

Outcome Questionnaire (OQ-45.2): avaliação das propriedades psicométricas via modelo bifactor e TRI*

Sonia Maria da Silva

Universidade Guarulhos, SP, Brasil

Iraí Cristina Boccato Alves

Instituto de Psicologia USP, SP, Brasil

Evandro Moraes Peixoto

Universidade São Francisco, SP, Brasil

Glauca Mitsuko Ataka Rocha

Instituto de Psicologia USP, SP, Brasil

Tatiana de Cássia Nakano

Pontifícia Universidade Católica de Campinas, SP, Brasil

Resumo

O objetivo foi avaliar as propriedades psicométricas da versão brasileira do *Outcome Questionnaire (OQ-45.2)*, instrumento designado a avaliar o progresso de pacientes em psicoterapia. Comparou-se o ajuste de diferentes modelos de medidas propostos ao *OQ-45.2* através da Análise Fatorial Confirmatória e verificou-se os parâmetros dos itens e dos participantes por meio do Modelo de Resposta Gradual. 419 adultos responderam ao instrumento (32,18±14,3; 62,8%, mulheres). Os resultados demonstram a adequação da estrutura Bifactor composta por três fatores específicos (desconforto subjetivo, relações interpessoais e desempenho do papel social) e um fator geral (desajustamento global) quando comparados aos outros modelos: unifatorial e com três fatores correlacionados. Verificou-se a invariância da estrutura interna do *OQ-45.2* ao avaliar homens e mulheres. Em relação às propriedades dos itens (dificuldade e ajustamento) foram obtidos parâmetros psicométricos apropriados para a avaliação de resultados psicoterapêuticos. Conclui-se que o *OQ-45.2* é uma medida adequada para avaliação destas características na população brasileira.

Palavras-chave: Teoria de Resposta ao Item; Análise fatorial; Avaliação psicológica; Validade do teste; Psicoterapia.

Outcome Questionnaire (OQ-45.2): assessment of the psychometric properties using bifactor model and IRT

Abstract

The aim of this study was to evaluate the Psychometric properties of the Brazilian version of the Outcome Questionnaire (OQ-45.2), an instrument that assesses the progress of patients in psychotherapy. The adjustments of different measure models proposed to OQ-45.2 were compared through Confirmatory Factor Analysis, and the parameters of items and participants were estimated through the Andrich's rating scale model. The sample comprised 419 adults (mean age: 32.18±14.3; 62.8% women). The results demonstrated the suitability of the bifactor model with three specific factors (subjective discomfort, interpersonal relationships and performance in social roles) and a general factor (overall maladjustment) when compared to others in the literature, like the one-factor and the three related factors models. It was observed a model invariance of the OQ-45 internal structure for both men and women assessment. Regarding the items' properties (adjustment and difficulty) were obtained the appropriate psychometric parameters for the assessment of psychotherapeutic outcome. In conclusion, the OQ-45 is an appropriate measure tool for assessing these characteristics in Brazilian population.

Keywords: Item Response Theory; Factorial analysis; Psychological assessment; Test validity; Psychotherapy.

Outcome Questionnaire (OQ-45.2): evaluación de las propiedades psicométricas via modelo bifactor y TRI

Resumen

El objetivo fue evaluar las propiedades psicométricas de la versión brasileña del *Outcome Questionnaire (OQ-45.2)*, instrumento diseñado para evaluar el progreso de los pacientes en psicoterapia. Ha sido comparado el ajuste de los diferentes modelos de medidas para *OQ-45.2* mediante análisis factorial confirmatorio y ha sido encontrado los parámetros de los ítems y de los participantes a través del Modelo de Respuesta Gradual. 419 adultos respondieron el instrumento (32,18±14,3; 62,8% mujeres). Los resultados demuestran la adecuación del modelo *bifactor* formado por tres factores específicos (el malestar subjetivo, las relaciones interpersonales y el desempeño del papel social) y un general (desajuste global) cuando se compara con otros modelos de la literatura: un factor de tres y factores relacionados. Se observó la invariación de la estructura interna del *OQ-45.2* cuando se evaluaron personas del sexo femenino y masculino. Con respecto a las propiedades de los elementos (ajuste de dificultad) se obtuvieron parámetros psicométricos apropiados para la evaluación de los resultados psicoterapêuticos. Se concluyó que el instrumento presenta buenas propiedades psicométricas para evaluación de la población brasileña.

Palabras clave: Teoría de Respuesta al Item; Análisis factorial; Evaluación psicológica; Validación de test; Psicoterapia.

* Ver ANEXO versão em inglês deste texto disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico>>.

Nos últimos anos observa-se no Brasil um aumento de pesquisas de processo e resultados de psicoterapia (Brum et al., 2012; Del Prette, 2011; Pieta, Siegmund, Gomes, & Gauer, 2015), especialmente na clínica-escola, o que leva a necessidade de atender a três objetivos: prestação de serviço à comunidade, formação profissional dos alunos do curso de Psicologia e o desenvolvimento de pesquisas. As demandas dos usuários exigem um constante “pensar” e “fazer” uma Psicologia que atenda às necessidades destes, aliando teoria e prática por meio da produção de conhecimento, especialmente o desenvolvimento e/ou adaptação de instrumentos de avaliação adequados à população atendida e que possibilitem mensurar a mudança em processos psicoterapêuticos (Honda, Peixoto, Rocha, & Enéas, 2015).

Há a preocupação em se buscar práticas avaliativas que tenham base em evidências empíricas (Antony & Barlow, 2010). Propriedades psicométricas, tais como validade e precisão, passaram a ser tema de discussões, especialmente com a publicação de Resolução do Conselho Federal de Psicologia (CFP) nº 002/2003, que considera a necessidade de aprimorar os instrumentos e procedimentos técnicos do trabalho dos psicólogos bem como de fazer a revisão periódica dos mesmos.

Dentro da gama de instrumentos de avaliação psicológica, aqueles chamados de autorrelato têm recebido destaque, pois, por serem respondidos pelo próprio indivíduo, permitem aferir a maneira pela qual ele relata a percepção sobre o próprio estado. Esses instrumentos vêm sendo cada vez mais usados na área de avaliação de processo e resultados de psicoterapia, uma vez que essas medidas parecem contribuir de modo mais significativo para determinar mudanças no paciente (Harmon, Hawkins, Lambert, Slade, & Whipple, 2005). São escalas simples e que acarretam menos esforços e que conseguem aproximar a pesquisa da prática clínica (Lambert, Hansen, & Finch, 2001).

