevalence of subsyndromal depression in the elderly is more than double the one for major depression (D. G. Blazer, 2003; Geiselman & Bauer, 2000; Geiselman, Linden, & Helmchen, 2001; Snowdon, 2001; Watson, Lewis, Kistler, Amick, & Boustani, 2004). In the United States, around 15% of the older adults are expected to present subthreshold depression, defined as depressive symptoms that fail to meet the full diagnostic criteria for a major depression episode (Judd, Schettler, & Akiskal, 2002). Opposite to major depression, the prevalence of subsyndromal depression seems to increase with advancing age (D. Blazer & Williams, 1980; Ernst & Angst, 1995; Lavretsky & Kumar, 2003; Tannock & Katona, 1995; VanItallie, 2005). Although number and severity of symptoms are lower than the full-blown syndrome, subsyndromal depression is associated with significant functional impairment and psychosocial disability (Penninx, Leveille, Ferrucci, van Eijk, & Guralnik, 1999; Wilms, Kanowski, & Baltes, 2000), increased risk of developing major depression at some point, and suicidal ideation (Geiselman & Bauer, 2000; Greden, 2001; VanItallie, 2005). However, there are not any reports on the impact that subthreshold depression may have on quality of life and/or attitudes to aging.

The WHOQOL-BREF instrument is a generic measurement of quality of life ("The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties," 1998). Its development was based on a transcultural simultaneous methodology (Bullinger M; Power M; Aaronson NK; Cella D; Anderson RT, 1996; Guillemin, 1995), which ensures its applicability in different cultural settings. Recent reports have described its satisfactory psychometric properties in older adult populations (Chachamovich et al., 2007; Hwang, Liang,

Chiu, & Lin, 2003). In addition, the WHOQOL-OLD and the AAQ instruments have been recently developed to provide more reliable and comprehensive tools in assessing quality of life and attitudes to aging in older adults (Laidlaw, Power, & Schmidt, 2007; Power et al., 2005). They are in line with the need of broader conceptual outcome measures, beyond the classic ones, such as mortality and morbidity. Since they were developed as specific instruments to elderly, particular attention was paid to guarantee that the specificities of old age would be properly covered by the instruments. Both instruments were developed using a combination of classical and modern psychometric approaches, which ensures the satisfactory psychometric performance of them (Laidlaw et al., 2007; Power et al., 2005). The WHOQOL Group recommends that the WHOQOL-OLD module should be applied together with one generic quality of life instrument (WHOQOL-100 or WHOQOL-BREF) to enhance validity (Power et al., 2005).

The aim of this study is assessing the association of depression symptoms on the quality of life and attitudes to aging in a large international sample of older adults. Furthermore, it aims at exploring to what extent subclinical depression symptoms also determine alterations in the quality of life and attitudes to aging in this population.

DESIGN AND METHODS

SUBJECTS

Data were obtained from a larger study on quality of life in older adults (WHOQOL-OLD). This project was carried out by the World Health Organization Quality of Life Group (WHOQOL) and involved 20 countries around the world (Power et al., 2005).

Opportunistic samples were recruited in participant centers across Europe, Asia, South America and North America. A total of 4316 adults aged 60 years old or older were interviewed and completed the 15-item Geriatric Depression Scale, WHOQOL-OLD, AAQ and a socio-demographic form. Recruitment was carried out in university hospitals, nursing homes, and in community groups for older adults.

PROCEDURES

The assessment was run in a cross-sectional design. The subjects were required to complete a socio-demographic form, the 15-item Geriatric Depression Scale (Sheik, 1986), the WHOQOL-BREF instrument ("The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties," 1998), the WHOQOL-OLD module (Power et al., 2005) and the Attitudes to Aging Questionnaire (Laidlaw et al., 2007).

The Geriatric Depression Scale (GDS) is a widely used instrument for assessing depression specifically for older adults. Its theoretical background is based on the ability to distinguish depressed elders from the ones with 'normal aging' neurovegetative symptoms. Thus, the GDS focuses on nonsomatic symptomatology, emphasizing the affective and cognitive areas (de Craen, Heeren, & Gussekloo, 2003; McDonald et al., 2006; Yesavage et al., 1982). It has not been validated cross-nationally up to the present (this issue is being addressed in another paper currently). Thus, each center has applied a nationally validated version. The optimal cut-off score of 5/6 presents satisfactory sensitivity and specificity (Almeida & Almeida, 1999; Rinaldi et al., 2003; Sutcliffe et al., 2000; Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006), although alternative cut-points were suggested for specific

conditions (Cullum, Tucker, Todd, & Brayne, 2006; Lim et al., 2000; Malakouti, Fatollahi, Mirabzadeh, Salavati, & Zandi, 2006). Since the instrument has not been validated internationally yet, its cut-point is not consensual. For the present study, the 5/6 cut-point was adopted.

The WHOQOL-BREF instrument is a generic questionnaire to assess quality of life. Developed under a simultaneous transcultural approach (Bullinger M; Power M; Aaronson NK; Cella D; Anderson RT, 1996; Guillemin, 1995), it is suitable for multi-centric studies and is also validated for older adults (Chachamovich et al., 2007; Hwang et al., 2003; Naumann & Byrne, 2004). It is composed of four domains (Physical, Psychological, Social and Environmental) and an Overall score.

The WHOQOL-OLD module is a specific complementary set of items recently developed to increase specificity of quality of life measurement in older adults (Power et al., 2005). It comprises six domains (Sensory Abilities, Autonomy, Past-Present and Future Activities, Social Participation, Death and Dying and Intimacy) and an Overall score. Furthermore, it proved to be suitable and adequate to measuring comprehensively QOL in older adults, since an intense qualitative phase ensured that relevant issues were properly covered by this instrument (Chachamovich et al., 2007; Power et al., 2005).

The Attitudes to Aging Questionnaire (AAQ) is an instrument developed by the World Health Organization Quality of Life Group to assess the subjective perception of aging. It also followed a transcultural methodology and shows good psychometric performance (Laidlaw et al., 2007). It is composed of three facets (Psychological Growth, Physical Change and Psychosocial Loss) and was also based in an intense theoretical debate among international experts, as well as in focus groups carried out with older adults to confirm or adjust the instrument items. It is also able

to assess attitudes to aging from the perspective of the elderly, the one considered the most capable of informing about the experience of aging.

STATISTICAL ANALYSIS

Statistical analyses were run in SPSS 14.0 software (SPSS, 2006). The total sample was analyzed through descriptive statistics for both clinical and subsyndromal depression. If distribution of demographic variables was found to be significantly different, an analysis of covariance (ANCOVA) was run to partial out the variance for which other predictors (age, educational level, gender and marital status) could account (Field, 2005). Independent t-tests were carried out to compare the quality of life and attitudes to aging scores according to the predictors.

Hierarchical regression was applied in each domain of WHOQOL-OLD, WHOQOL-BREF and AAQ instruments to detect the impact of depression. Age, gender, educational level and marital status were included as independent variables together with depression levels.

Secondly, the subsample with subsyndromal depression levels was selected and divided into two groups (one with GDS scores from 0 to 2, and another with scores from 3 to 5). This cut-point was based on the median of the GDS score distribution. Again, demographic distribution verification and Ancova analysis were applied to ensure comparability. Independent t-tests were run for each domain. At this stage, these comparisons had the aim of detecting the impact of small increases of depression scores even in a subsyndromic level. Hierarchical regression was also carried out to detect the relative weight of subsyndromic depression on quality of life and attitudes to aging.

Results are described through means and p-values (chi-square and t-tests), and R^2 changes and standardized β -values (for hierarchical multiple linear regression).

RESULTS

Descriptives. 4316 subjects were assessed. The sample was predominantly comprised of highly educated married women. Recruitment was carried out in homes primarily (76.6%), followed by community sites (14.7%), and hospitals (8.7%). No relevant correlation was found between the site of recruitment and the quality of life or attitudes to aging domains scores (Pearson coefficients from 0.065 to 0.155). Table 1 describes the demographic characteristics of the total sample divided into subsyndromal group and clinical depression group. There is a marked predominance of subjects with depression scores lower than the diagnostic cut-point (73.84%).

INSERT TABLE 1 HERE

The WHOQOL-BREF, WHOQOL-OLD and AAQ domain scores were compared across the clinically depressed and the subsyndromic groups. All quality of life and attitudes to aging scores are significantly lower in the depressed group ($p < .001$, data not shown). These findings point out a wide and intense association of major depression in quality of life outcomes.

This association, however, could be partially related to the discrepancy in the demographics distribution. To control this potential bias, an ANCOVA analysis was run. Table 2 describes the ANCOVA results. As shown in this table, the interaction among these variables assumes statistical significance in only four out of the 16

scores. Furthermore, the interaction explains not more than 3% of the total variance of the score in these four domains. Thus, the difference in quality of life and attitudes to aging scores between the two samples appears to be strongly related to depression levels, and not to demographic dissimilarities.

INSERT TABLE 2 HERE

Potential multicollinearity was tested through the variance inflation factor (VIF) in each regression model. VIF values ranged from .903 to 1.09, which indicate that the predictors do not have strong linear relationships among them (Field, 2005; Hair, 1998).

