

Desenvolvimento de um Banco de Itens para Avaliar o Transtorno Depressivo Maior

Hemerson Fillipy Silva Sales¹

Universidade Federal da Paraíba, João Pessoa-PB, Brasil

Josemberg Moura de Andrade

Universidade de Brasília, Brasília-DF, Brasil

Vivianne Freitas Oliveira Asfora

Universidade Federal da Paraíba, João Pessoa-PB, Brasil

RESUMO

O objetivo desta pesquisa foi desenvolver um banco de itens para avaliar o Transtorno Depressivo Maior (TDM). Para isso, foram elaborados 150 itens divididos em dois conjuntos (A e B) e utilizou-se uma amostra de conveniência de 501 participantes das cidades de Parnaíba-PI e João Pessoa- PB. Quanto aos resultados, as análises fatoriais e o método Hull mostraram que os conjuntos apresentaram apenas um fator com valores alfa excelentes (0,93 e 0,96, respectivamente), assegurando assim a unidimensionalidade e, conseqüentemente, a independência local destes. Já em relação às calibrações, as primeiras estimativas mostraram a necessidade de exclusão e de alteração das categorias de resposta de alguns itens, os quais foram recategorizados e reestimados seus parâmetros. Ao final da pesquisa, o banco de itens apresentou bons parâmetros psicométricos, estando calibrado e pronto para ser utilizado para prover testes adaptativos computadorizados.

Palavras-chave: transtorno depressivo maior; banco de itens; teoria de resposta ao item.

ABSTRACT – Development of an item pool to assess major depressive disorder

The purpose of this research is to develop an item pool to assess Major Depressive Disorder (MDD). To do so, we elaborated 150 items divided into two sets (A and B), and then had those items rated by a sample 501 adults from the cities of Parnaíba-PI and João Pessoa- PB. As for the results, the factor analysis using the Hull Method revealed the salience of a large general factor that produced excellent alpha values in both item sets (.93 and .96, respectively), thus ensuring the unidimensionality and local independence of the items. Regarding the calibrations, the obtained estimates indicated the need to exclude and change the response categories of some items, which were recategorized, and had their parameters reestimated. The item pool presented sound psychometric properties, and it can now be included in computerized adaptive testing.

Keywords: major depressive disorder; items bank; item response theory.

RESUMEN – Elaboración de un banco de ítems para evaluar el trastorno depresivo mayor

El objetivo de esta investigación fue elaborar un banco de ítems para evaluar el Trastorno Depresivo Mayor (TDM). Para ello, se elaboraron 150 ítems divididos en dos conjuntos (A y B) y se utilizaron una muestra de 501 participantes de las ciudades de Parnaíba-PI y João Pessoa-PB. Sobre los resultados, los análisis factoriales y el método Hull revelaron que los conjuntos presentaron solamente un factor con valores alfa excelentes (0,93 y 0,96, respectivamente). Todavía cabe señalar la unidimensionalidad y consecuentemente, la independencia de los conjuntos. Por lo que se refiere a las calibraciones, las primeras estimaciones mostraron una necesidad de exclusión y de alteración de las categorías de respuesta de algunos ítems, los cuales, dichos ítems fueron recategorizados y sus parámetros reestimados. Al final de la investigación, el banco de ítems presentó buenos parámetros psicométricos, estando el mismo calibrado y listo para ser utilizado para probar Tests Adaptativos Informatizados.

Palabras clave: trastorno depresivo mayor; banco de ítems; teoría de respuesta al ítem.

A depressão é um dos transtornos psiquiátricos que mais afetam pessoas no mundo todo. Além de ter como principais características o seu potencial incapacitante e o risco de suicídio (*World Health Organization* [WHO], 2015), a depressão também tem sido associada a uma

variedade de outras doenças, como câncer (Gorman, Su, Roberts, Dominick, & Malcarne, 2015), diabetes (Holt et al., 2014), demência (Cherbuin, Kim, & Anstey, 2016), distúrbios do sono (Nyer et al., 2013) e acidente vascular cerebral (Fei, et al., 2016).

¹ Endereço para correspondência: Universidade Federal da Paraíba, Campus I, Centro de Ciências Humanas e Letras – Departamento de Psicologia. Programa de Pós-graduação em Psicologia Social, 58051-900, João Pessoa, PB. Tel.: (83) 3216-7337. E-mail: fillipsilva@hotmail.com

Os tipos de tratamento para a depressão podem ser bastante diversificados. Para intervenção inicial, por exemplo, em pacientes com depressão leve ou moderada, são indicadas tanto medicações antidepressivas quanto psicoterapia focada. Especificamente, nos casos de pacientes com depressão moderada ou grave são indicadas medicações antidepressivas combinadas a intervenções psicoterápicas. A eletroconvulsoterapia é recomendada somente como tratamento de escolha para pacientes com depressão grave que não respondem às intervenções psicoterápicas e/ou farmacológicas (*American Psychological Association [APA], 2010*).

Além dos citados, outros tipos de tratamento menos convencionais têm ganhado espaço nos últimos anos, como a terapia cognitivo-comportamental aliada à prática de exercícios físicos (Cuijpers, Wit, & Taylor, 2014; Stanton & Reaburn, 2014); e uma variedade de técnicas de neuromodulação, como a estimulação cerebral profunda (Fenoy et al., 2016), a estimulação elétrica transcraniana por corrente contínua (Brunoni et al., 2014; Shiozawa et al., 2014), a estimulação do nervo vago (Tisi, Franzini, Messina, Savino, & Gambini, 2014) e a estimulação magnética transcraniana (Anderson, Hoy, Daskalakis, & Fitzgerald, 2016).

Como o tipo de tratamento vai depender da gravidade e da reatividade da doença, para que seja realizado de forma apropriada, é importante que o rastreamento e a avaliação da gravidade da depressão sejam precisos, evidenciando-se assim a importância da utilização de instrumentos psicológicos para avaliar o problema. Na grande maioria dos instrumentos disponíveis, a gravidade da doença é avaliada por meio do escore total, ou seja, a soma de todas as respostas atribuídas a cada um dos sintomas apresentados. Logo, a interpretação dos escores dessas medidas são dependentes de comparações com um grupo com as mesmas características daquelas do indivíduo avaliado (grupo normativo). É por meio de estudos realizados com o grupo normativo que são estabelecidos pontos de corte para que os resultados encontrados possam ser contextualizados.

Os procedimentos descritos anteriormente foram desenvolvidos, sobretudo, com base na Teoria Clássica dos Testes (TCT), conhecida como a teoria psicométrica tradicional. Tal teoria considera que a magnitude do construto que está sendo medido é representada pela pontuação total de um indivíduo a um conjunto de estímulos comportamentais, a saber, os itens do teste (Pasquali, 2013). Destaca-se, porém, que a utilização da TCT implica em uma série de dificuldades que, apesar de não impedirem sua utilização em larga escala, podem se apresentar como limitações, a saber: (a) o fato dos parâmetros dos itens dependerem da amostra de sujeitos utilizada, o que pressupõe que, se a amostra não for representativa da população, os parâmetros dos itens não podem ser considerados válidos; (b) o fato da fidedignidade ser frequentemente definida como a correlação

entre escores obtidos em termos de formas paralelas de um teste, e formas paralelas são difíceis de serem construídas e; (c) por fim, a suposição de que os erros de medida sejam iguais para todos os sujeitos, algo considerado improvável (Hambleton, & Swaminathan, 1985; Pasquali, 2007, 2013).