Serralta, Nunes e Eizirik (2007), ressaltaram que, no Brasil, havia pouca sistematização quanto à utilização de instrumentos para avaliar os processos psicoterapêuticos, assim como em função das formas de atendimento e aos resultados de psicoterapias. Os mesmos autores acrescentam que a pesquisa em psicoterapia se encontra em estado inicial, se comparada à pesquisa no contexto internacional, e ainda faltam estudos sistemáticos e em profundidade sobre esse tema. A fim de contribuir para o preenchimento desta lacuna esta pesquisa teve como objetivo principal estimar evidências de validade com base na estrutura interna e precisão da Escala de Avaliação de Resultados – *Outcome Questionnaire (OQ-45.2)* (Lambert et al., 2004) para a população brasileira.

O *OQ-45.2* é um dos instrumentos mais populares entre os profissionais norte-americanos que objetivam a avaliação de resultado de processos psicoterapêuticos (Hatfield & Ogles, 2004). Tal popularidade também pode ser observada pela quantidade de idiomas para os quais o instrumento foi traduzido: japonês, coreano, francês, italiano, alemão, holandês, português (Portugal), espanhol, sueco, norueguês e russo (Huag, Puschner, Lambert, & Kordy, 2004; Jong et al., 2007; Lambert et al., 2004, Machado & Fassnacht, 2015).

Desenvolvido por Lambert, Lunnen, Umphress, Hansen e Burlingame em 1994 (Lambert & Finch, 1999) o *OQ-45* foi concebido para o uso em três níveis distintos de aplicação: 1) mensurar a dimensão do sofrimento psíquico atual; 2) avaliar resultados antes e após as intervenções no tratamento ou para monitorar a resposta do paciente ao tratamento; 3) acompanhar por meio de um instrumento padronizado a tomada de decisões, por parte do psicoterapeuta, para melhorar a qualidade do tratamento (Lambert et al., 2004). O autor relata que o instrumento não foi idealizado para o diagnóstico de pacientes, pois esta tarefa pode ser realizada com instrumentos desenvolvidos para esta finalidade, como o MMPI-2 nos EUA, e que demandam mais tempo para aplicação.

Segundo o modelo de medida proposto por Lambert (1983), o progresso do paciente na psicoterapia deve ser avaliado por meio de três aspectos da sua vida: o desconforto subjetivo, as relações interpessoais, e o desempenho do papel social. Estas áreas de funcionamento sugerem um *continuum* dos sentimentos e sensações percebidas do seu estado interior, o que vivencia no seu íntimo, como se relaciona com as pessoas significativas da sua vida e como lida com tarefas relacionadas à produtividade, por exemplo, o trabalho, a escola ou qualquer outra atividade, inclusive o lazer. O desenvolvimento e seleção dos itens para o *OQ-45* foram determinados por uma série de considerações. Inicialmente foram selecionadas questões relativas a disfunções ou problemas mais comuns em uma ampla variedade de distúrbios. A seguir, itens que abordavam os sintomas com maior probabilidade de ocorrência, independente da particularidade dos problemas. E, na terceira etapa, os itens que avaliavam características pessoais e sociais relevantes quanto à qualidade de vida. A partir destas considerações os itens foram agrupados teoricamente em três subescalas denominadas: Desconforto Subjetivo (SD), Relações Interpessoais (IR) e Papel Social (SR) (Lambert et al., 2004).

Diversas pesquisas foram realizadas, em diferentes países, com o objetivo de avaliar as propriedades psicométricas do *OQ-45*. Dentre estas, os estudos mais recentes encontraram bons indicadores de precisão,

tanto para as subescalas quanto para o escore total. O estudo de Boswell, White, Sims, Harrist, & Romans (2013) avaliou amostra de 220 universitários nos EUA reportou α de Cronbach para $SD=0,93$; $IR=0,78$; $SR=0,70$ e Escore Total = $0,94$. Estudo com a versão alemã da escala aplicada em pacientes psiquiátricos ($N=294$) (Puschner, Cosh, & Becker, 2015) também observou coeficientes semelhantes (α de Cronbach para $SD=0,93$; $IR=0,74$; $SR=0,68$ e Escore Total= $0,93$), assim como o estudo de Machado & Fassnacht (2015) realizados com diferentes estratos da população portuguesa em que se observou coeficientes entre $0,61$ e $0,92$ nas amostras de universitários, entre $0,59$ e $0,92$ na amostra da comunidade geral e entre $0,56$ e $0,93$ para a amostra clínica.

A versão adaptada para o português brasileiro por Carvalho e Rocha (2009) foi avaliada por Silva (2013) que verificou α de Cronbach para $SD=0,91$ na amostra não-clínica e de $0,90$ na amostra clínica; $IR=0,71$ e $0,64$; $SR=0,65$ e $0,68$; Escore Total = $0,93$ e $0,92$ nas respectivas amostras. Além disso, a autora realizou um estudo de precisão da versão brasileira pelo método de teste e reteste com amostra de universitários ($n=33$) e amostra clínica, pacientes em clínicas escola ($n=55$), o intervalo de aplicação variou entre sete e 14 dias. Os resultados indicaram correlações de Person entre as diferentes aplicações que variaram entre $0,58$ e $0,91$ para os estudantes universitário, e entre $0,75$ e $0,89$ para a amostra de pacientes.

Se por um lado, a precisão da escala frequentemente é avaliada como adequada, por outro, alguns estudos têm encontrado dificuldades em sustentar empiricamente a estrutura com três fatores. Jong et al. (2007) em estudo com a versão holandesa, através da Análise Fatorial Exploratória (AFE) encontraram dois fatores adicionais: um composto pelos itens do papel social e outro que reflete ansiedade e sintomas somáticos. Já, Mueller, Lambert e Burlingame (1998) verificaram por meio de Análise Fatorial Confirmatória (AFC) melhores índices de ajustamento para a estrutura unidimensional, quando comparadas com as estruturas de três fatores. Mais recentemente, estudos têm encontrado evidências de validade em uma estrutura *bifactor* para o instrumento em questão. Neste modelo pressupõe-se que os itens da escala capturam duas fontes diferentes de variância, variância específica, oriundas de três fatores específicos e variância comum compartilhada entre todos os itens, também denominado fator geral ou Desajustamento Global (DG) (Lambert et al., 2004; Lo Coco, et al., 2008).