A hierarchical multiple linear regression was conducted to detect the increase of the coefficient of determination as new independent variables were included in the model. Data distribution was evaluated by the Kolmogorov-Smirnov test, showing normal distribution of all dependent variables (QOL and AAQ domains scores). Linearity and variance were also checked and met the linear regression assumptions. Gender, age, marital status, and educational level were selected to compose the regression model together with depression levels since they represent important demographic variables and are widely reported in several studies (Chan et al., 2006; Low & Molzahn, 2007). The standardized β -coefficients were also examined to compare the impact of the independent variables in the QOL and attitudes to aging domains. Tables 3 and 4 illustrate the R^2 increase and the β -coefficients for each domain, as well as for the overall scores in the three scales.

The inclusion of depression in the model determined a marked increase in the coefficient of determination. In fact, almost all of the variance is explained by

depression. The magnitude of the effect caused by depression in the model is greatly higher than the ones by the other variables, as can be seen by both the R^2 change and the β -coefficients.

Interestingly, the Death and Dying and the Psychological Growth domains seem to be less related to depression. In fact, the proposed model explains only 8.4 and 7.3% of the total variance, respectively. As expected, depression is negatively associated with all scores in different intensities.

INSERT TABLES 3 AND 4 HERE

Besides testing the association of major depression on quality of life and attitudes to aging, the subsyndromal symptomatology was also explored. The main objective of these further analyses were to investigate whether an increase of depression levels can be related to quality of life and attitudes to aging impairments, even when they are not considered clinically relevant.

A subsample of 3187 subjects was selected from the total sample. All subjects with GDS scores of 5 or lower were included in this analysis stage. To test the effect of subsyndromal depression in the QOL and AAQ scores, the sample was divided into two. The first group was composed of 1782 subjects with GDS scores equal to or lower than 2. The second group included subjects with GDS scores from 3 to 5.

Demographics are described in Table 5 bellow.

INSERT TABLE 5 HERE

Demographic variables were tested through chi-square and independent t-test statistics. Following the same strategy applied for the total sample and already described, an ANCOVA test was used to assess a potential effect of this different distribution. Results indicated that the interaction among gender, educational level and marital status is not significant for any domain. Partial Eta Squares were 0% for all the 16 scores (data not shown).

Table 6 shows the comparison of mean scores for all domains. It is possible to observe that the quality of life and attitudes to aging scores significantly is decreased as depression levels are increased.

INSERT TABLE 6 HERE

In fact, the analysis of the scores of all domains indicates a progressive impairment with increasing depression levels, even when non-clinical depression is included. Figure I bellow illustrates the decreasing QOL and AAQ overall scores in four different levels of depressive symptoms. An ANOVA analysis showed that the differences are statistically significant at all points ($p < .0001$). In addition, it seems that the AAQ scores are less intensely associated to the increase of depression levels when compared to the quality of life scores.

INSERT FIGURE 1 HERE

Hierarchical multiple linear regression was also applied in this subsample. Again, the increase in the coefficient of determination was assessed, as well as the

standardized β -coefficients. As expected, the variances of the domain scores were lower than the one observed in the total sample. As a result, total R^2 and changes in the coefficient were smaller.

The results indicate the same pattern observed in the findings from the total sample. Again, the magnitude of the effect caused by the inclusion of depression in the model is greatly higher than the ones caused by the other variables. This suggests that slight increase in depression level (in a subsyndromal scenario) is a major predictor of quality of life and attitudes to aging impairments.

The Psychological Growth and the Death and Dying domains again showed low coefficient of determination. However, the majority of this small coefficient was due to depression rather than any other independent variable in the model.

Tables 7 and 8 describe the hierarchical regression findings.

INSERT TABLES 7 AND 8 HERE

DISCUSSION

This study aimed at exploring the association of clinical and subclinical depression in the quality of life and attitudes to aging in a large international sample of older adults. The results suggest that even relatively minor levels of depression are associated with significant decrease in quality of life and with negative attitudes to aging pattern. These findings assume particular relevance due to the fact that the used instruments are adequate for investigations in older adults. This ensures the reliability

of the results. Furthermore, the large international sample could be seen as an important factor of external validity.

Studies regarding attitudes to aging have demonstrated the effect of culture in the perception of aging. Interestingly, stereotypes of aging proved not to be adequate to describe the aging process (Yun & Lachman, 2006). Quality of life is also intimately linked to cultural aspects, since culture and value systems of the society are part of the QOL definition (Schmidt & Bullinger, 2003; "The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties," 1998). The AAQ and the WHOQOL instruments were constructed under a simultaneous methodology, which included input from several cultural contexts in the development phase. Thus, they seem to represent what Draper nominates as a "culture-proof variable", since they are able to assess meaningfully abstract concepts (Draper, 2007).

The link between QOL and attitudes to aging is yet to be further explored. Some similarities should be highlighted. First, both concepts are based predominately on the subjective perception rather than objective conditions. Secondly, both are multidimensional and include physical and psychological aspects. On the other hand, it is possible to state that the attitudes to aging questionnaire assesses a more stable perception, while the QOL instrument would be related to a more circumstantial aspect. The combination of the WHOQOL-OLD and the AAQ is suggested as an effective methodology for assessing psychological and/or psychiatric interventions (Laidlaw et al., 2007).

The sequential comparisons of the mean scores of quality of life and attitudes to aging domains indicate progressive impairment on both outcomes with increasing depression levels. This finding is clinically relevant and is in line with a series of

recent studies (Chan et al., 2006; Netuveli et al., 2006; Sobocki et al., 2007; Stafford et al., 2007). However, the consistent result of impairment in all quality of life and AAQ domains with a slight increase in depression levels (in a subsyndromal stage) has not been reported up to date. When analyzed as a whole, the present results show that there is a decrease in QOL and AAQ as depression levels increase. This phenomenon happens not only for the clinical depressed group (i.e., subjects with 6-9 symptoms and subjects with 10 or more symptoms on GDS scale), but also for subsyndromic subjects.

Thus, findings suggest that classifying a subject as non-depressed is not sufficient and still is not informative about his/her quality of life and attitudes to aging status. Historically, depression has been under-diagnosed and under-treated in later life (Alexopoulos, 2005), and this is evidently more likely to occur if a person does not meet diagnostic criteria (Jeste, Blazer, & First, 2005). These results would raise the hypothesis that subsyndromal depression could be treated because of the impact of minor scores on the GDS scale on appraisals of aging and quality of life. Potential increase on quality of life and attitudes to aging of patients treated for subsyndromic depression symptomatology appears as a question to be further addressed in longitudinal studies.

The hierarchical multiple regressions produced interesting results. Firstly, depression accounted for the vast majority of coefficients of determination in all domains. The standardized β -coefficients also showed a high difference of magnitude between depression and age, gender, marital status and educational level. Furthermore, depression was the only predictor with significant values for all outcomes in both total and subsyndromal samples. In fact, depression has been described as the predominant factor contributing to morale (Woo, Ho, & Wong,

2005), as well as the strongest predictor for subjective perception of quality of life (Chan et al., 2006; Demura & Sato, 2003). The present findings corroborate these previously reported results.

Secondly, there is a marked difference in the extent to which the regression model explained the scores variances. It is suggested that moderate coefficients of determination are expected for models including quality of life since it is a wide and comprehensive construct, which would require complex models to provide adequate explanation (Bowling, Banister, Sutton, Evans, & Windsor, 2002). For example, recent studies using linear regression methodology on quality of life reported model R^2 values of .568 (Low & Molzahn, 2007) and .475 (Netuveli et al., 2006). Since the model applied in the present study is simpler than the ones cited above, it is expected the R^2 values to be lower. In fact, the values in the present study ranged from .214 to .476 (with the exception of Death and Dying and Psychological Growth) in the total sample. These values are considered sizeable given the complex nature of the dependent variables (Bowling et al., 2002). On the other hand, the coefficient of determination was markedly smaller for the Death and Dying and Psychological Growth domains, which suggest that there may be other important and specific factors which account for the variation of these areas. Our data were not able to address this question, though. It is important to note that the Death and Dying domain seems to present some potential psychometric weaknesses, as demonstrated by other reports (Chachamovich et al., 2007; Power et al., 2005). It is hypothesized that this domain would perform better when applied to unhealthy subjects, but would loose power when healthy subjects are included in the sample. Studies on quality of life in HIV patients support this hypothesis (O'Connell, Skevington, & Saxena, 2003; "WHOQOL-HIV for quality of life assessment among people living with HIV and

AIDS: results from the field test," 2004). Regarding the Psychological Growth domain, the only paper published up to the present describes its satisfactory psychometric performance in an international dataset (Laidlaw et al., 2007), which indicates that there must be another cluster of variables for determining Psychological Growth, and that these variables are more specific for this domain. A close look to the eight items included in this domain suggests that generativity is one major topic (i.e., passing on the benefits of growing older, giving good examples to younger people) (Laidlaw et al., 2007). It is possible to hypothesize that older adults without the opportunity to be in contact with younger generations (e.g., without children or grandchildren) may not be adequately represented by this domain. It is extremely relevant to design investigations on this topic in order to implement specific interventions to improve psychological aspects of growing older.