Um problema típico das medidas de depressão desenvolvidas com base na TCT é o fato delas não fornecerem informações mais precisas do quanto cada sintoma contribui para a gravidade da doença. Nesse tipo de medida, por exemplo, é possível um sujeito que relatou pensamentos ou tentativa de suicídio ter uma pontuação igual ou mesmo menor que a pontuação de um outro sujeito que não relatou. Nesse caso, a igualdade das pontuações pode distorcer o fato de que o primeiro sujeito apresenta uma sintomatologia mais grave da depressão do que o segundo sujeito.

Considerando essa problemática, nas últimas décadas, muitos pesquisadores têm se voltado para a adequação de testes psicológicos e educacionais por meio de outra teoria psicométrica que, embora não substitua por completo a teoria tradicional, tem resolvido algumas de suas limitações. Trata-se da Teoria de Resposta ao Item (TRI). A TRI é uma teoria psicométrica que propõe modelos matemáticos e métodos estatísticos para avaliar o desempenho de um sujeito em virtude do seu traço latente, ou seja, uma característica do indivíduo que não pode ser observada diretamente, mas sim inferida a partir da observação de variáveis secundárias relacionadas a ela (Andrade, Tavares, & Valle, 2000).

Alguns estudos, tanto na literatura internacional (Doty et al., 2018; Uher et al., 2008; Yan & Kao, 2014) quanto nacional (Andrade, 2008; Castro, Trentini, & Riboldi, 2010), têm considerado as vantagens da TRI para investigar aspectos que não poderiam ser testados apenas por meio da TCT. Por exemplo, Uher et al. (2008) examinaram três medidas de depressão por meio da TRI, a saber: a *Hamilton Rating Scale for Depression (HRS-D)*, a *Montgomery-Asberg Depression Rating Scale (MADRS)* e o *Beck Depression Inventory (BDI)*. No que diz respeito à qualidade das medidas, foi relatado que o MADRS e o BDI apresentaram estimativas internamente consistentes, mas que eram mutuamente distintas quanto à gravidade da depressão. Logo, os pesquisadores recomendaram o uso destes como medidas complementares. Por outro lado, a HRS-D não foi internamente consistente e apresentou itens pouco apropriados para pacientes clínicos.

Do mesmo modo, no Brasil, Castro, Trentini e Riboldi (2010) evidenciaram algumas das vantagens provenientes da TRI por meio da aplicação de um dos modelos estatísticos dessa teoria ao BDI. Um dos resultados encontrados pelos autores foi que os respondentes do BDI não conseguiam distinguir algumas das afirmações nas categorias de respostas da maioria dos itens. A partir desse aspecto, os autores indicaram a

necessidade de se repensar o dimensionamento da escala de respostas.

Descobertas desse tipo só são possíveis porque a TRI utiliza o item como unidade de análise e não o teste como ocorre na TCT. Assim, diferente dessa, o principal objetivo da TRI não é o desenvolvimento de testes, mas sim a construção de bancos de itens, os quais devem ser utilizados, sobretudo, para prover testes adaptativos computadorizados (*computerized adaptive testing* – CATs). Esse tipo de testagem tem por característica fundamental aplicar testes específicos para cada sujeito, os quais devem se assemelhar ao seu nível de aptidão (Pasquali, 2007, 2013). Os CATs têm ganhado espaço nos últimos anos, na literatura internacional, e têm apresentado uma série de vantagens quando comparados aos testes tradicionais do tipo lápis e papel (Aybek & Demirtasli, 2017).

Apesar disso, no Brasil os estudos sobre CATs vêm se desenvolvendo de forma bastante tímida, principalmente no que diz respeito à área da Avaliação Psicológica. Considerando a revisão realizada por Manseira e Misaghi (2015), por exemplo, até o ano de 2013 não havia sido encontrado nenhum trabalho nacional direcionado ao uso de CATs com fins de avaliação de transtornos psicológicos.

Partindo dessa revisão, realizou-se uma nova busca no período entre 2014 e 2016 nas bases de dados: SciELO, Pepsic, Index Psi, Google acadêmico e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Para isso, considerou-se os estudos de construção de CATs, até iniciativas de incentivo a esse propósito, como a elaboração de bancos de itens e o desenvolvimento de algoritmos e *softwares* para serem utilizados na área. Dentre os estudos encontrados, apenas dois estavam relacionados à avaliação psicológica, a saber, duas adaptações de bancos de itens, sendo uma delas inclusive, para avaliar a depressão. Trata-se do Banco de Ansiedade e Depressão do *Patient-Reported Outcomes Measurement Information System* (PROMIS-D; Castro, Rezende, Mendonça, Silva, & Pinto, 2014), o qual foi adaptado culturalmente para a língua portuguesa.

O resultado dessa busca evidenciou uma carência da literatura nacional quanto a esse nível de testagem. Assim, o objetivo desta pesquisa foi desenvolver um banco de itens para avaliar o Transtorno Depressivo Maior (TDM). Como esse procedimento é a parte mais árdua e demorada no processo de desenvolvimento de CATs, acredita-se que disponibilizar um banco de itens para avaliação da depressão é um passo essencial para motivar a construção de CATs para avaliar esse transtorno.

Método

Elaboração dos Itens

Para a elaboração dos itens, foram utilizados como referências tanto os critérios diagnósticos do *Diagnostic and Statistical Manual of Mental Disorders* – DSM-5

(APA, 2014) para o Transtorno Depressivo Maior quanto os itens do *Beck Depression Inventory* – BDI (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Especificamente, o DSM-5 foi utilizado como base para a categorização dos itens em termos dos sintomas que representam, enquanto que o BDI foi utilizado como modelo para orientar no desenvolvimento do enunciado dos itens. Foram elaborados, ao todo, 150 itens.

Em seguida os itens foram submetidos à uma análise teórica, a qual foi organizada em duas fases. Na primeira, com objetivo de avaliar a pertinência de cada item com o construto estudado (depressão), foi realizada uma análise de juízes (três especialistas na área). Nessa fase, foram excluídos 14 itens. Na segunda fase, foi realizada a análise semântica para avaliar a inteligibilidade dos itens remanescentes para a população alvo. Nessa situação os itens foram apresentados a dois grupos focais que tiveram por intuito avaliar a clareza e a compreensão de cada um dos itens (Pacico, 2015). Nessa análise, a grande maioria dos itens mostrou-se clara e compreensível para os participantes. Somente em quatro itens foram apontadas algumas dificuldades de compreensão, os quais foram alterados em concordância com os participantes.

Após essa análise, os itens puderam, enfim, partir para a próxima etapa dessa pesquisa que foi o pré-teste. Para esse propósito, os itens foram divididos em dois conjuntos (Conjunto A e Conjunto B). A ideia era que cada participante respondesse somente metade dos itens do banco, diminuindo, dessa forma, o efeito de fadiga. Para tentar controlar a equivalência dos questionários quanto a cada critério diagnóstico, os itens foram organizados de acordo com o critério que representavam, sendo em seguida alocados de forma aleatória a um dos questionários. Complementando esse procedimento, os questionários foram organizados de forma alternada e aplicados ao mesmo tempo na população-alvo para que cada sujeito tivesse a mesma chance de responder qualquer uma das duas versões. Tal procedimento foi realizado para satisfazer o tipo de equalização desejada (equalização via população), a qual exige, apenas, que os itens dos dois questionários sejam calibrados simultaneamente (Andrade et al., 2000).