O uso de instrumentos psicológicos em diferentes países, adaptados para outras línguas diferente do original, requer uma versão equiva-

lente, uma vez que os construtos psicológicos subjacentes precisam ser equivalentes (Butcher, Derksen, Sloore, & Sirigatti, 2003). Para o uso em nosso país, Carvalho e Rocha (2009) traduziram e adaptaram o *OQ-45.2* com base na versão original e na de Portugal, tendo elaborado versão brasileira para o desenvolvimento de futuras pesquisas das propriedades psicométricas.

Portanto, faz-se necessária a realização de pesquisas para estimar evidências de validade com base na estrutura interna deste instrumento para esta população, haja vista que a falta de tais informações compromete qualquer inferência a partir dos escores do teste (AERA, APA, & NCME, 2014). Outra importante lacuna a respeito deste instrumento diz respeito falta de informação de sua capacidade em avaliar de forma semelhante pessoas dos diferentes sexos. Esta lacuna fica ainda mais evidente à medida que os participantes das pesquisas são comumente comparados através dos escores do *OQ-45.2* em função desta variável (Lambert et al., 2004; Jong et al., 2007; Machado & Fassnacht, 2015; Rodríguez, 2000). Desta maneira ter evidências empíricas de que as variáveis observadas, os itens da escala, estão relacionadas aos construtos latentes da mesma maneira para os diferentes grupos torna-se um imperativo aos pesquisadores da área, uma vez que este é um pré-requisito à comparação destes grupos através dos resultados brutos provenientes da escala (Milfont & Fisher, 2010).

A fim de oferecer contribuições para o preenchimento destas lacunas, a presente pesquisa teve como principal objetivo a) estimar as primeiras evidências de validade da estrutura interna e precisão do *OQ-45.2* para a população brasileira, b) avaliar a invariância da estrutura interna dessa medida em relação ao sexo dos participantes, e c) estimar os parâmetros dos itens (dificuldade e ajuste) e dos participantes (nível de intensidade no construto).

Métodos

Participantes

A amostra obtida foi composta por 419 participantes, dos quais 263 (62,8%) eram mulheres, com idades que variavam entre 18 e 78 anos ($M=32,87\pm 15,6$). Quanto à escolaridade, na amostra total 46% dos participantes tinham nível superior completo ou incompleto, 39% nível médio completo ou incompleto, 8% ensino fundamental completo ou incompleto e 7% não responderam. A amostra total ($N=419$) subdividida

em dois grupos: pacientes (N=59), pacientes em tratamento em uma clínica-escola de Psicologia, situada em uma cidade da grande São Paulo, e não pacientes (N=360). Quanto à faixa etária, 59% tinham idade entre 20 e 39 anos, 21% entre 40 e 59 anos, 12% entre 18 e 19 anos, 7% entre 60 e 78 anos e 1% não responderam/não declararam a idade.

A amostra também foi descrita quanto aos recursos econômicos. Para esta classificação foi utilizado o Critério de Classificação Econômica Brasil da Associação Brasileira de Empresas de Pesquisa (ABEP), que é um critério que tem como função estimar o poder de compra das pessoas. A classificação de classes econômicas da ABEP é dividida em oito classes e cada uma corresponde a uma renda média familiar (valor bruto em R\$) A1=12,926; A2=8.418; B1=4.418; B2=2.565; C1=1.541; C2=1.024; D=714; e E=477. Na classificação da ABEP, com base no Levantamento Sócio Econômico de 2010, 35,0% classificam-se na Classe B2; 25,5% na C1; 13,0% na C2; 11,0% na B1; 5,0% na A2; 3,0% na D; 0,5% na E; e 7,0% não responderam.

Instrumentos

Escala de Avaliação de Resultados – OQ 45.2 (Lambert et al. 1994): trata-se de um instrumento de autorrelato, composto por 45 itens cujas respostas são dadas numa escala tipo Likert de cinco pontos que variam entre “nunca” (zero ponto), “raramente” (um ponto), “às vezes” (dois pontos), “frequentemente” (três pontos) e “sempre” (quatro pontos). Segundo o estudo original (Lambert, 1996), o instrumento é composto por três dimensões: Desconforto subjetivo (25 itens), Relacionamentos interpessoais (11 itens), e Papel social (9 itens). No entanto, estudos mais recentes apresentam evidências da estrutura bifactor para o instrumento em questão (Lambert et al., 2004; Lo Coco, et al., 2008), nesta perspectiva a estrutura interna é composta por um fator geral denominado Desajustamento Global e três fatores específicos, supracitados. A versão utilizada para esta pesquisa é resultado de uma tradução e adaptação cultural do *OQ-45.2* realizada por Carvalho e Rocha (2009), após a autorização dos autores americanos para tradução, adaptação cultural e equivalência semântica para o uso no Brasil.

Questionário sociodemográfico: instrumento desenvolvido para acessar as principais características da amostra como sexo, idade, escolaridade e nível socioeconômico.

Considerações éticas

O Projeto desta Pesquisa foi aprovado pelo Comitê de Ética da universidade em que foi realizado

sob o protocolo (CAAE 0041.0.132.000-110). Os instrumentos foram aplicados nos participantes, sendo que para todos foi entregue o Termo de Consentimento Livre e Esclarecido (TCLE) no qual constavam os objetivos da pesquisa e a forma de divulgação dos resultados de acordo com as normas éticas. Somente participaram da pesquisa as pessoas que concordaram com os procedimentos da pesquisa e assinaram o TCLE. Vale ressaltar que para os participantes referentes ao grupo de pacientes os instrumentos foram aplicados pelos pesquisadores, sendo estes externos ao processo psicoterapêutico.

Procedimentos e análises estatísticas

A fim de obter evidências de validade da estrutura interna do *OQ-45* empregou-se a Análise Fatorial Confirmatória (AFC). Assim, foram testados e comparados, através do teste de diferença do χ^2 , os seguintes modelos: unidimensional, de três fatores, e *bifactor*. Para tanto foi utilizado o método de estimação Robust Weighted Least Square (WLSMV). A escolha de tais procedimentos foi fundamentada na adequação destes métodos ao nível de mensuração ordinal (Lara & Alexis, 2014). Desta forma, os modelos foram testados a partir dos índices recomendados por Muthén, L. e Muthén, B. (2012): WLSMV χ^2 , gl, CFI, TLI e RMSEA. Tais análises foram realizadas no apoio do pacote estatístico Mplus 7.11 (Muthén, L. & Muthén, B., 2012). Uma vez obtido evidências de validade da estrutura interna do instrumento avaliou-se a invariância do modelo de medida entre participantes do sexo feminino e masculino. Milfont & Fisher, 2010).