As predicted, the same model produced lower coefficients of determination in the subsyndromal sample. However, the results shown for the total sample (i.e., elevated and significant standardized β -coefficients for depression in all domains and the pattern of R^2 values) remained similar for the subsyndromic sample. This finding reinforces the role of even minimal depression levels in the QOL and AAQ domain scores. The question that arises from these findings would be whether offering treatment for subsyndromal patients (either psychopharmacological or psychotherapy) could improve their quality of life and attitudes to aging. Specific designed studies to address this issue are needed.

Some limitations should be observed in the present study. The cross-sectional design does not allow inference of causality. It is, thus, possible to detect association between depression and QOL/AAQ, but not to state that the latter is caused by the former. Longitudinal studies are required to address this issue. In addition, clinical or

structured psychiatric interviews were not used to provide a clinical diagnosis of depression. However, the GDS was chosen due to the fact that it has been extensively studied for diagnostic purposes. Furthermore, we decided to use level of depression rather than the clinical diagnosis of depression in the analysis to avoid this potential bias. It is also important to observe that the tested model did not include other demographic, clinical and social variables, which could be potentially related to depression and QOL (such as impaired functioning, life stressors, social support and physical health). These non-evaluated variables may play a role in the relationship between depression levels and quality of life.

In summary, the present study showed that the increase of depression levels, both in clinical and subsyndromic stages, is a major predictor of quality of life and attitudes to aging impairments for older adults.

Acknowledgements: First author was partially funded by CAPES for a PhD Scholarship (number 3604/06-3). Authors thank all investigators from the twenty WHOQOL-OLD Centers involved.

Table 1 – Total sample characteristics (n=4316)

	GDS 0-5 n=3187	GDS 6-15 n=1129	P
	N (%) or M (SD)	N (%) or M (SD)	
Age			.000^a
Mean (SD)	71.8 (7.9)	73.06 (3.13)	
Gender			.000^b
Male	1399 (43.2)	402 (35.6)	
Female	1818 (56.8)	727 (64.4)	
Marital Status			.000^b
Single	172 (5.2)	73 (6.5)	
Married	2032 (62.3)	498 (44.0)	
Separated	245 (7.5)	117 (10.4)	
Widowed	811 (24.8)	441 (39.1)	
Educational Level			.000^b
Illiterate	63 (1.9)	91 (8.1)	
Basic Level	925 (28.3)	800 (70.9)	
College or Higher	2312 (70.9)	225 (19.9)	
Depression Level			.000^a
GDS 15 (Mean, SD)	2.57 (1.36)	8.49 (2.23)	

a= Independent t-test; b= Chi-Square test

Table 2- Ancova results of the interaction of Gender, Education Level and Marital Status for each domain

Interaction	Gender		Education Lev		Marital Status		Gender x Ed Lev x Mar		
	F	p	F	p	F	p	F	P	PES*
WHOQOL-OLD									
Sensory Ability	19.6	.000	6.9	.008	53.2	.000	5.6	.018	.001
Autonomy	2.6	.106	39.4	.000	.13	.713	1.74	.187	.000
Past, Pres and Fut Activities	.166	.683	7.47	.006	4.80	.028	.000	.987	.000
Social Participation	12.8	.000	14.0	.000	6.9	.008	.296	.587	.000
Death and Dying	32.0	.000	.249	.618	.000	.998	6.1	.014	.002
Intimacy	26.1	.000	7.1	.008	194.9	.000	1.39	.238	.000
Overall Score	.247	.620	19.1	.000	49.0	.000	9.6	.002	.003
AAQ									
Psychological Growth	5.0	.024	13.2	.000	.910	.340	.041	.840	.000
Physical Change	3.3	.069	.406	.524	11.0	.001	1.2	.237	.000
Psychosocial Loss	.294	.588	28.8	.000	44.2	.000	4.3	.037	.001
Overall Score	3.6	.060	16.3	.000	2.0	.154	.077	.782	.000
WHOQOL-BREF									
Physical	.269	.604	17.5	.000	7.25	.007	.230	.631	.000
Psychological	4.33	.038	1.05	.304	4.06	.044	.001	.973	.000
Social	30.4	.000	.208	.648	4.60	.032	.661	.416	.000
Environmental	1.14	.285	4.24	.039	1.08	.297	1.08	.298	.000
Overall Score	.769	.380	4.97	.026	2.56	.110	.481	.488	.000

* *Partial Eta Square, values with $p < .05$ in bold*

Table 3 – Description of R² changes in the hierarchical regressions for each WHOQOL-OLD, WHOQOL-BREF and AAQ domains

Independent Variables	WHOQOL-OLD							WHOQOL-BREF					AAQ			
	Se Ab	Auton	PPF	SP	DD	Int	Over	Phys	Psyc	Soc	Env	Over	PG	PL	PC	Over
	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch
Gender*	.003	.000	.000	.001	.013	.010	.001	.002	.005	.004	.004	.000	.000	.000	.000	.000
Age	.062	.013	.003	.017	.007	.014	.017	.048	.008	.000	.002	.006	.002	.038	.014	.026
Marital Status**	.000	.000	.001	.001	.000	.007	.001	.007	.007	.010	.005	.007	.000	.000	.000	.007
Educational Level	.010	.017	.008	.010	.004	.002	.022	.005	.002	.000	.006	.004	.002	.014	.006	.002
Depression	.159	.212	.324	.327	.060	.181	.421	.362	.454	.222	.319	.363	.069	.352	.222	.367
Total R²	.231	.242	.336	.356	.084	.214	.462	.424	.476	.236	.336	.380	.073	.404	.242	.402

1=male, 2=female; ** 1 = non-married, 2=married; Se Ab= sensory ability, Auton=autonomy, PPF=past,present and future activities, SP=social participation, DD=death and dying, Int=intimacy, Over=overall, Phys=physical, Psyc=psychological, Soc=social, Env=environmental, PG=psychological growth, PL=psychosocial loss, PC=physical change.

Table 4 – Standardized Beta Coefficients for each variable in the final model

Independent Variables	WHOQOL-OLD							WHOQOL-BREF				AAQ				
	Se Ab	Auton	PPF	SP	DD	Int	Over	Phys	Psyc	Soc	Env	Over	PG	PL	PC	Over
	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β
Gender*	.073	.028	.010	.061	-.079	-.087	-.006	-0.23	-.033	.094	-.023	.008	.026	.024	.026	.033
Age	-.195	-.042	.019	-.062	.112	-.081	-.066	-.134	-.004	.048	.080	-.004	-0.17	-.122	-.049	-.080
Marital Status**	.019	.038	-.017	.006	.030	-.083	-.10	-.017	-.016	-.063	-.041	-.025	.000	.025	.018	-.036
Educational Level	.059	.062	.009	.041	.044	-.022	.058	.050	-.002	-.018	.017	.015	-.0.81	.033	.021	-.010
Depression	-.399	-.485	-.581	-.576	-.252	-.394	-.655	-.607	-.684	-.479	-.575	-.611	-.270	-.595	-.479	-.614

*1=male, 2=female; ** 1 = non-married, 2=married; bolded values with p<.05

Table 5 – Demographics of the subsyndromal sample (n=3187)

	GDS 0-2 n=1782	GDS 3-5 n=1405	P
	N (%) or M (SD)	N (%) or M (SD)	
Age			
Mean (SD)	70.77 (7,6)	73.35 (8.1)	.000 ^a
Gender			
Male	789 (44.3)	587 (41.8)	.129 ^b
Female	993 (55.7)	817 (58.2)	
Marital Status			
Single	81 (4.0)	91 (6.2)	.000 ^b
Married	1240 (68.6)	782 (54.4)	
Separated	141 (7.8)	104(7.1)	
Widowed	344 (19.0)	467 (32.1)	
Educational Level			
Illiterate	10 (0.5)	53 (3.5)	.000 ^b
Basic Level	401 (22.0)	524 (35.4)	
College or Higher	1411 (77.4)	901 (60.9)	

a= independent t test; b= Chi-Square test

Table 6 - Comparison of means between subsamples

Domains	GDS 0-2	GDS 3-5	P value ^a
	Mean (SD)	Mean (SD)	
WHOQOL-BREF			
Physical	78.33 (13.2)	64.99 (15.7)	.000
Psychological	75.98 (10.8)	66.80 (11.7)	.000
Social	73.10 (14.3)	65.39 (15.6)	.000
Environmental	76.46 (13.1)	65.39 (15.6)	.000
Overall	76.86 (14.2)	64.68 (16.2)	.000
AAQ			
Psychological Growth	72.30 (11.0)	68.60 (10.7)	.000
Physical Change	73.77 (12.2)	65.05 (12.0)	.000
Psychosocial Loss	80.45 (11.7)	70.40 (12.0)	.000
Overall	75.47 (8.6)	68.04 (7.7)	.000
WHOQOL-OLD			
Sensory Abilities	84.90 (13.5)	77.75 (15.5)	.000
Autonomy	79.60 (12.0)	72.45 (12.9)	.000
Past, Pres and Future Activities	78.45 (10.5)	71.40 (11.0)	.000
Death and Dying	75.80 (17.5)	69.60 (19.5)	.000
Intimacy	77.20 (17.0)	69.70 (18.0)	.000
Overall	95.70 (9.7)	86.48 (9.8)	.000

a= independent t-tests

Table 7 – Description of R² changes in the hierarchical regressions for each WHOQOL-OLD, WHOQOL-BREF and AAQ domains (subsyndromal sample)