Participantes

Utilizou-se uma amostra de conveniência de 501 participantes das cidades de Parnaíba-PI e João Pessoa-PB, sendo que 251 responderam ao Conjunto A e 250 ao Conjunto B. A maior parte era do sexo feminino (61,9%), solteiros (58,5%) e com idades entre 18 e 76 anos ($M=30,09$, $DP=12,01$). Ressalta-se ainda que 58,5% tinham uma renda de no máximo um salário mínimo e 72,9% tinham pelo menos o ensino médio completo. Quanto ao diagnóstico de depressão, 6% informaram ter sido diagnosticados em algum momento nos últimos 12 meses. Destes, 63,3% tomavam medicamento controlado e apenas 50% afirmaram receber atendimento psicológico e/ou psiquiátrico.

Instrumentos

Foram utilizados dois Questionários designados “Questionário A” e “Questionário B”. Cada um deles composto por um conjunto de itens e por perguntas sociodemográficas. Os participantes deveriam responder a um dos questionários (Conjunto A ou Conjunto B), bem como às perguntas sociodemográficas.

Conjunto A – composto por 68 itens e equivalente à metade dos itens totais remanescentes da análise de conteúdo. Todos os itens foram respondidos pelos participantes em termos de concordância, tendo como base as últimas duas semanas. Para isso, os itens foram precedidos do enunciado “Nas últimas duas semanas...” e respondidos em uma escala de resposta Likert de 5 pontos que varia de 0 a 4, sendo 0 “discordo totalmente” e 4 “concordo totalmente”.

Conjunto B – similar ao Conjunto A – composto por 68 itens – esse também foi respondido pelos participantes em termos de concordância, utilizando-se da mesma escala de resposta. A diferença ficou por conta dos itens utilizados que consistiu na outra metade de itens remanescentes da análise de conteúdo.

Procedimento

A aplicação dos questionários aconteceu tanto de forma individual quanto coletiva, sendo as respostas exclusivamente individuais. Os participantes foram abordados em suas residências e em ambientes de grande circulação de pessoas, como shoppings, praças e universidades. As recomendações da Resolução n. 510/16 do Conselho Nacional de Saúde que regulamenta pesquisas com seres humanos foram seguidas. Os sujeitos assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) para iniciar a pesquisa e, foi informado, que os mesmos poderiam abandonar a pesquisa em qualquer momento. O projeto desta pesquisa foi aprovado pelo Comitê de Ética do Centro de Ciências da Saúde da Universidade Federal da Paraíba com o protocolo n.º 55307216.3.0000.5188.

Análise dos Dados

Foram utilizados os softwares *Statistical Package for the Social Sciences* (SPSS, versão 23.0), o *Factor* (versão 10.3.01; Ferrando & Lorenzo-Seva, 2013) e o *R* (versão 3.3.2; *R Development core Team*, 2011). Enquanto que o primeiro foi usado para descrever a amostra, utilizando-se de estatísticas descritivas (frequência, média e desvio padrão), o segundo foi empregado para verificar a dimensionalidade dos conjuntos de itens por meio do método *Hull – Comparative Fit Index* (CFI; Lorenzo-Seva, Timmerman, & Kiers, 2011), e estimar a consistência interna desses com o alfa de Cronbach. O *R*, por sua vez, foi usado para realizar análises fatoriais exploratórias com o estimador WSM (*weighted least squares*) e para estimar os parâmetros dos itens, os quais foram testados por meio do GRM (Samejima, 1969). Para essas duas últimas

análises foram considerados os pacotes *psych* (Revelle, 2014) e *ltm* (Rizopoulos, 2006), respectivamente.

Resultados

Análise Fatorial e Consistência Interna

As análises dessa seção foram realizadas para verificar os pressupostos de unidimensionalidade e independência local dos conjuntos de itens. Especificamente, foi utilizada uma das técnicas mais empregadas para esse propósito, a análise fatorial (AF). Adicionalmente também foi verificada a consistência interna por meio do alfa de Cronbach.

Para o conjunto A, inicialmente buscou-se verificar se a matriz de correlações dos itens era adequada para a realização de uma AF. Os valores encontrados tanto para o KMO (0,924) quanto para o Teste de Esfericidade de Bartlett (910733,8 [df: 2278; $p < 0,001$]) mostraram-se apropriados para esse propósito (Tabachnick & Fidell, 2013). Na sequência, para se obter evidências da dimensionalidade do conjunto de itens, utilizou-se o método *Hull – Comparative Fit Index* (CFI). Esse método indicou a estrutura unifatorial como a melhor solução. Considerando esse resultado, realizou-se uma AF exploratória baseada em correlações Policóricas, utilizando o método de estimação *Weighted least squares* (WLS). Os resultados dessa análise apontaram que os itens A5, A17, A18, A22, A40 e A54, não apresentaram carga fatorial suficiente para serem retidos no fator geral, a saber o valor mínimo de 0,35. Esse ponto de corte baseou-se no tamanho amostral, visto que para amostras de 250 participantes, cargas iguais ou acima de 0,35 são consideradas significativas (Hair, Black, Babin, Anderson, & Tatham, 2009). Por esse motivo, os itens foram excluídos e solicitada uma nova análise, na qual todos os itens apresentaram cargas fatoriais acima do ponto de corte estabelecido no fator geral (0,35), com valores variando entre 0,37 (itens A24, A47 e A50) e 0,85 (itens A56 e A60). Este fator, denominado Depressão Maior, obteve um valor próprio de 20,51, o qual explicava 41% da variância total. Por último, para verificar a consistência interna do fator considerou-se o alfa de Cronbach para correlações policóricas, o qual apresentou um valor de 0,93.

Em relação ao Conjunto B, foram empregadas as mesmas análises do conjunto anterior. Assim, inicialmente buscou-se verificar matriz de correlações, a qual se mostrou adequada para realizar uma AF (KMO=0,934; Teste de Esfericidade de Bartlett=10664,0 (df=1953; $p < 0,001$). Em seguida obteve-se evidências da dimensionalidade desse conjunto de itens. Utilizou-se o método *Hull-CFI* que assim como no conjunto anterior, apresentou a estrutura unifatorial como a melhor solução. A primeira AF apontou que os itens B1, B14, B15, B42, B51 e B59 não apresentaram cargas fatoriais suficientes (0,35) para retenção no fator geral. Tais itens foram excluídos e solicitada uma nova análise. Todos

os itens apresentaram cargas fatoriais acima de 0,35, no fator geral variando entre 0,38 (item B05) e 0,86 (item B49). O fator, denominado “depressão maior”, obteve um valor próprio de 22,8, o qual explicou 46% da variância total. O alfa de Cronbach do fator apresentou um valor de 0,96. Satisfeito o pressuposto de unidimensionalidade, os itens puderam, enfim, seguir para a calibração.

Calibração dos Itens

No Conjunto A, para a primeira estimação foram considerados todos os itens retidos na AF. Foi estabelecido

como ponto de corte para avaliar a discriminação dos itens valores a iguais ou acima de 0,60 (Nakano, Primi, & Nunes, 2015). Dentre todos os itens da primeira análise apenas o item A21 não apresentou discriminação satisfatória. Após a retirada desse item, o modelo foi testado novamente e nesta oportunidade todos os itens atingiram a qualidade esperada, com valores a entre 0,613 (item A34) e 2,347 (item A57). Assim partiu-se então para a análise das Curvas Características dos Itens (CCIs) de cada um dos itens. Os gráficos demonstraram quatro padrões de disposição das categorias (curvas), os quais podem ser visualizados na Figura 1.