Ainda em relação a estrutura interna, verificou-se o modelo apresentava uma dimensionalidade essencial dentro da estrutura *bifactor*. Para tanto, foram empregados os procedimentos descritos por Rios e Wells (2014) Expected common variance (ECV) e Percentage of uncontaminated correlations (PUC). De acordo com os autores o primeiro procedimento proporciona a razão entre a força do fator geral e dos fatores específicos no modelo *bifactor*. Valores de ECV próximos de 1 evidenciam que uma forte dimensão geral está presente nos dados *bifactor*. Enquanto PUC pode ser definido como o número de correlações não contaminadas, dividido pelo número de correlações únicas. Valores de PUC próximos de 1 evidenciam que uma forte dimensão geral está presente nos dados bifactor. A última etapa consistiu na avaliação do grau em que o score total reflete uma variável comum. Para tanto empregou-se avaliação da precisão controlando o efeito do fator geral, através do procedimento denominado por Reise (2012) de Ômega hierárquico ω_H .

Para a descrição dos parâmetros dos itens e dos participantes utilizou-se Teoria de Resposta ao Item (TRI), mais especificamente o *Andrich's rating scale model*, com calibração estimada a partir do método *maximum likelihood*, disponível no *software Winsteps* (Linacre, 2015). A fim de identificar a métrica da escala, ancorou-se a média dos índices de dificuldade dos itens em 0. Desta maneira, diferentes procedimentos estatísticos foram utilizados para obter o nível de traço latente apresentado pelos sujeitos (*theta*), os índices de dificuldade dos itens (*b*), os índices de ajustes dos itens (*infit* e *outfit*), e curvas as características de resposta e os índices de confiabilidade.

Resultados

Os índices de ajustes dos modelos unidimensional, com três fatores e *bifactor* são apresentados na Tabela 1.

Conforme a Tabela 1 observa-se que o modelo bifactor apresentou índices de ajustamentos significativamente superior quando comparado aos outros modelos. De acordo com Rios e Wells (2014) outra evidência de que os dados se adequam a estrutura bifactor deriva da comparação entre o modelo bifactor e unidimensional através de mudanças no índice CFI (ΔCFI), uma vez que o primeiro modelo apresenta valor de ΔCFI superior 0,01. As cargas fatoriais dos itens no modelo bifactor são apresentadas na Tabela 2.

Quanto às cargas fatoriais apresentadas pelos itens, observa-se na Tabela 2 que, no geral, estas são mais elevadas no fator geral quando comparadas as cargas fatoriais apresentadas pelos fatores específicos, o que denota maior variância do fator geral na estrutura proposta. Contudo algumas exceções foram observadas: itens 7, 17, 27, 38. Notadamente, alguns itens apresentaram cargas negativas elevadas, igual ou superior a 0,30, nos fatores específicos, são eles os itens 1 e 30. Destaca-se o fato dos itens 11, 14 e 26 não apresentarem cargas fatoriais igual ou superior 0,30 tanto no fator geral quanto no respectivo fator específico.

Considerando que os valores de ECV e PUC foram 0,75 e 0,61, respectivamente, sugere-se a presença de unidimensionalidade essencial do modelo em questão. Resultados que são complementários aos obtidos no procedimento Ômega hierárquico: igual a 0,92 para o fator geral, e igual 0,02, 0,00004 e 0,01, para os fatores específicos, respectivamente. Sugerindo que uma quantidade muito elevada da variância nos escores podem ser atribuídas ao fator geral.

Quanto a avaliação da invariância do modelo os resultados indicam equivalência do modelo configural entre os grupos, ($\chi^2/gf=1,39$; CFI=0.938; TLI=0.935; RMSEA=0.044 I.C. 90%=0.039-0.048), ou seja, similaridades da estrutura interna geral proposta para o modelo de medida, número de variáveis latentes. Assim como invariância do modelo escalar, que avalia, além da equivalência descrita no modelo configural, se média dos itens são equivalentes entre os grupos, controlado possíveis diferenças a variáveis latentes, são equivalentes para ambos os grupos ($\chi^2/gf=1,32$; CFI=0,941; TLI=0,942; RMSEA=0,040 I.C. 90%= 0,035-0,044), haja vista que o ΔCFI e ΔTLI entre os dois modelos são inferiores a 0,01. Mais especificamente, esses resultados indicam que homens e mulheres com mesma média na variável latente não apresentam médias diferentes nos itens. De acordo com Santos e Primi (2014), o teste do modelo escalar é equivalente à análise de Funcionamento Diferencial do Item (DIF) no parâmetro de dificuldade "b" da TRI. Portanto, pode ser inferir a capacidade do instrumento em avaliar homens e mulheres de maneira semelhante, e que possíveis diferenças entre os escores brutos no teste, ao compara estes grupos, estariam associadas as características psicológicas dos sujeitos avaliados, e não a erros de medida associados ao instrumento.

Em relação aos parâmetros dos participantes e dos itens estimados através da TRI, os resultados são apresentados na Tabela 3, onde se verifica índices de dificuldade dos itens, índices de ajustes dos itens (*infit* e *outfit*), índices de correlação entre item e theta apresentados pelos participantes, e as estatísticas descritivas do parâmetro dos participantes (*theta*).

TABELA 1
Índices de ajustes confirmatórios dos modelos unidimensional, com três fatores e *Bifactor* do OQ-45

Modelo avaliado	χ^2	Df	χ^2/gf	$\Delta\chi^2$	Δgf	CFI	TLI	RMSEA	RMSEA CI 90%
Unidimensional	1960,298	945	2,07			0,90	0,89	0,051	0,048-0,054
Três fatores	1862,425	942	1,98	97,973 <i>p</i> <.001	3	0,91	0,90	0,049	0,045-0,052
<i>Bifactor</i>	1665,158	900	1,85	197,267 <i>p</i> <.001	42	0,93	0,92	0,045	0,042-0,049