Independent Variables	WHOQOL-OLD							WHOQOL-BREF				AAQ				
	Se Ab	Auton	PPF	SP	DD	Int	Over	Phys	Psyc	Soc	Env	Over	PG	PL	PC	Over
	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch	R ² Ch
Gender*	.009	.001	.000	.002	.009	.006	.000	.001	.001	.009	.002	.000	.000	.001	.001	.001
Age	.064	.013	.003	.016	.005	.020	.027	.050	.006	.000	.000	.004	.000	.046	.008	.021
Marital Status**	.000	.004	.001	.001	.001	.005	.007	.001	.002	.005	.003	.002	.001	.002	.002	.004
Educational Level	.003	.008	.002	.006	.003	.001	.006	.007	.000	.001	.007	.005	.002	.003	.003	.001
Depression	.063	.101	.113	.170	.037	.042	.215	.193	.192	.088	.129	.170	.031	.165	.141	.204
Total R²	.139	.127	.119	.195	.055	.074	.255	.252	.201	.103	.141	.181	.033	.217	.154	.230

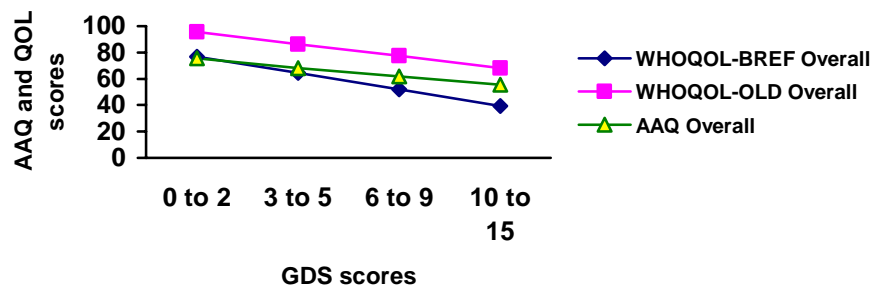
*1=male, 2=female; ** 1 = non-married, 2=married

Table 8 – Standardized Beta Coefficients for each variable in the final model (subsyndromal sample)

Independent Variables	WHOQOL-OLD							WHOQOL-BREF								
	Se Ab	Auton	PPF	SP	DD	Int	Over	Phys	Psyc	Soc	Env	Over	PG	PL	PC	Over
	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β	Std β
Gender*	.091	.038	.012	.058	-.082	-.087	.000	-.017	-.034	.101	-.026	.011	.004	.043	.031	.036
Age	-.201	-.042	.008	-.040	.122	-.093	-.058	-.134	.008	.038	.090	-.020	.013	-.130	-.011	-.054
Marital Status**	.012	.048	-.023	.015	.026	-.074	-.058	-.009	.017	-.028	-.018	.005	.017	.019	.033	-.035
Educational Level	.030	.054	.013	.030	.031	.008	.043	.059	.005	-.004	.070	.044	-.072	.014	.018	-.020
Depression	-.253	-.326	-.345	-.423	-.190	-.212	-.474	-.446	-.441	-.299	-.364	-.415	-.176	-.417	-.384	-.462

*1=male, 2=female; ** 1 = non-married, 2=married; bolded values with $p < .05$

Figure 1 - Impact of depression levels on QOL and AAQ scores



*ANOVA results with $p < .0001$ at all points

REFERENCES

- Alexopoulos, G. S. (2005). Depression in the elderly. *Lancet*, 365(9475), 1961-1970.
- Almeida, O. P., & Almeida, S. A. (1999). [Reliability of the Brazilian version of the abbreviated form of Geriatric Depression Scale (GDS) short form]. *Arq Neuropsiquiatr*, 57(2B), 421-426.
- Blazer, D., & Williams, C. D. (1980). Epidemiology of dysphoria and depression in an elderly population. *Am J Psychiatry*, 137(4), 439-444.
- Blazer, D. G. (2003). Depression in late life: review and commentary. *J Gerontol A Biol Sci Med Sci*, 58(3), 249-265.
- Bowling, A., Banister, D., Sutton, S., Evans, O., & Windsor, J. (2002). A multidimensional model of the quality of life in older age. *Aging Ment Health*, 6(4), 355-371.
- Brazier, J. E., Walters, S. J., Nicholl, J. P., & Kohler, B. (1996). Using the SF-36 and Euroqol on an elderly population. *Qual Life Res*, 5(2), 195-204.
- Bullinger M; Power M; Aaronson NK; Cella D; Anderson RT. (1996). Creating and evaluating cross-cultural instruments. . In S. B (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 659-668). Hagerstown, MD: Lippincott-Raven.
- Chachamovich, E., Trentini, C., & Fleck, M. P. (2007). Assessment of the psychometric performance of the WHOQOL-BREF instrument in a sample of Brazilian older adults. *Int Psychogeriatr*, 19(4), 635-646.

- Chan, S. W., Chien, W. T., Thompson, D. R., Chiu, H. F., & Lam, L. (2006). Quality of life measures for depressed and non-depressed Chinese older people. *Int J Geriatr Psychiatry, 21*(11), 1086-1092.
- Covinsky, K. E., Kahana, E., Chin, M. H., Palmer, R. M., Fortinsky, R. H., & Landefeld, C. S. (1999). Depressive symptoms and 3-year mortality in older hospitalized medical patients. *Ann Intern Med, 130*(7), 563-569.
- Cullum, S., Tucker, S., Todd, C., & Brayne, C. (2006). Screening for depression in older medical inpatients. *Int J Geriatr Psychiatry, 21*(5), 469-476.
- de Craen, A. J., Heeren, T. J., & Gussekloo, J. (2003). Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. *Int J Geriatr Psychiatry, 18*(1), 63-66.
- Demura, S., & Sato, S. (2003). Relationships between depression, lifestyle and quality of life in the community dwelling elderly: a comparison between gender and age groups. *J Physiol Anthropol Appl Human Sci, 22*(3), 159-166.
- Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J. P., et al. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Jama, 291*(21), 2581-2590.
- Dozeman, E., van Schaik, D. J., Beekman, A. T., Stalman, W. A., Bosmans, J. E., & van Marwijk, H. W. (2007). Depression and anxiety, an indicated prevention (DIP) protocol in homes for the elderly: feasibility and (cost) effectiveness of a stepped care programme. *BMC Geriatr, 7*, 6.

- Draper, P. (2007). Conducting cross-cultural research in teams and the search for the "culture-proof" variable. *Menopause, 14*(4), 680-687.
- Ernst, C., & Angst, J. (1995). Depression in old age. Is there a real decrease in prevalence? A review. *Eur Arch Psychiatry Clin Neurosci, 245*(6), 272-287.
- Field, A. (2005). *Discovering Statistics using SPSS*. London: Sage.
- Geiselmann, B., & Bauer, M. (2000). Subthreshold depression in the elderly: qualitative or quantitative distinction? *Compr Psychiatry, 41*(2 Suppl 1), 32-38.
- Geiselmann, B., Linden, M., & Helmchen, H. (2001). Psychiatrists' diagnoses of subthreshold depression in old age: frequency and correlates. *Psychol Med, 31*(1), 51-63.
- Greden, J. F. (2001). The burden of disease for treatment-resistant depression. *J Clin Psychiatry, 62 Suppl 16*, 26-31.
- Guillemin, F. (1995). Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol, 24*(2), 61-63.
- Hair, J. A., R; Tatham, R; Black, W. (1998). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2005). Quality of life in older people: a structured review of generic self-assessed health instruments. *Qual Life Res, 14*(7), 1651-1668.
- Hwang, H. F., Liang, W. M., Chiu, Y. N., & Lin, M. R. (2003). Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age Ageing, 32*(6), 593-600.

- Jeste, D. V., Blazer, D. G., & First, M. (2005). Aging-related diagnostic variations: need for diagnostic criteria appropriate for elderly psychiatric patients. *Biol Psychiatry*, *58*(4), 265-271.
- Judd, L. L., Schettler, P. J., & Akiskal, H. S. (2002). The prevalence, clinical relevance, and public health significance of subthreshold depressions. *Psychiatr Clin North Am*, *25*(4), 685-698.
- Laidlaw, K., Power, M. J., & Schmidt, S. (2007). The Attitudes to Ageing Questionnaire (AAQ): development and psychometric properties. *Int J Geriatr Psychiatry*, *22*(4), 367-379.
- Lavretsky, H., & Kumar, A. (2003). Practical geriatrics: clinically significant nonmajor geriatric depression. *Psychiatr Serv*, *54*(3), 297-299.
- Lim, P. P., Ng, L. L., Chiam, P. C., Ong, P. S., Ngui, F. T., & Sahadevan, S. (2000). Validation and comparison of three brief depression scales in an elderly Chinese population. *Int J Geriatr Psychiatry*, *15*(9), 824-830.
- Low, G., & Molzahn, A. E. (2007). Predictors of quality of life in old age: a cross-validation study. *Res Nurs Health*, *30*(2), 141-150.
- Malakouti, S. K., Fatollahi, P., Mirabzadeh, A., Salavati, M., & Zandi, T. (2006). Reliability, validity and factor structure of the GDS-15 in Iranian elderly. *Int J Geriatr Psychiatry*, *21*(6), 588-593.
- McDonald, W. M., Holtzheimer, P. E., Haber, M., Vitek, J. L., McWhorter, K., & Delong, M. (2006). Validity of the 30-item geriatric depression scale in patients with Parkinson's disease. *Mov Disord*, *21*(10), 1618-1622.