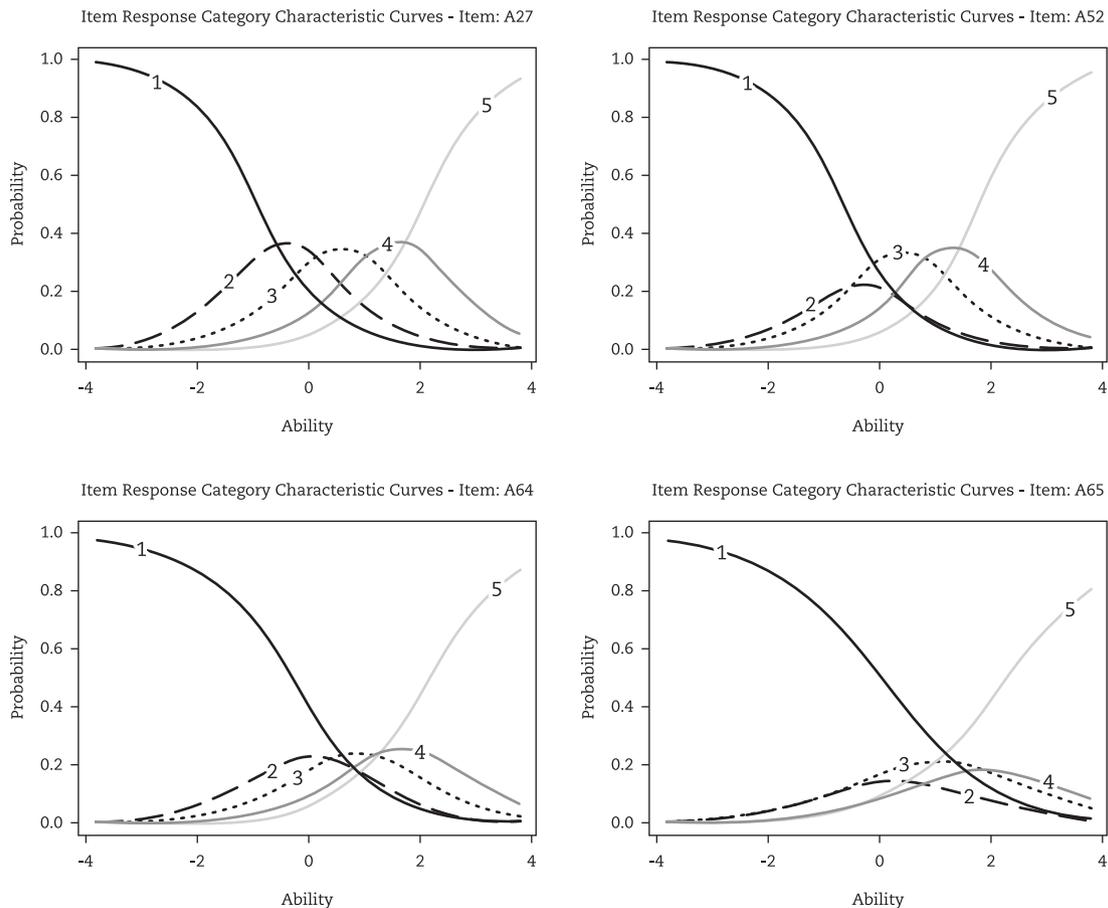


Figura 1. Padrões de disposição das curvas do Conjunto A

O primeiro padrão (e.g. item A27 – superior esquerdo) consistiu dos itens em que todas as categorias de resposta em algum ponto da escala de θ (gravidade da depressão) tinha maior probabilidade de endosso do que as demais. O item A52 (superior direito) exemplifica o segundo padrão encontrado. Este correspondeu aquele no qual somente uma categoria foi completamente sobreposta pelas demais. O terceiro padrão está representado pelo item A54 (inferior esquerdo), e se configura nas CCIs em

que apenas uma das categorias de resposta, dentre as três categorias intermediárias (1, 2 e 3), apresentava probabilidade de endosso maior que as demais em algum ponto do nível de θ . Por último, o quarto padrão (e.g., item A65 – inferior direito) representa os itens em que as três categorias de resposta intermediárias foram sobrepostas pelas categorias das extremidades (valores 0 e 4).

Os itens similares ao item A27 foram considerados bons e, conseqüentemente não foram alterados. Já em

relação aos padrões em que pelo menos uma das categorias não teve probabilidade maior de ser endossada em relação às demais, em nenhum ponto de θ apontam para o fato de que alguns itens podem funcionar melhor se forem retiradas algumas categorias. Desse modo, o próximo passo foi redefinir as categorias da escala de resposta. Foram realizadas basicamente três tipos de recategorizações. Nos itens que apresentavam o segundo padrão foram feitos dois tipos de recategorizações, que consistiram na junção da categoria do meio (2) à categoria 1 ou à categoria 3. A escolha por qual das categorias seria unida foi baseada na sobreposição das curvas. No item A57, por exemplo, as categorias 2 e 3, foram recategorizadas para serem uma só, totalizando ao final quatro categorias nesse item. Em relação aos itens que obedeciam ao terceiro e quarto padrões, todas as categorias intermediárias foram recategorizadas em uma só,

totalizando três categorias para esse item. Depois disso, o banco passou a apresentar itens com 3, 4 e 5 categorias de resposta.

A análise do modelo recategorizado indicou que os itens A24 e A45 não apresentavam discriminação suficiente para serem retidos e o item A4, mesmo recategorizado, teve uma de suas categorias de resposta completamente sobreposta pelas categorias das extremidades. Os três itens citados foram eliminados e realizada outra estimação. Todos os itens apresentaram discriminação satisfatória ($a > 0,60$) e todas as categorias de resposta dos itens foram bem representadas. As estimativas dos parâmetros abrangeram uma ampla porção do θ , apresentando valores b (*thresholds*) entre -2,013 (b_1 do item A48) e 3,717 (b_2 do item B67). Na Figura 2, são apresentados os padrões de disposição das curvas após a recategorização.