TABELA 2
Cargas fatoriais dos itens

Itens	Fator geral	Fatores específicos			θ^2
		1	2	3	
2	0,308	0,220			0,857
3	0,308	0,123			0,890
5	0,613	0,027			0,624
6	0,618	0,176			0,587
8	0,756	0,033			0,427
9	0,765	-0,036			0,413
10	0,674	0,082			0,539
11	0,177	0,233			0,914
13	0,711	-0,223			0,444
15	0,790	-0,099			0,366
22	0,464	0,207			0,742
23	0,673	0,140			0,527
24	0,656	-0,184			0,535
25	0,646	0,272			0,509
27	0,307	0,387			0,756
29	0,538	0,360			0,581
31	0,561	-0,209			0,642
33	0,645	0,070			0,579
34	0,423	0,233			0,767
35	0,462	0,242			0,728
36	0,653	0,311			0,477
40	0,716	0,346			0,367
41	0,595	0,282			0,566
42	0,777	0,141			0,377
45	0,386	0,309			0,756
1	0,370		-0,423		0,684
7	0,368		0,396		0,708
16	0,337		-0,017		0,886
17	0,403		0,488		0,599
18	0,688		0,087		0,519
19	0,472		-0,034		0,776
20	0,612		0,152		0,603
26	0,255		-0,094		0,926
30	0,528		-0,316		0,621
37	0,521		0,320		0,626
43	0,582		-0,058		0,658
4	0,404			0,281	0,758
12	0,362			0,314	0,770
14	-0,024			-0,046	0,997
21	0,502			-0,055	0,744
28	0,496			0,431	0,568
32	0,271			0,304	0,834
38	0,438			0,554	0,501
39	0,392			0,405	0,682
44	0,586			0,409	0,488
$\Sigma\lambda^2$	12,92	1,2	0,82	1,10	
$\Sigma\lambda$	518,9284	11,8336	0,25	6,76	

$\Sigma\lambda^2$ = soma dos quadrados das cargas fatoriais; $\Sigma\lambda$ = soma das cargas fatoriais ao quadrado; θ^2 = variância residual do item.

TABELA 3
Estatísticas descritivas dos índices de dificuldade e ajuste dos itens

<i>Itens</i>	<i>b</i>	<i>S.E.</i>	<i>Infit</i>	<i>Outfit</i>	<i>r</i>	<i>Theta</i>
1	0,55	0,07	1,11	1,24	0,39	
2	-0,52	0,06	1,09	1,07	0,35	
3	-0,28	0,05	1,23	1,40	0,32	
4	-0,66	0,05	1,06	1,08	0,45	
5	-0,37	0,05	0,88	0,86	0,59	
6	-0,57	0,05	0,84	0,84	0,60	
7	-0,24	0,04	1,21	1,42	0,36	
8	0,71	0,07	0,78	0,69	0,53	
9	0,12	0,05	0,75	0,76	0,66	
10	0,08	0,05	0,83	0,78	0,61	
11	0,80	0,09	1,48	2,88	0,14	
12	-0,16	0,05	1,18	1,19	0,37	
13	0,19	0,05	0,83	0,91	0,58	
14	-1,15	0,05	1,64	1,96	0,05	
15	0,17	0,05	0,73	0,70	0,67	
16	-0,90	0,05	1,23	1,28	0,36	
17	-0,05	0,04	1,27	1,43	0,38	
18	-0,26	0,04	0,85	0,92	0,61	
19	-0,13	0,05	1,03	1,02	0,45	
20	0,13	0,05	0,91	0,89	0,55	
21	-0,19	0,05	1,08	1,09	0,47	
22	-0,58	0,05	1,00	0,99	0,48	
23	0,08	0,05	0,84	0,86	0,60	
24	0,23	0,05	1,01	1,15	0,49	
25	-0,09	0,05	0,86	0,87	0,60	
26	0,88	0,08	1,39	1,53	0,19	
27	0,10	0,05	1,21	1,29	0,36	
28	-0,04	0,05	1,00	1,04	0,51	
29	0,17	0,05	0,92	0,91	0,54	
30	0,35	0,05	1,01	1,04	0,47	
31	-0,01	0,05	1,07	1,34	0,45	
32	1,42	0,13	1,10	1,42	0,15	
33	0,24	0,05	0,86	0,82	0,58	
34	-0,18	0,05	1,08	1,11	0,45	
35	0,33	0,05	1,14	1,22	0,41	
36	-0,28	0,05	0,77	0,77	0,66	
37	-0,48	0,04	1,07	1,12	0,50	
38	-0,10	0,05	1,05	1,11	0,45	
39	0,47	0,06	1,07	1,05	0,39	
40	0,12	0,05	0,75	0,72	0,65	
41	-0,01	0,05	0,90	0,94	0,56	
42	-0,32	0,05	0,69	0,69	0,72	
43	0,01	0,05	0,96	1,01	0,52	
44	0,68	0,06	0,90	0,82	0,50	
45	-0,29	0,05	1,07	1,11	0,44	
M	0,00	0,05	1,02	1,10	0,47	-0,71
SD	0,47	0,01	0,20	0,37	0,14	0,57
Mín.	1,42	0,04	1,64	2,88	0,05	1,12
Max.	-1,15	0,13	0,69	0,69	0,72	-2,59

b = índices de dificuldades; *r* = correlação entre item e *theta*; *theta* = nível de traço latente apresentado pelos participantes.

Conforme apresentado na Tabela 3, as análises foram realizadas considerando o fator geral avaliado pelo *OQ-45.2*, haja vista os indicativos de unidimensionalidade essencial verificados através do EVC e PUC. Quanto aos índices de dificuldade dos itens observa-se pequena variação em torno da média, vale ressaltar que através do esquema de ancoragem a dificuldade média dos itens foi centrada em zero. Tais resultados indicam que nenhum item se apresentou muito fácil ou muito difícil de ser endossado pelos participantes. Em outra perspectiva indica melhor capacidade do instrumento em avaliar porção central do contínuo Desajustamento global.

Em relação aos índices de ajuste *Infit* observa-se com exceção dos itens 11, 14 e 26, os itens apresentaram índices considerados adequados (entre 0,7 e 1,3), conforme recomendado por Bond & Fox (2001). O que indica ajustamento dos itens aos padrões de respostas esperado pelo modelo quando os valores de dificuldade dos itens encontram-se próximos aos valores de *theta* das pessoas. Quanto aos índices de *Outfit* a notadamente os itens 7, 11, 17 14 26 e 32 apresentam valores fora do intervalo estabelecido (entre 0,7 e 1,3). Indicando assim padrões de resposta inesperados pelo modelo quando a diferença entre o *theta* e a dificuldade das categorias é muito grande. De maneira semelhante os índices de correlação entre item e nível de *theta* dos participantes indicam a inadequação dos itens 11, 14, 26 e 32, haja vista a incapacidade destes itens, ao contrário dos outros que compuseram a escala, em recuperar o nível de *theta* dos participantes. Em relação aos parâmetros dos participantes estatística descritivas de *theta* indicam que os sujeitos tendem a endossar categoriais de resposta mais baixa nos itens do instrumento ($M=-0,71$ e $DP=0,57$). Através das estatísticas Máximo e Mínimo observa-se alta variabilidade no nível de *theta* dos participantes (entre -2,59 e 1,12), o que indica que a amostra foi composta por pessoas com diferentes níveis de desajustamento global.