- Naumann, V. J., & Byrne, G. J. (2004). WHOQOL-BREF as a measure of quality of life in older patients with depression. *Int Psychogeriatr, 16*(2), 159-173.
- Netuveli, G., Wiggins, R. D., Hildon, Z., Montgomery, S. M., & Blane, D. (2006). Quality of life at older ages: evidence from the English longitudinal study of aging (wave 1). *J Epidemiol Community Health, 60*(4), 357-363.
- O'Connell, K., Skevington, S., & Saxena, S. (2003). Preliminary development of the World Health Organisation's Quality of Life HIV instrument (WHOQOL-HIV): analysis of the pilot version. *Soc Sci Med, 57*(7), 1259-1275.
- Pearlman, R. A., & Uhlmann, R. F. (1988). Quality of life in chronic diseases: perceptions of elderly patients. *J Gerontol, 43*(2), M25-30.
- Penninx, B. W., Leveille, S., Ferrucci, L., van Eijk, J. T., & Guralnik, J. M. (1999). Exploring the effect of depression on physical disability: longitudinal evidence from the established populations for epidemiologic studies of the elderly. *Am J Public Health, 89*(9), 1346-1352.
- Power, M., Quinn, K., & Schmidt, S. (2005). Development of the WHOQOL-old module. *Qual Life Res, 14*(10), 2197-2214.
- Rinaldi, P., Mecocci, P., Benedetti, C., Ercolani, S., Bregnocchi, M., Menculini, G., et al. (2003). Validation of the five-item geriatric depression scale in elderly subjects in three different settings. *J Am Geriatr Soc, 51*(5), 694-698.
- Rovner, B. W. (1993). Depression and increased risk of mortality in the nursing home patient. *Am J Med, 94*(5A), 19S-22S.
- Schmidt, S., & Bullinger, M. (2003). Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil, 84*(4 Suppl 2), S29-34.

- Sheik, J. Y., JA. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontology*, 37, 819-820.
- Snowdon, J. (2001). Is depression more prevalent in old age? *Aust N Z J Psychiatry*, 35(6), 782-787.
- Sobocki, P., Ekman, M., Agren, H., Krakau, I., Runeson, B., Martensson, B., et al. (2007). Health-related quality of life measured with EQ-5D in patients treated for depression in primary care. *Value Health*, 10(2), 153-160.
- SPSS, I. (2006). Statistical Package for the Social Sciences Chicago: Inc SPSS.
- Stafford, L., Berk, M., Reddy, P., & Jackson, H. J. (2007). Comorbid depression and health-related quality of life in patients with coronary artery disease. *J Psychosom Res*, 62(4), 401-410.
- Street, H., O'Connor, M., & Robinson, H. (2007). Depression in older adults: exploring the relationship between goal setting and physical health. *Int J Geriatr Psychiatry*, 22(11), 1115-1119.
- Sutcliffe, C., Cordingley, L., Burns, A., Mozley, C. G., Bagley, H., Huxley, P., et al. (2000). A new version of the geriatric depression scale for nursing and residential home populations: the geriatric depression scale (residential) (GDS-12R). *Int Psychogeriatr*, 12(2), 173-181.
- Tannock, C., & Katona, C. (1995). Minor depression in the aged. Concepts, prevalence and optimal management. *Drugs Aging*, 6(4), 278-292.
- The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. (1998). *Soc Sci Med*, 46(12), 1569-1585.

- VanItallie, T. B. (2005). Subsyndromal depression in the elderly: underdiagnosed and undertreated. *Metabolism, 54*(5 Suppl 1), 39-44.
- Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: a systematic review. *Acta Psychiatr Scand, 114*(6), 398-410.
- Watson, L. C., Lewis, C. L., Kistler, C. E., Amick, H. R., & Boustani, M. (2004). Can we trust depression screening instruments in healthy 'old-old' adults? *Int J Geriatr Psychiatry, 19*(3), 278-285.
- WHOQOL-HIV for quality of life assessment among people living with HIV and AIDS: results from the field test. (2004). *AIDS Care, 16*(7), 882-889.
- Wilms, H. U., Kanowski, S., & Baltes, M. M. (2000). Limitations in activities of daily living: towards a better understanding of subthreshold mental disorders in old age. *Compr Psychiatry, 41*(2 Suppl 1), 19-25.
- Woo, J., Ho, S. C., & Wong, E. M. (2005). Depression is the predominant factor contributing to morale as measured by the Philadelphia Geriatric Morale Scale in elderly Chinese aged 70 years and over. *Int J Geriatr Psychiatry, 20*(11), 1052-1059.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res, 17*(1), 37-49.
- Yun, R. J., & Lachman, M. E. (2006). Perceptions of aging in two cultures: Korean and American views on old age. *J Cross Cult Gerontol, 21*(1-2), 55-70.

11.2 Termo de consentimento

CONVITE À PARTICIPAÇÃO DE UM ESTUDO SOBRE QUALIDADE DE VIDA EM IDOSOS

Estamos realizando um trabalho de pesquisa sobre como é que as pessoas acima de 60 anos acham que está a sua vida nas últimas duas semanas. Para isto, gostaríamos de contar com a sua colaboração durante alguns minutos para responder a um questionário. Serão feitas várias perguntas sobre diferentes aspectos de sua vida: sua saúde física, sua vida emocional, sua relação com amigos e familiares, seu meio-ambiente.

Esta pesquisa está sendo feita em conjunto com mais 20 países em todo mundo, e é coordenada pelas Organização Mundial de Saúde. Em Porto Alegre, o grupo coordenador pertence ao Hospital de Clínicas de Porto Alegre e tem como responsável o Dr. Marcelo Fleck.

Gostaríamos de deixar claro que o(a) senhor(a) deve ter toda a liberdade para interromper o questionário se desejar ou negar-se a responder, caso o(a) senhor(a) não esteja de acordo em participar. Asseguramos que todas as informações prestadas pelo senhor (a) são sigilosas e serão utilizadas somente para esta pesquisa. A divulgação das informações serão anônimas e em conjunto com as respostas de um grupo de 300 pessoas.

Para algumas pessoas que concordarem, estaremos fazendo novo questionário no prazo de duas semanas da data do primeiro encontro, com o objetivo de melhor avaliar os instrumentos deste estudo.

Se você tiver alguma pergunta a fazer antes de decidir, sinta-se a vontade para fazê-la.

Data: ____/____/2004

Nome do entrevistado: _____

Assinatura: _____

Nome do entrevistador: _____

Assinatura: _____

11.3 Ficha de datos sócio-demográficos

Nº _____

INFORMAÇÕES SÓCIODEMOGRÁFICAS

Método de administração:

1. Por E-mail 3. Feita pelo entrevistador 5. Outro: _____
 2. Auto-administrada 4. Feita pelo telefone _____

1 Nome: _____
 1 Sexo: 1. Masculino 2. Feminino

2 Idade: _____

3 Estado civil atual:

1. Solteiro (nunca casou)
 2. Casado
 3. Vivendo maritalmente
 4. Separado/Divorciado (sem companheiro atual)
 5. Viúvo

4 Mais alto nível de escolaridade alcançado:

1. Ensino fundamental
 2. Ensino médio
 3. Curso técnico
 4. Curso profissionalizante
 5. Curso superior
 6. Analfabeto (incapaz de ler ou escrever)
 7. Outra: Por favor, especifique

5 Situação de moradia:

1. Morando na sua casa (apoiado pela família, cuidador ou companheiro)
 2. Morando com sua família, mas não na sua própria casa
 3. Morando na sua casa (sem apoio da família/cuidador ou companheiro)
 4. Morando em um lar de idosos
 5. Morando em um abrigo protegido/cuidado pela comunidade
 6. Morando num lar com cuidados de enfermagem
 7. Morando numa enfermaria de longa permanência (hospital)
 8. Outro (por favor, dê detalhes abaixo)

--

- 6 Qual era a sua ocupação anterior?
- 1. **Profissional graduado ou Administrador** (por ex., médico, professor, engenheiro, artista, contador, executivo, funcionário público graduado, etc.)
 - 2. **Vendas e Serviços** (por ex., gerente de vendas, proprietário de loja, vendedor, corretor de seguros, policial, garçom, zelador, cabeleireiro, etc.)
 - 3. **Trabalhador técnico e Auxiliar** (por ex., mecânico, auxiliar de gráfica, costureira, eletricitista, secretária, auxiliar de escritório, técnico de contabilidade, gerente de escritório, etc.)
 - 4. **Trabalhador semi ou Não qualificado** (por ex., trabalhador de construção civil, motorista de ônibus, operário, padeiro, porteiro, etc.)
 - 5. **Fazenda** (por ex., fazendeiro ou trabalhador rural, etc.)
 - 6. **Militar**
 - 7. **Afazeres domésticos**
 - 8. **Trabalho não remunerado**
 - 9. **Outro: Por favor, especifique**
-
- 7 Por favor, assinale **qualquer** das alternativas abaixo que melhor se aplique a você.
- 1. Empregado em tempo integral
 - 2. Empregado em tempo parcial (um turno)
 - 3. Empregado eventualmente
 - 4. Trabalhador autônomo
 - 5. Desempregado, procurando trabalho
 - 6. Aposentado
 - 7. Permanentemente incapacitado para o trabalho
 - 8. Estudante
 - 9. Afazeres domésticos
 - 10. Trabalho voluntário
- 8 Se você está empregado atualmente, qual é a sua ocupação no momento?
- 1. **Profissional graduado ou Administrador** (por ex., médico, professor, engenheiro, artista, contador, executivo, funcionário público graduado, etc.)