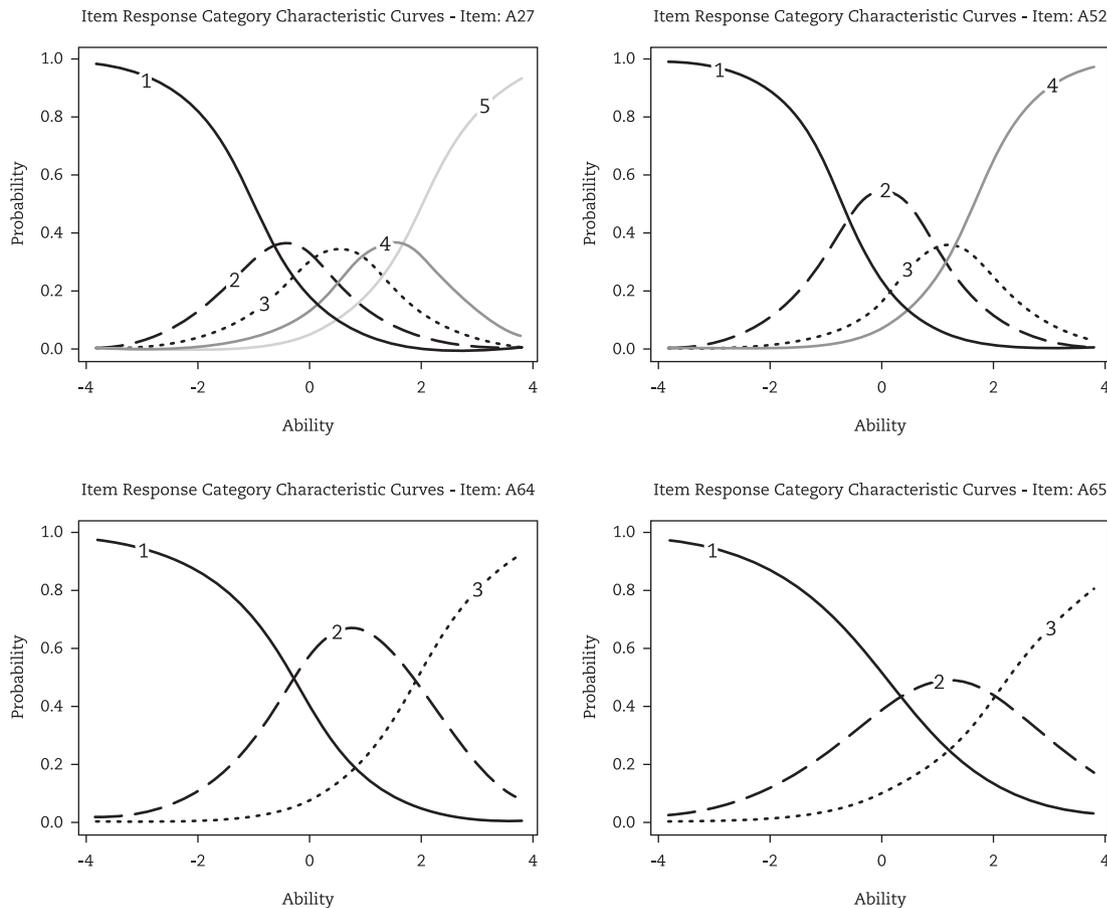


Figura 2. Padrões de disposição das curvas do Conjunto A após a recategorização

É possível observar na Figura 2 que houve uma melhora significativa no modelo. Nenhum dos itens apresentou curvas completamente sobrepostas. Além das

CCIs, a TRI também fornece as Curvas de Informação dos Itens (CIIs) e a Curva de Informação do Teste (CIT). Essa última pode ser visualizadas na Figura 3.

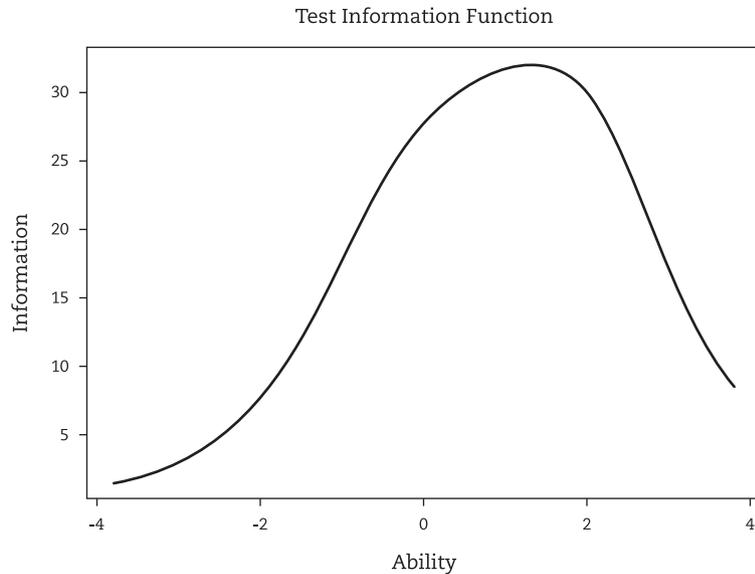


Figura 3. Curva de Informação do Teste do Conjunto A

Quanto ao conjunto B, a calibração seguiu os mesmos passos realizados com o conjunto A. Desse modo, inicialmente procedeu-se à primeira estimação dos itens. A partir dessa análise pôde-se constatar que os itens B5, B13 e B60 não tiveram discriminação satisfatória ($a > 0,60$). Estes foram retirados e o modelo foi novamente testado. Desta vez todos os itens apresentaram discriminação satisfatória com valores a variando entre 0,639 (item B67) e 2,439 (item B67). Em relação às CCIs desse modelo, foram observados os mesmos padrões encontrados na organização das categorias do conjunto A. Assim, algumas categorias de

respostas de itens foram recategorizadas. Após esse procedimento, os parâmetros dos itens do modelo foram novamente estimados.

Todos os itens apresentaram discriminação satisfatória com valores a entre 0,60 (Item B18) e 2,494 (Item B49) e não tiveram suas categorias totalmente sobrepostas. As estimativas dos parâmetros dos itens abrangeram uma ampla porção do θ , apresentando valores b entre -1,775 ($b1$ do item B4) e 3,717 ($b2$ do item B67). Houve uma melhora significativa nos itens, de modo que, no novo modelo, não houveram curvas completamente sobrepostas. A Figura 4 mostra a CIT do conjunto B.

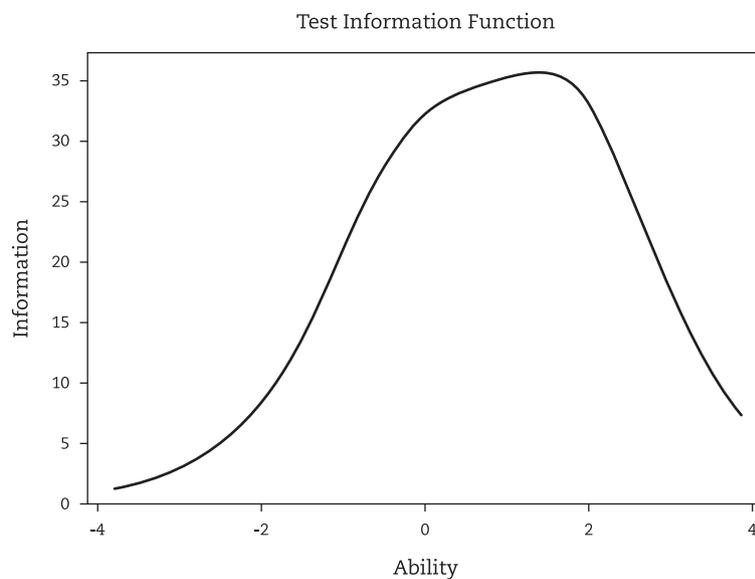


Figura 4. Curva de Informação do Teste do Conjunto B

A soma da informação dos itens compreendeu uma faixa de *theta* muito semelhante àquela mostrada no Conjunto A (entre -2 e 4) com a maior parte da informação concentrada aproximadamente entre 0 e 2. Nesse ponto a informação do conjunto de itens é, ainda, maior que a informação concentrada no mesmo ponto no Conjunto A. Assim como aconteceu na CIT do Conjunto 1, nessa também houve um pico entre 1 e 2 onde a informação é maior que em toda a faixa de *theta*.

Discussão

O presente estudo partiu de uma necessidade da literatura nacional de contar com medidas mais adequadas para avaliar a gravidade do TDM. Para isso buscou-se utilizar técnicas mais robustas que àquelas comumente utilizadas nos processos de construção de medidas. Especificamente, empregaram-se análises da TRI, as quais tem apresentado vantagens comparadas às técnicas baseadas na TCT.