Quanto à probabilidade de os respondentes endossarem cada categoria de resposta dos itens que compuseram as dimensões do *OQ-45.2*, análise gráfica das Curvas características de resposta indicaram uma relação monotômica crescente entre os valores de *theta* e as categorias escalares, ou seja, entre o nível de habilidade dos sujeitos e o nível de dificuldade apresentado por cada uma das categorias de respostas. O que demonstra, no geral, a funcionalidade das categorias de resposta em escala Likert de cinco pontos utilizada no instrumento. Por fim, avaliou-se os índices de consistências do instrumento sem o controle do fator geral, os resultados indicaram índices alfa de Cronbach igual a 0,94 para o fator geral DG, e índices iguais a

0,91, 0,72 e 0,70 para os fatores específicos SD, IR e SR, respectivamente.

Discussão

Esta pesquisa teve como objetivo avaliar as propriedades psicométricas do *OQ-45* para a população amostra brasileira. Para tanto, recorreu-se a verificação dos índices de ajuste de três diferentes estruturas fatorais comumente empregadas na literatura (Lo Coco et al., 2008) são estas: estrutura unidimensional, com três fatores e bifactor. Confirmando a hipótese teórica, os resultados sugerem a adequação da estrutura bifactor do *OQ-45.2*, composta por três fatores específicos: SD, IR e SR e fator geral: DG (Lambert et al., 2004). Vale ressaltar que os três fatores específicos correspondem à proposta original da escala proposta por Lambert et al. (1994).

Embora esta pesquisa corrobore os resultados de estudos anteriores referentes à estrutura bifactor do *OQ-45.2*, observa-se que os índices de ajuste obtidos na AFC realizada nestas pesquisas são relativamente superiores aos obtidos por outros pesquisadores, por exemplo: Lo Coco et al. (2008) ao avaliar a estrutura *bifactor* diante população italiana ($df/\chi^2=2.99$; CFI 0.830; AGFI 0.806; SRMR 0.054; RMSEA 0.049). Tais evidências indicam a adequação do método de estimação utilizado, WLSMV, o qual se baseia em matriz de correlações policóricas que, portanto, é mais adequado ao nível de mensuração ordinal, como no caso dos testes psicológicos respondidos em escala do tipo Likert (Cook, Kallen, & Amtmann, 2009; Muthén, L. & Muthén, B., 2012), em contraste ao método de estimação maximum likelihood utilizado na pesquisa supracitada.

Ainda em relação à estrutura interna, destaca-se a invariância da estrutura fatorial quando comparada à avaliação dos diferentes grupos em função dos sexos dos participantes, O que evidencia a capacidade da escala em avaliar tais grupos de maneiras semelhantes, de modo a possibilitar a comparação entre os escores destes grupos em estudos futuros (Cook, Kallen, & Amtmann, 2009; Milfont & Fisher, 2010). Diferentes autores vêm destacando a importância deste procedimento para os testes psicológicos haja vista que a falta de tais informações pode levar pesquisadores/profissionais práticos a fazerem comparações inadequadas, infringindo, inclusive, questões éticas ligadas a avaliação psicológica (Milfont & Fisher, 2010).

Através da TRI foi possível verificar os parâmetros dos itens que compõem a escala, resultados que evidenciam forças e limitações dos itens que a compõem. Dentre os aspectos positivos, encontram-se os índices

de *infit* e *outfit* que em grande parte dos itens se adequam ao padrão de resposta esperado pelo modelo, bem como a adequação da escala de tipo Likert adotada no instrumento. Em relação às fragilidades da escala pode-se apontar quatro itens, são eles: 11, 14, 26 e 32, com índices de ajustes fora do esperado, assim como baixa capacidade em recuperar o theta dos participantes.

De acordo com Wright & Linacre, (1994) a estatística Outfit uma estatística baseada em valores inesperados periféricos e, portanto, mais sensível à influência de outlier. Desta maneira, infere-se que estas são situações menos preocupantes no que diz respeito aos índices de ajustes dos itens, haja vista a maior relevância da capacidade do item de se ajustar ao padrão de resposta esperado pelo modelo, quando os níveis de traços latentes dos sujeitos estão próximos ao nível de intensidade dos itens (*Infit*). O que justifica a manutenção do item 32.

Por fim, as análises sugerem a exclusão dos itens 11, 14 e 26, da escala, uma vez que não apresentaram cargas fatoriais iguais ou superior a 0,30 nos fatores específicos e nem no fator geral, conforme esperado teoricamente. Tais indicações foram confirmadas através da TRI, já os itens em questão não apresentaram bons índices de ajustes e correlação com a estimativa do nível de traço latente dos sujeitos. Vale ressaltar que incompatibilidade destes itens ao modelo proposto vem sendo observado por diferentes autores, Lambert et al. (2004), Lo Coco et al. (2008), Vacarezza, Florenzano, & Trapp, 2008, entre outros. Entretanto os itens 11 e 26 são utilizados como indicadores de uso de álcool e drogas. Assim os autores supracitados sugerem a manutenção destes itens para avaliação qualitativa destas características. Com relação ao

item 14 sobre trabalho, a hipótese é que os resultados tenham relação com a adaptação do item “Eu trabalho muito”, que poderia ser modificado para “Eu trabalho exageradamente”. Haja vista que para grande parte da população brasileira o trabalho é muito valorizado, e culturalmente não é reconhecido como desencadeador de desconforto subjetivo.

Quanto as evidências de precisão do OQ-45, os resultados obtidos são coerentes com a proposta original do instrumento (Lambert et al., 1996) que obteve índices de consistência interna, alpha de Cronbach, que variavam entre 0,70 e 0,93 para os fatores específicos e 0,92 para a escala total. Ressalta-se que índices similares foram alcançados por diferentes pesquisadores que avaliaram versões adaptadas para outros países (Chiappelli, Lo Coco, Gullo, Bensi, & Prestano, 2008; Wennberg, Philips, & de Jong, 2010). Contudo, entende-se que os objetivos estabelecidos nesta pesquisa foram satisfatoriamente atingidos. Conclui-se que o instrumento apresentou boas propriedades psicométricas para avaliação da população brasileira, mostrando-se promissor para o uso profissional.