- 2. **Vendas e Serviços** (por ex., gerente de vendas, proprietário de loja, vendedor, corretor de seguros, policial, garçoneiro, zelador, cabeleireiro, etc.)
 - 3. **Trabalhador técnico e Auxiliar** (por ex., mecânico, auxiliar de gráfica, costureira, eletricitista, secretária, auxiliar de escritório, técnico de contabilidade, gerente de escritório, etc.)
 - 4. **Trabalhador semi ou Não qualificado** (por ex., trabalhador de construção civil, motorista de ônibus, operário, padeiro, porteiro, etc.)
 - 5. **Fazenda** (por ex., fazendeiro ou trabalhador rural, etc.)
 - 6. **Militar**
 - 7. **Afazeres domésticos**
 - 8. **Trabalho não remunerado**
 - 9. **Outro: Por favor, especifique**
-

- 9 Em comparação a outras pessoas da sua idade, como você classificaria a sua situação financeira e bens (posses)?
- 1. Bem acima da média
 - 2. Levemente acima da média
 - 3. Na média
 - 4. Levemente abaixo da média
 - 5. Bem abaixo da média
- 10 Sobre seu relacionamento com o seu companheiro:
- 1. Eu não tenho companheiro
 - 2. Eu não tenho relacionamento íntimo e próximo
 - 3. Raramente íntimo e próximo
 - 4. Algumas vezes íntimo e próximo
 - 5. Muito íntimo e próximo
- 11 Quantos filhos você tem? _____
- 12 Quão satisfeito você está com o seu relacionamento com os seus filhos?
- 1. Muito satisfeito
 - 2. Satisfeito
 - 3. Nem satisfeito e nem insatisfeito
 - 4. Insatisfeito
 - 5. Muito insatisfeito
 - 6. Não se aplica

13 Quantos netos você tem? _____

14 Quanto satisfeito você está com o seu relacionamento com os seus netos?

- 1.Muito satisfeito
 2.Satisfeito
 3.Nem satisfeito e nem insatisfeito
 4.Insatisfeito
 5.Muito insatisfeito
 6.Não se aplica

15 Você se considera, de um modo geral, uma pessoa saudável ou doente?

- 1.Saudável
 2.Doente

16 Por favor, dê detalhes de qualquer problema de saúde que você tenha que possa interferir com sua qualidade de vida.

--

17a Caso você tenha algum problema de saúde, você usa medicação em função dele?

- 1.Sim
 2.Não
 3.Nenhum problema de saúde devido à medicação
 4.Não se aplica

17b Se sim, por favor, dê detalhes das medicações que você usa:

Medicação	Para quê usa?
1	
2	
3	
4	
5	
6	
7	

18a Você usa algum remédio complementar (por ex., vitaminas, ou outras substâncias não prescritas por um médico)?

- 1.Sim 2.Não

18b Se sim, por favor dê detalhes das substâncias que você usa e quanto frequentemente você as usa.

19 Você fuma cigarros?

- 1.Fumante
 - 1a.Quantos por semana:_____
- 2.Ex-fumante
 - 2a.Há quanto tempo você deixou de fumar?
____anos ____ meses
- 3.Nunca fumou

20 Quão freqüentemente você consome bebida alcoólica?

- 1.Diariamente
- 2.Muitas vezes na semana
- 3.Aproximadamente uma vez por semana
- 4.Aproximadamente duas vezes por mês
- 5.Uma vez por mês ou menos
- 6.Nunca

21 Quantos copos de bebida alcoólica consome nos dias em que bebe?

- 1) 1 ou 2
- 2) 3 ou 4
- 3) 5 ou 6
- 4) 7 a 9
- 5) 10 ou mais
- 6) Não se aplica

22 Condição:

- 1.Internação 2.Ambulatório 3.Comunitário

O Sr.(a) poderia nos fornecer um telefone para contato? _____

Obrigado(a) por completar este formulário!

11.4 Escala de Avaliação Sócio-Demográfica

ESCALA PARA NÍVEL SOCIOECONÔMICO

A Associação Brasileira de Institutos de Pesquisa de Mercado criou, em 1978, um novo sistema de classificação socioeconômica. O conceito básico desta classificação é discriminar as pessoas sócio-economicamente mediante informações sobre sua escolaridade e a posse de determinados 'itens de conforto', tais como televisor, geladeira, rádio, automóvel e empregados domésticos. É levado em consideração o número de entidades possuídas, item por item ao invés de simplesmente atribuírem-se pontos conforme a presença ou ausência de cada item. A soma dos pontos obtidos vai incluir a pessoa entrevistada nas classes A, B, C, D e E, conforme mostrado a seguir.

Critério

Item	Não tem	1	2	3	4	5	6 ou mais
TV	0	2	4	6	8	10	12
Rádio	0	1	2	3	4	5	6
Banheiro	0	2	4	6	8	10	12
Carro	0	4	8	12	16	16	16
Empregada	0	6	12	18	24	24	24
Telefone	0	5	5	5	5	5	5
Geladeira	0	2	2	2	2	2	2

Instrução do chefe da família	Pontos
Analfabeto/Primário incompleto	0
Primário completo/Ginasial incompleto	1
Ginasial completo/Colegial incompleto	3
Colegial completo/Superior incompleto	5
Superior completo	10

Por exemplo:

O sujeito X possui 1 televisão, 3 rádios, 1 automóvel, 1 telefone e 1 geladeira.

X tem nível superior incompleto. Assim, X tem a seguinte pontuação:

$$2 + 3 + 4 + 5 + 2 + 5 = 21$$

Com isto, X é classificado na classe B.

Classe	Pontos
A	35 ou mais
B	21 a 34
C	10 a 20
D	5 a 9
E	0 a 4

Obrigado(a)!

13.5 Instrumento WHOQOL-BREF

WHOQOL - ABREVIADO

Versão em Português

PROGRAMA DE SAÚDE MENTAL ORGANIZAÇÃO MUNDIAL DA SAÚDE GENEBRA

Somente para uso da coordenação

	Equações para computação dos escores dos domínios	Escore Bruto	Escore Transformado *	
			4-20	0-100
Domínio 1	$(6-Q3) + (6-Q4) + Q10 + Q15 + Q16 + Q17 + Q18$ ____ + ____ + ____ + ____ + ____ + ____ + ____	=		
Domínio 2	$Q5 + Q6 + Q7 + Q11 + Q19 + (6-Q26)$ __ + __ + __ + ____ + ____ + ____	=		
Domínio 3	$Q20 + Q21 + Q22$ __ + __ + __	=		
Domínio 4	$Q8 + Q9 + Q12 + Q13 + Q14 + Q23 + Q24 + Q25$ __ + __ + __ + __ + __ + __ + __	=		

* Ver Tabela 4 da página 10 do manual, para converter escores brutos em escores transformados

Instruções

Este questionário é sobre como você se sente a respeito de sua qualidade de vida, saúde e outras áreas de sua vida. **Por favor, responda a todas as questões.** Se você não tem certeza sobre que resposta dar em uma questão, por favor, escolha entre as alternativas a que lhe parece mais apropriada. Esta, muitas vezes, poderá ser sua primeira escolha.

Por favor, tenha em mente seus valores, aspirações, prazeres e preocupações. Nós estamos perguntando o que você acha de sua vida, tomando como referência as **duas últimas semanas**. Por exemplo, pensando nas últimas duas semanas, uma questão poderia ser:

	nada	muito pouco	médio	muito	completamente
Você recebe dos outros o apoio de que necessita?	1	2	3	4	5

Você deve circular o número que melhor corresponde ao quanto você recebe dos outros o apoio de que necessita nestas últimas duas semanas. Portanto, você deve circular o número 4 se você recebeu "muito" apoio como abaixo.

	nada	muito pouco	médio	muito	completamente
Você recebe dos outros o apoio de que necessita?	1	2	3	4	5

Você deve circular o número 1 se você não recebeu "nada" de apoio.