A TRI, diferente da teoria tradicional, tem como propósito a construção de bancos de itens. No Brasil, não foi encontrado nenhum banco de itens calibrado com o objetivo de avaliar o TDM. O único banco disponível, é um conjunto de itens que teve sua adaptação apenas a nível transcultural, não disponibilizando informações sobre as características psicométricas dos itens (Castro et al., 2014). Assim, devido essa carência e pela relevância do problema, o objetivo dessa pesquisa foi construir um banco de itens para avaliar o TDM. Com base nas análises realizadas e nos seus respectivos resultados confia-se que esse objetivo tenha sido satisfeito.

A pré-testagem dos itens consistiu em duas etapas importantes: a verificação do pressuposto da unidimensionalidade e independência local dos conjuntos de itens e a estimação dos parâmetros dos mesmos. A unidimensionalidade e a independência local são dois pressupostos básicos para realizar análises de TRI, de modo que satisfazê-los foi necessário para prosseguir com as análises posteriores. Para isso foram empregadas análises fatoriais baseadas em estimadores adequados para escalas de resposta a nível categórico (WLS), acompanhadas do método Hull, uma técnica bastante robusta para verificar a retenção de fatores. Essas duas análises apoiaram a solução de um fator para os conjuntos de itens A e B, os quais apresentaram valores alfa excelentes (0,93 e 0,96, respectivamente; Pasquali, 2013).

Esses dados, além de corroborarem a hipótese de unidimensionalidade dos dois conjuntos de itens, também sustentam o pressuposto de independência local, visto que, como assegura Hambleton e Swaminathan (1985), esse pressuposto estaria implicado no primeiro. Além disso, os resultados das análises somam evidências de validade de construto, a forma mais fundamental de validade de medidas psicológicas (Pasquali, 2003). Quanto aos valores dos alfas, esses também

podem ser usados para demonstrar a precisão dos conjuntos de itens, que nos dois casos foram considerados muito bons. Esses achados indicam que os itens retidos são explicados por uma única variável latente, a depressão. Os resultados foram consistentes com o esperado, visto que os itens foram elaborados para representar adequadamente esse construto por meio dos seus diferentes sintomas.

A estimação dos parâmetros dos itens foi, sem dúvidas, a parte mais importante desse estudo. Trata-se de um procedimento de análises de dados baseado na TRI, que fornece muito mais informação para os itens, de forma individual, do que o seria feito com a metodologia clássica. O modelo usado foi o GRM, que foi escolhido pela sua capacidade de utilizar itens com categorias de respostas variados, acreditando-se que, como realmente ocorreu, os itens careceriam de redefinições nas suas categorias de resposta.

As modificações realizadas proporcionaram melhorias nas estimações feitas; primeiramente no que diz respeito às modificações sofridas pela retirada de itens que não apresentavam discriminação satisfatória antes da recategorização e após as recategorizações. A redefinição das categorias de resposta da maioria dos itens foi a maior alteração realizada nos conjuntos de itens. Isso foi feito devido algumas dessas categorias não apresentarem probabilidade de endosso maior que as demais em nenhum ponto da escala *theta*. Em geral, durante a tabulação dos dados pôde-se perceber que era comum muitos dos sujeitos endossar apenas a categoria do meio (2) e as categorias das extremidades (0 e 4) da escala de resposta. Acredita-se que esse problema esteja relacionado à própria natureza de alguns sintomas da depressão, já que itens relacionados a humor deprimido, fadiga e perda de energia, por exemplo, necessitaram de muito menos alterações do que àqueles referentes a comportamentos suicidas, alterações de peso e apetite, nos quais, praticamente, todos os itens precisaram ser recategorizados.

No estudo de Castro, Trentini e Riboldi (2010), foi verificado um problema semelhante. Foi identificado que pelo menos uma das categorias de resposta de 13 itens do BDI não apresentavam, em nenhuma faixa de *theta*, probabilidade de endosso maior do que as outras categorias. Os achados desses autores, bem como os que foram encontrados na presente pesquisa sugerem que esse problema também pode ocorrer em outras medidas de depressão. Nesse caso, seria interessante que essas medidas também tivessem seus parâmetros estimados por meio de modelos da TRI, buscando adaptar suas escalas de resposta para compor o número de categorias que representassem de forma mais adequada cada item ou sintoma.

Após as modificações realizadas pôde-se avaliar os itens de forma individual quanto aos seus parâmetros e às suas curvas de informação no que diz respeito

aos dois conjuntos. A avaliação dos itens em relação a quantidade de informação e o nível do *theta* mensurado, demonstram a grande vantagem que os modelos da TRI apresentam em comparação à TCT em análises desse tipo, possibilitando identificar os sintomas depressivos com maior peso para avaliar a gravidade da depressão (Castro et al., 2010).

Os itens referentes a humor deprimido foram os mais discriminativos com valores de parâmetro *a* variando a entre 0,72 (item A11) e 2,382 (item A57), média de 1,665. As CIIs dos itens que representam esse sintoma, em geral mostraram que os mesmos apresentam bastante informação em uma ampla faixa de *theta* que vai de -2 a 3. Os itens que apresentaram menores índices de discriminação foram aqueles relacionados a alterações no peso e no apetite que apresentaram valores *a* entre 0,618 (item A34) a 1,236 (item B45), média de 0,858; seguidos dos itens referentes a alterações do sono com valores entre 0,646 (item A20) e 1,36 (item B36), média de 0,994. A avaliação das CIIs dos itens referentes à alteração do sono apresentou curvas com menos informação, mas com uma ampla faixa de *theta* (entre -2 e 3). Já as CIIs dos itens referentes a alterações de peso, embora também tenham apresentado pouca informação em uma ampla faixa de *theta*, houve uma maior concentração de informação entre 0 e 3.

O fato das CIIs desses sintomas abrangerem grandes faixas de *theta*, principalmente o humor deprimido, evidencia que os mesmos são precisos para avaliar sujeitos que apresentam graus diversos de depressão. Todavia, o mesmo não aconteceu com os itens referentes à ideiação suicida. A CIIs desses itens, em geral, apresentam mais informação e se concentraram em uma faixa de *theta* menor, entre 1 e 3. Junto aos itens referentes ao sintoma de alterações de peso ou apetite, os sintomas de ideiação suicida foram considerados os mais graves, mostrando assim, que são mais precisos para avaliar indivíduos com níveis mais graves de depressão e pouco serviriam para avaliar níveis mais baixos da doença. Esses achados já eram esperados e já tinham sido observados por Castro et al. (2010) em seu estudo com o BDI.

Essa característica, encontrada, sobretudo, nos itens de ideiação suicida foi considerada por Gibbons et al. (2012) em um dos poucos estudos de desenvolvimento de CATs para depressão encontrados na literatura. Esses pesquisadores desenvolveram um algoritmo específico para os itens de suicídio, que consistia na administração de um item dessa categoria ao final da avaliação, quando nenhum desses itens fosse previamente apresentado. Caso o item fosse endossado positivamente, o avaliador recebia um relatório de alerta de suicídio gerado pelo teste.

Em relação às CITs, os resultados foram bastante similares nos dois conjuntos de itens, embora tenham apontado uma leve vantagem do Conjunto B em termos de informação. O conjunto B compreendeu uma faixa

de *theta* extensa entre -2 e 4, com informação um pouco mais direcionada para região do lado direito da curva. Isso indica que o banco fornece mais informação para níveis de moderados a grave do TDM, o que é bastante normal, uma vez que os itens foram construídos com o propósito de avaliar a gravidade da doença.