Novas pesquisas se fazem necessárias, a fim de estabelecer normas de interpretação dos escores brutos da escala, imputando assim significado aos resultados obtidos através da aplicação do OQ-45.2. Nesta direção, sugere-se a utilização da TRI através do procedimento Mapa de Itens-pessoas (Peixoto & Nakano, 2014). Assim como a realização de novos estudos que busquem outras evidências de validade para OQ-45 para a população brasileira (AERA, APA, & NCME, 2014). Por fim, é importante continuar a avaliação do desempenho do instrumento com amostras clínicas e mais diversificadas regionalmente.

Referências

- Abishe, D. R. (2008). *Cross-cultural comparison of the Outcome Questionnaire*. Tese de doutorado em filosofia. Universidade de Oklahoma. Norman: Oklahoma. ProQuest: UMI 3304221.
- AERA – American Educational Research Association, APA – American Psychological Association, NCME – National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Antony, M. M. & Barlow, D. H. (2010). *Handbook of assessment and treatment planning for psychological disorders* (2ª ed.). New York: Guilford Press.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model*. Mahwah, NJ: LEA.
- Boswell, D. L., White, J. K., Sims, W. D., Harrist, R. S., & Romans, J. S. (2013). Reliability and validity of the Outcome Questionnaire-45.2. *Psychological Reports: Mental & Physical Health*, 112(3), 689-693. <http://dx.doi.org/10.2466/02.08.PR0.112.3.689-693>
- Brum, E. H. M., Frizzo, G. B., Gomes, Grill. A., Silva, M. R., Sousa, D. D., & Piccinini, C. A. (2012). Evolução dos modelos de pesquisa em psicoterapia. *Estudos de Psicologia*, 9(2), 259-269. <http://dx.doi.org/10.1590/S0103-166X2012000200012>

- Butcher, J., Derksen, J., Sloore, H., & Sirigatti, S. (2003). Objective personality assessment of people in diverse cultures: European adaptations of the MMPI-2. *Behaviour Research and Therapy*, *41*(7), 819-840. [http://dx.doi.org/10.1016/S0005-7967\(02\)00186-9](http://dx.doi.org/10.1016/S0005-7967(02)00186-9)
- Carvalho, L. F. & Rocha, G. M.A. (2009). Tradução e adaptação cultural do Outcome Questionnaire (OQ-45). *Psico-USF*, *14*(3), 309-316. <http://dx.doi.org/10.1590/S1413-82712009000300007>
- Chiappelli, M., Lo Coco, G., Gullo, S., Bensi, L., & Prestano, C. (2008). L'Outcome-Questionnaire 45.2. The Outcome Questionnaire 45.2. Italian validation of an instrument for the assessment of psychological treatments. *Epidemiologia E Psichiatria Sociale*, *17*, 152-216.
- Cook K. F., Kallen, M. A., & Amtmann D. (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res*, *18*(4), 447-460. <http://dx.doi.org/10.1007/s11136-009-9464-4>
- Del Prette, G. (2011). *Objetivos analítico-comportamentais e estratégias de intervenção nas interações com a criança em sessões de duas renomadas terapeutas infantis*. Tese de doutorado não publicada. Universidade de São Paulo.
- Doerfler, L. A., Addis, M. E., & Moran, P. W. (2002). Evaluating mental health outcomes in an inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32. *The Journal of Behavioral Health Services & Research*, *29*(4), 394-403. <http://dx.doi.org/10.1007/BF02287346>
- Harmon, C., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J.S. (2005). Improving outcomes for poorly responding clients: The use of Clinical Support Tools and feedback to clients. *A Journal of Clinical Psychology*, *61*(2), 175-185. <http://dx.doi.org/10.1002/jclp.20109>
- Hatfield, D. & Ogles, B. M. (2004). The Use of Outcome Measures by Psychologists in Clinical Practice. *Professional Psychology: Research & Practice*, *35*(5), 485-491. <http://dx.doi.org/10.1037/0735-7028.35.5.485>
- Honda, G. C., Peixoto, E. M., Rocha, G. M., & Enéas, M. L. E. (2015). Psicoterapia breve psicodinâmica no processo de formação profissional. In T. V. Santeiro & G. M. A. Rocha (Orgs.). *Clínica de orientação psicanalítica: compromissos, sonhos e inspirações no processo de formação* (pp. 76-80). São Paulo: Vetor.
- Huang, S., Puschner, B., Lambert, M. J., & Kordy, H. (2004). Veraderungs messung in der psychotherapie mit dem Ergebnis Frage Bogen (EB-45) Assessment of change in psychotherapy with the German version of the Outcome Questionnaire (OQ-45). *Zeitschrift für Differentielle und Diagnostische Psychologie*, *25*, 141-151.
- Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy*, *14*(4), 288-301. <http://dx.doi.org/10.1002/cpp.529>
- Lambert, M. J. (1983). Introduction to assessment of psychotherapy outcome: Historical perspective and current issues. In: M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 3-32). New York: John Wiley and Sons.
- Lambert, M. J., Burlingame, G. N., Umphress, V., Hanse, N.B., Vermeersh, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, *3*, 249-258. [http://dx.doi.org/10.1002/\(SICI\)1099-0879\(199612\)3:4<249::AID-CPP106>3.0.CO;2-S](http://dx.doi.org/10.1002/(SICI)1099-0879(199612)3:4<249::AID-CPP106>3.0.CO;2-S)
- Lambert, M. J. & Finch, A. E. (1999). The Outcome Questionnaire. M. E. Maruish (Ed.). *The use of psychological testing for treatment planning and outcomes assessment* (pp. 831-869). New Jersey: Lawrence Erlbaum Associates, Publishers.
- Lambert, M.J., Hansen, N.B., & Finch, A.E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*, 159-172. <http://dx.doi.org/10.1037/0022-006X.69.2.159>
- Lambert, M. J. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin and S. L. Garfield (Eds.). *Handbook of psychotherapy and behavior change*, *4*, 72-113. New York: John Wiley and Sons.
- Lambert, M. J., Morton, J. J., Hartfield, D., Harmon, C., Hamilton, S. Reid, R. C., Shimokawa, K., Christopherson, C., & Burlingame, G. M. (2004). *Administration and Scoring Manual for the OQ-45.2 Outcome Questionnaire*. Salt Lake City: American Professional Credentialing Services.
- Lambert, J. M., Smart, D. W., Campbell, M. P., Hawkins, E. J., Harmon, C., & Slade, K. L. (2006). Psychotherapy outcome, as measured by the OQ-45, in African American, Asia/Pacific Islander, Latino/a, and Native American clients compared with matched caucasian clients. *Journal of College Student Psychotherapy*, *20*(4), 17-29. http://dx.doi.org/10.1300/J035v20n04_03
- Lara, D. & Alexis, S. (2014). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. *Revista Argentina de Ciencias del Comportamiento*, *6*(1), 39-48.
- Lara, C., Cruz, C., Vacarezza, A., Florenzano U. R., & Trapp, A. (2008). Análisis comparativo de dos instrumentos de evaluación clínica: OQ45 e InterRAI – Salud Mental. *Revista Chilena de NeuroPsiquiatria*, *46*(3), 192-198. <http://dx.doi.org/10.4067/S0717-92272008000300004>
- Linacre, J. M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.
- Linacre J. M. (2015). A user's guide to Winsteps Ministep: Rasch-model computer programs. Retrieved from: <http://www.winsteps.com>