Por favor, leia cada questão, veja o que você acha e circule no número e lhe parece a melhor resposta.

		muito ruim	Ruim	nem ruim nem boa	boa	muito boa
1(G1)	Como você avaliaria sua qualidade de vida?	1	2	3	4	5

		muito insatisfeito	Insatisfeito	nem satisfeito nem insatisfeito	satisfeito	muito satisfeito
2(G4)	Quão satisfeito(a) você está com a sua saúde?	1	2	3	4	5

As questões seguintes são sobre **o quanto** você tem sentido algumas coisas nas últimas duas semanas.

		nada	muito pouco	mais ou menos	bastante	extremamente
3(F1.4)	Em que medida você acha que sua dor (física) impede você de fazer o que você precisa?	1	2	3	4	5
4(F11.3)	O quanto você precisa de algum tratamento médico para levar sua vida diária?	1	2	3	4	5
5(F4.1)	O quanto você aproveita a vida?	1	2	3	4	5
6(F24.2)	Em que medida você acha que a sua vida tem sentido?	1	2	3	4	5
7(F5.3)	O quanto você consegue se concentrar?	1	2	3	4	5
8(F16.1)	Quão seguro(a) você se sente em sua vida diária?	1	2	3	4	5
9(F22.1)	Quão saudável é o seu ambiente físico (clima, barulho, poluição, atrativos)?	1	2	3	4	5

As questões seguintes perguntam sobre **Quão completamente** você tem sentido ou J capaz de fazer certas coisas nestas últimas duas semanas.

		nada	muito pouco	médio	muito	completamente
10(F2.1)	Você tem energia suficiente para seu dia-a-dia?	1	2	3	4	5
11(F7.1)	Você J capaz de aceitar sua aparência física?	1	2	3	4	5
12(F18.1)	Você tem dinheiro suficiente para satisfazer suas necessidades?	1	2	3	4	5
13(F20.1)	Quão disponíveis para você estão as informações que precisa no seu dia-a-dia?	1	2	3	4	5
14(F21.1)	Em que medida você tem oportunidades de atividade de lazer?	1	2	3	4	5

As questões seguintes perguntam sobre **Quão bem ou satisfeito** você se sentiu a respeito de vários aspectos de sua vida nas últimas duas semanas.

		muito ruim	ruim	nem ruim	bom	muito bom
15(F9.1)	Quão bem você é capaz de se locomover?	1	2	3	4	5

		muito insatisfeito	Insatisfeito	nem satisfeito nem insatisfeito	satisfeito	muito satisfeito
16(F3.3)	Quão satisfeito(a) você está com o seu sono?	1	2	3	4	5
17(F10.3)	Quão satisfeito(a) você está com sua capacidade de desempenhar as atividades do seu dia-a-dia?	1	2	3	4	5
18(F12.4)	Quão satisfeito(a) você está com sua capacidade para o trabalho?	1	2	3	4	5
19(F6.3)	Quão satisfeito(a) você está consigo mesmo?	1	2	3	4	5
20(F13.3)	Quão satisfeito(a) você está com suas relações pessoais (amigos, parentes, conhecidos, colegas)?	1	2	3	4	5
21(F15.3)	Quão satisfeito(a) você está com sua vida sexual?	1	2	3	4	5
22(F14.4)	Quão satisfeito(a) você está com o apoio que você recebe de seus amigos?	1	2	3	4	5
23(F17.3)	Quão satisfeito(a) você está com as condições do local onde mora?	1	2	3	4	5
24(F19.3)	Quão satisfeito(a) você está com o seu acesso aos serviços de saúde?	1	2	3	4	5
25(F23.3)	Quão satisfeito(a) você está com o seu meio de transporte?	1	2	3	4	5

As questões seguintes referem-se a **com que frequência** você sentiu ou experimentou certas coisas nas últimas duas semanas.

		nunca	Algumas vezes	freqüentemente	muito freqüentemente	sempre
26(F8.1)	Com que frequência você tem sentimentos negativos tais como mau humor, desespero, ansiedade, depressão?	1	2	3	4	5

Alguém lhe ajudou a preencher este questionário?.....

Quanto tempo você levou para preencher este questionário?.....

OBRIGADO PELA SUA COLABORAÇÃO

13.6 Módulo WHOQOL-OLD



WHOQOL-OLD

Instruções

Este questionário pergunta a respeito dos seus pensamentos, sentimentos e sobre certos aspectos de sua qualidade de vida, e aborda questões que podem ser importantes para você como membro mais velho da sociedade.

Por favor, responda todas as perguntas. Se você não está seguro a respeito de que resposta dar a uma pergunta, por favor escolha a que lhe parece mais apropriada. Esta pode ser muitas vezes a sua primeira resposta.

Por favor tenha em mente os seus valores, esperanças, prazeres e preocupações. Pedimos que pense na sua vida nas duas últimas semanas.

Por exemplo, pensando nas duas últimas semanas, uma pergunta poderia ser :

O quanto você se preocupa com o que o futuro poderá trazer?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

Você deve circular o número que melhor reflete o quanto você se preocupou com o seu futuro durante as duas últimas semanas. Então você circularia o número 4 se você se preocupou com o futuro “Bastante”, ou circularia o número 1 se não tivesse se preocupado “Nada” com o futuro. Por favor leia cada questão, pense no que sente e circule o número na escala que seja a melhor resposta para você para cada questão.

Muito obrigado(a) pela sua colaboração!

As seguintes questões perguntam sobre o quanto você tem tido certos sentimentos nas últimas duas semanas.

F25.1 Até que ponto as perdas nos seus sentidos (por exemplo, audição, visão, paladar, olfato, tato), afetam a sua vida diária?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F25.3 Até que ponto a perda de, por exemplo, audição, visão, paladar, olfato, tato, afeta a sua capacidade de participar em atividades?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F26.1 Quanta liberdade você tem de tomar as suas próprias decisões?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F26.2 Até que ponto você sente que controla o seu futuro?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F26.4 O quanto você sente que as pessoas ao seu redor respeitam a sua liberdade?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F29.2 Quão preocupado você está com a maneira pela qual irá morrer?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F29.3 O quanto você tem medo de não poder controlar a sua morte?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F29.4 O quanto você tem medo de morrer?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F29.5 O quanto você teme sofrer dor antes de morrer?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

As seguintes questões perguntam sobre quão completamente você fez ou se sentiu apto a fazer algumas coisas nas duas últimas semanas.

F25.4 Até que ponto o funcionamento dos seus sentidos (por exemplo, audição, visão, paladar, olfato, tato) afeta a sua capacidade de interagir com outras pessoas?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

F26.3 Até que ponto você consegue fazer as coisas que gostaria de fazer?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

F27.3 Até que ponto você está satisfeito com as suas oportunidades para continuar alcançando outras realizações na sua vida?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

F27.4 O quanto você sente que recebeu o reconhecimento que merece na sua vida?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

F28.4 Até que ponto você sente que tem o suficiente para fazer em cada dia?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

As seguintes questões pedem a você que diga o quanto você se sentiu satisfeito, feliz ou bem sobre vários aspectos de sua vida nas duas últimas semanas.

F27.5 Quão satisfeito você está com aquilo que alcançou na sua vida?

Muito insatisfeito	Insatisfeito	Nem satisfeito nem insatisfeito	Satisfeito	Muito satisfeito
1	2	3	4	5

F28.1 Quão satisfeito você está com a maneira com a qual você usa o seu tempo?

Muito insatisfeito	Insatisfeito	Nem satisfeito nem insatisfeito	Satisfeito	Muito satisfeito
1	2	3	4	5

F28.2 Quão satisfeito você está com o seu nível de atividade?

Muito insatisfeito	Insatisfeito	Nem satisfeito nem insatisfeito	Satisfeito	Muito satisfeito
1	2	3	4	5

F28.7 Quão satisfeito você está com as oportunidades que você tem para participar de atividades da comunidade?

Muito insatisfeito	Insatisfeito	Nem satisfeito nem insatisfeito	Satisfeito	Muito satisfeito
1	2	3	4	5

F27.1 Quão feliz você está com as coisas que você pode esperar daqui para frente?

Muito infeliz	Infeliz	Nem feliz nem infeliz	Feliz	Muito feliz
1	2	3	4	5

F25.2 Como você avaliaria o funcionamento dos seus sentidos (por exemplo, audição, visão, paladar, olfato, tato)?

Muito ruim	Ruim	Nem ruim nem boa	Boa	Muito boa
1	2	3	4	5

As seguintes questões se referem a qualquer relacionamento íntimo que você possa ter. Por favor, considere estas questões em relação a um companheiro ou uma pessoa próxima com a qual você pode compartilhar (dividir) sua intimidade mais do que com qualquer outra pessoa em sua vida.

F30.2 Até que ponto você tem um sentimento de companheirismo em sua vida?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F30.3 Até que ponto você sente amor em sua vida?

Nada	Muito pouco	Mais ou menos	Bastante	Extremamente
1	2	3	4	5

F30.4 Até que ponto você tem oportunidades para amar?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

F30.7 Até que ponto você tem oportunidades para ser amado?

Nada	Muito pouco	Médio	Muito	Completamente
1	2	3	4	5

Você tem algum comentário sobre o questionário?

OBRIGADO(A) PELA SUA COLABORAÇÃO!