Esse resultado sugere que a composição final do banco cobre uma faixa ampla da gravidade da depressão, a qual pode até ser maior que a faixa coberta pelas medidas tradicionais baseadas na TCT. Embora não tenham sido feitas investigações nesse sentido, a crença nessa possibilidade sustenta-se em achados anteriores, como o estudo de Olini et al. (2013), no qual esses pesquisadores compararam medidas baseadas no PROMIS-D com duas medidas desenvolvidas com a TCT, o BDI e a *Center for Epidemiological Studies Depression Scale*. Os resultados mostraram que as medidas baseadas no PROMIS-D (versões de 8 e 28 itens) avaliaram a gravidade da depressão com maior precisão e maior amplitude do que as medidas comparadas. A versão de 28 itens do PROMIS-D forneceu, inclusive, mais que o dobro de informação que as demais.

Há de se destacar, entretanto, que assim como em todo e qualquer empreendimento científico, este estudo também está passível de limitações. A primeira delas diz respeito ao tamanho da amostra, o qual buscou-se apenas satisfazer as análises estatísticas empregadas. Assim, embora tenha-se considerado uma amostra um pouco maior (250 sujeitos) que o tamanho mínimo indicado pela literatura para realizar análises da TRI (200 sujeitos; Nunes & Primi, 2005), destaca-se que amostras maiores gerariam estimativas mais confiáveis. No estudo realizado por Nunes e Primi (2005), por exemplo, embora estes autores tenham evidenciado que amostras com 200 participantes apresentam resultados muito próximos àqueles gerados com amostra bem maiores, tais resultados são mais precisos para estimar os parâmetros de dificuldade (*b*) e a habilidade dos sujeitos (*theta*), mas menos eficazes para os parâmetros de discriminação (*a*).

Quanto ao pressuposto de unidimensionalidade, uma limitação importante a se destacar é que tal característica foi investigada em dois conjuntos de itens de forma separada. Esse procedimento foi empregado para que as pessoas que respondessem ao questionário não ficassem fadigadas ou entediadas ao ter que responder tantos itens e isso acabasse por enviesar as respostas dessas. A questão é que por conta desse procedimento, a inferência de unidimensionalidade do banco, contemplando os dois conjuntos, pode ser questionada. Para sanar esse problema, seria interessante, em uma nova pré-testagem, refazer a alocação aleatória dos itens nos conjuntos, ao invés de utilizar a mesma alocação feita nessa pesquisa.

É importante destacar, ainda, que, embora, a recategorização seja um procedimento válido para tratar a qualidade dos itens, tendo sido, inclusive, empregada

em uma diversidade de pesquisas, a recalibração ainda é a melhor opção. Isso por que não se pode garantir, por exemplo, que um indivíduo que endossou uma determinada categoria escolheria a nova categoria a qual foi unida na recategorização. Isso pode ocorrer, principalmente em relação aqueles indivíduos que endossaram as categorias intermediárias 1 e 3, devido à proximidade com as categorias das extremidades (0 e 4). Outro problema da recategorização é a tendência que tem de inflar as novas categorias.

Assim, sugere-se que esses itens sejam recalibrados em uma outra oportunidade, considerando as escalas de resposta como tratadas depois de recategorizadas e rotulando essas novas categorias. A rotulação será importante para reduzir o viés da falta de entendimento da escala de resposta, já que, como advertido por Nunes et al. (2008) a utilização de rótulos apenas nas extremidades pode favorecer a pouca utilização das categorias de resposta intermediárias.

Por fim, embora tenham sido apresentadas algumas limitações, como já elucidado, confia-se que o objetivo proposto tenha sido satisfeito. O banco de itens mostrou bons parâmetros psicométricos, estando o mesmo calibrado e pronto para ser utilizado para fornecer itens para compor testes curtos e personalizados e, sobretudo, para prover Testes Adaptativos Computadorizados. Com isso, espera-se que essa iniciativa contribua de forma significativa para a literatura nacional, estimulando a padronização dos procedimentos com medidas de autorrelato na pesquisa clínica, além de ajudar o(a) psicólogo(a) clínico a avaliar de forma rápida e precisa a resposta do paciente às intervenções empregadas e modificar os planos de tratamento (Cella, Gershon, Lai, & Choi, 2007). As limitações dos instrumentos diagnósticos utilizados atualmente impactam negativamente no tratamento do transtorno da personalidade. Avaliações bem embasadas cientificamente fomentam planos de intervenção mais eficazes, minimizando riscos para os pacientes.

Referências

- American Psychological Association [APA] (2010). *Practice Guideline for the Treatment of Patients with Major Depressive Disorder* (3ª ed.). Washington, DC: American Psychiatric Association.
- American Psychological Association [APA] (2014). *Diagnostic and statistical manual of mental disorders* (5ª ed.). Washington, DC: American Psychiatric Publishing.
- Andrade, J. M. de (2008). *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil* (Tese de doutorado em Psicologia Social, do Trabalho e das Organizações). Universidade de Brasília, Brasília. Recuperado de <http://repositorio.unb.br/handle/10482/1751>
- Anderson, R. J., Hoy, K. E., Daskalakis, Z. J., & Fitzgerald, P. B. (2016). Repetitive transcranial magnetic stimulation for treatment resistant depression: Re-establishing connections. *Clinical Neurophysiology*, 127(11), 3394-3405. doi: 10.1016/j.clinph.2016.08.015
- Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). *Teoria da Resposta ao Item: Conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística.
- Aybek, E.C., & Demirtasli, R. N. (2017). Computerized adaptive test (CAT) applications and item response theory models for polytomous items. *International Journal of Research in Education and Science*, 3(2), 475-487. doi: 10.21890/ijres.327907
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571. doi: 10.1001/archpsyc.1961.01710120031004
- Brunoni, A. R., Boggio, P. S., DeRaedt, R., Benseñor, I. M., Lotufo, P. A., Namur, V., & Vanderhasselt, M. A. (2014). Cognitive control therapy and transcranial direct current stimulation for depression: a randomized, double-blinded, controlled trial. *Journal of Affective Disorders*, 162(2014), 43-49. doi: 10.1016/j.jad.2014.03.026
- Castro, N. F. C., Rezende, C. H. A., Mendonça, T. M. S., Silva, C. H. M., & Pinto, R. M. C. (2014). Adaptação transcultural dos Bancos de Itens de Ansiedade e Depressão do Patient-Reported Outcomes Measurement Information System (PROMIS) para língua portuguesa. *Cadernos de Saúde Pública*, 30(4), 879-884. doi: 10.1590/0102-311X00096113
- Castro, S. M. J., Trentini, C., & Riboldi, J. (2010). Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. *Revista Brasileira de Epidemiologia*, 13(3), 487-501. doi: 10.1590/S1415-790X2010000300012
- Cella, D., Gershon, R., Lai, J. S., & Choi, S. (2007). The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*, 16(1), 133-141. doi:10.1007/s11136-007-9204-6
- Cherbuin, N., Kim, S., & Anstey, K. J. (2016). Late-life depression and dementia risk. *European Psychiatry*, 33(Supplement), S581. doi: 10.1016/j.eurpsy.2016.01.1707
- Cuijpers, P., Wit, L., & Taylor, A. (2014). The effects of psychological treatments for adult depression on physical activity: A systematic review. *Mental Health and Physical Activity*, 7(1), 6-8. doi: 10.1016/j.mhpa.2014.01.002
- Doty, S. B., Haroz, E. E., Singh, N. S., Bogdanov, S., Bass, J. K., Murray, L. K., Callaway, K. L., & Bolton, P. A. (2018). Adaptation and testing of an assessment for mental health and alcohol use problems among conflict-affected adults in Ukraine. *Conflict and Health*, (2018), 12-34. doi: 10.1186/s13031-018-0169-6
- Fei, K., Benn, E. K., Negron, R., Arniella, G., Tuhim, S., & Horowitz, C. R. (2016). Prevalence of depression among stroke survivors: racial-ethnic differences. *Stroke*, 47(2), 512-515. doi: 10.1161/STROKEAHA.115.010292
- Fenoy, A. J., Schulz, P., Selvaraj, S., Burrows, C., Spiker, D., Cao, B., ... & Soares, J. (2016). Deep brain stimulation of the medial forebrain bundle: distinctive responses in resistant depression. *Journal of affective disorders*, 203(2016), 143-151. doi: 10.1016/j.jad.2016.05.064
- Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona. Recuperado de: <http://psico.fccep.urv.es/utilitats/factor/documentacion/technicalreport.pdf>