- Lo Coco, G. L., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, M. J. (2008). The factorial structure of the Outcome Questionnaire-45: A study with an Italian sample. *Clinical Psychology and Psychotherapy*, 15, 418-423. <http://dx.doi.org/10.1002/cpp.601>
- Machado, P. P. P. & Fassnacht, D. B. (2015). The Portuguese version of the Outcome Questionnaire (OQ-45): Normative data, reliability, and clinical significance cut-offs scores. *Psychology and Psychotherapy: Theory, Research and Practice* (2015), 88, 427-437. <http://dx.doi.org/10.1111/papt.12048>
- Milfont, T. L. & Fisher, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-121.
- Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, 70(2), 248-262. http://dx.doi.org/10.1207/s15327752jpa7002_5
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA.
- Parra, G. & Bergen, V. (2002). OQ 45-2 Cuestionario para evaluación de resultados y evolución en psicoterapia: adaptación, validación, e indicaciones para su aplicación e interpretación. *Revista de Terapia*, 20, 161-176.
- Pasquali, L. (2010). Testes Referentes a Construto: Teorias e Modelos de Construção. In: L. Pasquali et al. (Orgs.), *Instrumentação Psicológica: Fundamentos e prática* (pp. 165-198). Porto Alegre: Artmed.
- Peuker, A. C., Habigzang, L. F., Koller, S. H., & Araujo, L. B. (2009). Avaliação de processo e resultado em psicoterapias: uma revisão. *Psicologia em Estudo*, 14(3), 439-445. <http://dx.doi.org/10.1590/S1413-73722009000300004>
- Pieta, M. A. M., Siegmund, G., Gomes, W. B., & Gauer, G. (2015). Desenvolvimento de protocolos para acompanhamento de psicoterapia pela Internet. *Contextos Clínicos*, 8(2), 128-140. <http://dx.doi.org/10.4013/ctc.2015.82.02>
- Peixoto, E. M. & Nakano, T. C. (2014). Problemas e perspectivas na utilização dos testes psicológicos em psicologia do esporte. In: C. R. Campos, T. C. Nakano. (Org.). *Avaliação Psicológica direcionada a populações específicas: técnicas, métodos e estratégias* (pp. 201-23). São Paulo: Vetor.
- Primi, R., Carvalho, L. F., Miguel, F. K., & Muniz, M. (2010). Resultado dos fatores da BFP por meio da Teoria de Resposta ao Item: interpretação referenciada no item. In: C. H. S. S., Nunes, C. S., Hutz, & M. F. O., Nunes. (Org.). *Bateria Fatorial de Personalidade (BFP): manual técnico* (pp. 153-170). São Paulo: Casa do Psicólogo.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696. <http://dx.doi.org/10.1080/00273171.2012.715555>
- Rios, J. & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 08-116. [10.7334/psicothema2013.260](https://doi.org/10.7334/psicothema2013.260)
- Rodríguez, M. A. von B. (2000). *Investigación Empírica em Psicoterapia: Validación del Cuestionario de Resultados Terapêuticos OQ-45.2*. Tesis para optar al grado de Licenciado em Psicologia. Universidad Nacional Andres Bello. Santiago do Chile.
- Santos, D. & Primi, R. (2014). *Social and emotional development and school learning: a measurement proposal in support of public policy*. Technical report for Organization for Economic Cooperation and Development (OCDE) Rio de Janeiro State Education Department (SEEDUC) and Ayrton Senna Institute. São Paulo: Ayrton Senna Institute.
- Serralta, F. B., Nunes, M. L. T., & Eizirik, C. L. (2007). Elaboração da versão em português do Psychotherapy Process Q-Set. *Revista Brasileira de Psiquiatria*, 29(1), 44-55. <http://dx.doi.org/10.1590/S0101-81082007000100011>
- Silva, S. M. (2013). *Escala de Avaliação de Resultados (Outcome Questionnaire) – OQ-45.2: Validade e precisão*. Tese de Doutorado [não publicada]. Universidade de São Paulo.
- Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. *Journal of Psychoeducational Assessment*, 15, 40-55. <http://dx.doi.org/10.1177/073428299701500104>
- Wennberg, P., Philips, B., & de Jong, K. (2010). The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample. *Psychology and Psychotherapy: Theory, Research and Practice*, 83, 325-329. <http://dx.doi.org/10.1348/147608309X478715>

Autores:

Sonia Maria da Silva – Universidade Guarulhos.
 Iraí Cristina Boccatto Alves – Doutora, Instituto de Psicologia da USP.
 Evandro Morais Peixoto – Doutor, Pontifícia Universidade Católica de Campinas.
 Gláucia Mitsuko Ataka Rocha – Doutora, Instituto de Psicologia da USP.
 Tatiana de Cássia Nakano – Doutora, Pontifícia Universidade Católica de Campinas.

Endereço para correspondência:

Evandro Morais Peixoto
 Rua Liliane Regina, 3
 06386-300 Carapicuíba, SP, Brasil
 <epeixoto_6@hotmail.com>

Recebido em: 15.07.2016
 Aceito em: 22.09.2016