13.7 Geriatric Depression Scale (GDS) – Versão em Português

ESCALA DE DEPRESSÃO GERIÁTRICA (GDS)

Nome: _____

Data de avaliação: _____ Avaliador(a): _____

Gds1. Você está em geral satisfeito(a) com a sua vida?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds2. Você abandonou muitas das suas atividades e interesses?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds3. Você sente que sua vida tem estado vazia?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds4. Você se aborrece freqüentemente?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds5. Você está otimista a respeito do futuro?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds6. Você se sente perturbado por pensamentos que não saem da sua cabeça?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds7. Você se sente bem humorado(a) a maior parte do tempo?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds8. Você tem a sensação de que algo ruim está prestes a acontecer para você?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds9. Você se sente feliz a maior parte do tempo?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds10. Você se sente desamparado com freqüência?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds11. Você se sente tenso(a) e inquieto(a)?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds12. Você prefere ficar em casa a sair e fazer coisas novas?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds13. Você tem se preocupado a respeito do futuro?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds14. Você tem tido mais problemas de memória do que as outras pessoas?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds15. Você acha maravilhoso estar vivo(a) agora?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds16. Você tem se sentido "para baixo" e triste freqüentemente?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds17. Você anda se sentindo sem valor (inútil) atualmente?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds18. Você se preocupa muito com o passado?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds19. Você sente que a vida é excitante?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds20. Tem sido difícil para você iniciar novos planos?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds21. Você se sente cheio(a) de energia?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds22. Você se sente sem esperanças de mudar a sua situação?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds23. Você sente que a maioria das pessoas estão melhores do que você?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds24. Você tem se aborrecido com pequenas coisas freqüentemente?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds25. Você tem sentido vontade de chorar freqüentemente?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds26. Você tem tido dificuldade para se concentrar?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds27. Você sente prazer ao acordar, pela manhã?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds28. Você prefere evitar encontros sociais?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds29. Você tem tido facilidade para tomar decisões?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Gds30. A sua mente está tão clara quanto antes?	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
ESCORE	(/30)	

CONCLUSÃO:

* Material elaborado para fins de pesquisa.

The Geriatric Scale Questionnaire.

Yesavage JA, Brink TL, Lum O, Huang V, Adey MB, Leirer VO: Development and validation of a geriatric depression scale: preliminary report. J. Psychiatric 17:37-39, 1993

Copyright 1981, J. Yesavage, T. Brink

13.8 Atitudes Frente ao Envelhecimento (AAQ)

Atitudes Frente ao Envelhecimento AAQ

Instruções

Este questionário pergunta a respeito de como você se sente em relação a envelhecer.

Por favor, responda todas as perguntas. Se você não está seguro a respeito de que resposta dar a uma pergunta, por favor escolha a que lhe parece mais apropriada. Esta pode ser muitas vezes a sua primeira resposta.

Por favor tenha em mente seus valores, esperanças, prazeres e preocupações. Pedimos que pense na sua vida em geral.

Por exemplo, pensando em como você se sente de um modo geral, uma pergunta poderia ser:

Eu não gosto de envelhecer

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

Você deve circular o número que melhor descreva o quanto você concorda com a afirmação. Então você circularia o número 4 se você não gosta de envelhecer "Bastante", ou circularia o número 1 se você está "Nada" preocupado sobre envelhecer. Por favor, leia cada questão, avalie os seus sentimentos e circule o número na escala que seja a melhor resposta para cada questão.

Muito obrigado pela sua colaboração!

As seguintes questões perguntam sobre o quanto você concorda com as seguintes afirmações. Se você concorda com as questões extremamente circule o número próximo de "Extremamente". Se você não concorda com as afirmações de jeito nenhum circule o número próximo de "Nada". Você deve circular um dos números entre esses, se você deseja indicar que a sua resposta está entre "Nada" e "Extremamente". As perguntas se referem aos idosos de um modo geral.

1. As pessoas são tão velhas quanto se sentem (A idade de uma pessoa não significa nada).

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

2. A medida em que as pessoas envelhecem elas estão mais capazes para lidar com a vida.

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

3. A velhice é um tempo de doença.

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

4. É um privilégio envelhecer.

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

5. As pessoas mais velhas são interessadas em novas tecnologias tais como computadores e a internet.

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

6. As pessoas mais velhas são interessadas em amor.

Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

7. A velhice é um tempo de solidão.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5
8. A sabedoria vem com a idade.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5
9. Há muitas coisas prazerosas no envelhecer.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5
10. A velhice é uma época depressiva da vida.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5
11. As habilidades e capacidades de uma pessoa diminuem com a idade.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5
12. É importante fazer exercício em qualquer idade.				
Discordo	Discordo	Não concordo	Concordo	Concordo
totalmente	2	nem discordo	4	totalmente
1		3		5

As seguintes questões perguntam sobre o quão verdadeiras as seguintes afirmações são para você. Se a afirmação é "Extremamente" verdadeira para você, circule o número próximo de "Extremamente". Se as afirmações não são nem um pouco verdadeiras para você, circule o número próximo de "Nada". Você deve circular um dos números entre esses, se você deseja indicar que a sua resposta está entre "Nada" e "Extremamente".

13. Envelhecer tem sido mais fácil do que pensava.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

14. Eu encontro mais dificuldade para falar sobre os meus sentimentos na medida em que envelheço.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

15. Eu me aceito mais agora na medida em que fui envelhecendo.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

16. Eu não me sinto velho.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

17. Eu vejo a velhice principalmente como um tempo de perda.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

18. Minhas crenças pessoais/espiritualidade tornam-se mais importantes para mim na medida em que fico mais velho.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

19. MINHA IDENTIDADE NÃO É DEFINIDA PELA MINHA IDADE.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

20. Eu tenho mais energia agora do que eu esperava para a minha idade.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

21. Eu estou perdendo minha independência física na medida em que fico mais velho.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

22. Problemas com minha saúde física não me impedem que eu faça o que quero.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

23. Eu estou infeliz com as mudanças em minha aparência física na medida em que fui envelhecendo.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

24. Na medida em que eu me torno mais velho, eu encontro mais dificuldades para fazer novos amigos.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

25. É muito importante passar os benefícios da minha experiência para os mais jovens.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

26. Eu temo perder minha independência financeira na medida em que fico mais velho.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

27. Agora eu tenho tempo para fazer coisas que realmente me interessam.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

28. Eu quero continuar fazendo algum tipo de trabalho (voluntário ou remunerado), até quando for possível.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

29. Eu estou preocupado em me tornar um fardo (um peso) financeiro para minha família.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

30. Eu acredito que a minha vida foi relevante.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

31. Minha vida tem tanto significado agora quanto ela sempre teve.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

32. Eu não me sinto envolvido na sociedade agora que estou mais velho.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

33. Eu quero dar um bom exemplo para as pessoas mais jovens.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

34. Eu me sinto excluído das coisas em função da minha idade.

Nada verdadeiro 1	Muito pouco verdadeiro 2	Mais ou menos verdadeiro 3	Bastante verdadeiro 4	Extremamente verdadeiro 5
-------------------------	--------------------------------	----------------------------------	-----------------------------	---------------------------------

35. O meu futuro me assusta.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

36. Minha saúde está melhor do que eu esperava para a minha idade.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

37. Eu me mantenho tão em forma e ativo quanto possível, fazendo exercícios.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

38. Na medida em que eu envelheço, meus relacionamentos mais importantes têm se tornado mais satisfatórios e íntimos.

Nada verdadeiro	Muito pouco verdadeiro	Mais ou menos verdadeiro	Bastante verdadeiro	Extremamente verdadeiro
1	2	3	4	5

OBRIGADO(A) PELA SUA COLABORAÇÃO!

13.9 Termo de Aprovação da Comissão Científica e Comissão de Pesquisa e Ética em Saúde do Hospital de Clínicas de Porto Alegre



HCPA - HOSPITAL DE CLÍNICAS DE PORTO ALEGRE
Grupo de Pesquisa e Pós-Graduação

COMISSÃO CIENTÍFICA E COMISSÃO DE PESQUISA E ÉTICA EM SAÚDE

RESOLUÇÃO

A Comissão Científica e a Comissão de Pesquisa e Ética em Saúde, que é reconhecida pela Comissão Nacional de Ética em Pesquisa (CONEP)/MS como Comitê de Ética em Pesquisa do HCPA e pelo Office For Human Research Protections (OHRP)/USDHHS, como Institutional Review Board (IRB0000921) analisaram o projeto:

Projeto: 01-374

Pesquisadores:

MARCELO PIO DE ALMEIDA FRECK

EDUARDO CHACHAMOVICH

CLARISSA TRENTIN

Título: DESENVOLVIMENTO DE UMA MEDIDA DE AVALIAÇÃO DE QUALIDADE DE VIDA PARA IDOSOS E SUA RELAÇÃO COM ENVELHECIMENTO SAUDÁVEL

Este projeto foi Aprovado em seus aspectos éticos e metodológicos, inclusive quanto ao seu Termo de Consentimento Livre e Esclarecido, de acordo com as Diretrizes e Normas Internacionais e Nacionais, especialmente as Resoluções 196/96 e complementares do Conselho Nacional de Saúde. Os membros do CEP/HCPA não participaram do processo de avaliação dos projetos onde constam como pesquisadores. Toda e qualquer alteração do Projeto, assim como os eventos adversos graves, deverão ser comunicados imediatamente ao CEP/HCPA.

Porto Alegre, 11 de janeiro de 2002.

Profa. Themis Reverbél da Silveira
Coordenadora do GPPG e CEP-HCPA