- Gorman, J. R., Su, H. I., Roberts, S. C., Dominick, S. A., & Malcarne, V. L. (2015). Experiencing reproductive concerns as a female cancer survivor is associated with depression. *Cancer*, 121(6), 935-942. doi: 10.1002/cncr.29133
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2012). Development of a Computerized Adaptive Test for Depression. *Archives General Psychiatry*, 69(11), 1104-1112. doi: 10.1001/archgenpsychiatry.2012.14
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (9ª ed.). Porto Alegre: Bookman.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Norwell: Kluwer Academic Publishers.
- Holt, R. I., De Groot, M., Lucki, I., Hunter, C. M., Sartorius, N., & Golden, S. H. (2014). NIDDK international conference report on diabetes and depression: current understanding and future directions. *Diabetes Care*, 37(8), 2067-2077. doi: 10.2337/dc13-2134
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364. doi: 10.1080/00273171.2011.564527
- Manseira, P. R. P., & Misaghi, M. (2015). Arquitetura de software para um Sistema de Gestão de Testes Adaptativos Computadorizado. *Produção em Foco*, 5(1), 1-25. doi: 10.14521/P2237-5163201500070001
- Nakano, T. C., Primi, R., & Nunes, C. H. S. S. (2015). Análise de itens e Teoria de Resposta ao Item (TRI). Em, Hutz, C. S., Bandeira, D. R. & Trentini, C. M. (Eds.), *Psicometria* (pp. 97-123). Porto Alegre: Artmed.
- Nunes, C. H. S. S., & Primi, R. (2005). Impacto do tamanho da amostra na calibração de itens e Estimativa de escores por teoria de resposta ao item. *Avaliação Psicológica*, 4(2), 141-153. Recuperado de <http://pepsic.bvsalud.org/pdf/avp/v4n2/v4n2a06.pdf>
- Nunes, C. H. S. S., Primi, R., Nunes, M. F. O., Muniz, M., Cunha, T. F., & Couto, G. (2008). Teoria de Resposta ao Item para otimização de escalas tipo Likert – um exemplo de aplicação. *Revista Iberoamericana de Diagnóstico y Evaluación e Avaliação Psicológica*, 1(25), 51-79. Recuperado de http://www.aidep.org/03_ridep/R25/R253.pdf
- Nyer, M., Farabaugh, A., Fehling, K., Soskin, D., Holt, D., Papakostas, G. I., ... & Mischoulon, D. (2013). Relationship between sleep disturbance and depression, anxiety, and functioning in college students. *Depression and anxiety*, 30(9), 873-880. doi: 10.1002/da.22064.
- Olino, T. M., Yu, L., McMakin, D. L., Forbes, E. E., Seeley, J. R., Lewinsohn, P. M., & Pilkonis, P. A. (2013). Comparisons across depression assessment instruments in adolescence and young adulthood: An item response theory study using two linking methods. *Journal of Abnormal Child Psychology*, 41(8), 1267-1277. doi: 10.1007/s10802-013-9756-6.
- Pacico, J. C. (2015). Como é feito um teste? Produção de itens. Em, Hutz, C. S., Bandeira, D. R. & Trentini, C. M. (Eds.), *Psicometria*. (pp. 55-70). Porto Alegre: Artmed.
- Pasquali, L. (2007). *Teoria de resposta ao item: Teoria, procedimentos e aplicações*. Brasília: LabPAM/UnB.
- Pasquali, L. (2013). *Psicometria: Teoria dos testes na psicologia e na educação* (5ª ed). Petrópolis: Editora Vozes.
- R Development Core Team. (2011). *A language an environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Revelle, W. (2014). *Psych: Procedures for personality and psychological research*. Recuperado de <http://cran.r-project.org/web/packages/psych/psych.pdf>
- Rizopoulos, D. (2006). LTM: An R package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, 17(5), 1-25. doi: 10.18637/jss.v017.i05
- Samejima, R. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No 17). Psychometric Society. Recuperado de <https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>
- Shiozawa, P., Fregni, F., Benseñor, I. M., Lotufo, P. A., Berlim, M. T., Daskalakis, J. Z., ... Brunoni, A. R. (2014). Transcranial direct current stimulation for major depression: an updated systematic review and meta-analysis. *International Journal of Neuropsychopharmacology*, 17(9), 1443-1452. doi: 10.1017/S1461145714000418
- Stanton, R., & Reaburn, P. (2014). Exercise and the treatment of depression: A review of the exercise program variables. *Journal of Science and Medicine in Sport*, 17(2), 177-182. doi: 10.1016/j.jsams.2013.03.010
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6a ed.). Needham Heights, MA: Allyn & Bacon.
- Tisi, G., Franzini, A., Messina, G., Savino, M., & Gambini, O. (2014). Vagus nerve stimulation therapy in treatment-resistant depression: a series report. *Psychiatry and Clinical Neurosciences*, 68(8), 606-611. doi: 10.1111/pcn.12166
- Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., ... Schmael, C. (2008). Measuring depression: Comparison and integration of three scales in the GENDEP study. *Psychological Medicine*, 38(2), 289-300. doi: 10.1017/S0033291707001730
- World Health Organization [WHO] (2015). *Facts sheets No 396 – Depression*. Recuperado de <http://www.who.int/mediacentre/factsheets/fs396/en/>
- Yan, F. N., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*, 26(3), 171-177. doi: 10.3969/j.issn.1002-0829.2014.03.010

recebido em setembro de 2017
aceito em dezembro de 2018

Sobre os autores

Hemerson Fillipy Silva Sales é psicólogo formado na UFPI. Possui título de Mestre em Neurociência e Comportamento pela UFPB. Atualmente é doutorando do Programa de Pós-Graduação de Psicologia Social da UFPB.

Josemberg Moura de Andrade possui o título de Doutor em Psicologia pela UnB. É professor Associado do Departamento de Psicologia Social e do Trabalho e professor do Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações, ambos da UnB.

Vivianne Freitas Oliveira Asfora é psicóloga formada na UFPB. Possui título de Mestre em Neurociência e Comportamento pela UFPB. Atualmente é servidora do Tribunal de Justiça de Pernambuco, onde atua no Núcleo de Conciliação e Justiça Restaurativa.