

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: validade convergente e estabilidade temporal

Makilim Nunes Baptista – Universidade São Francisco, Itatiba, São Paulo, Brasil

Hugo Ferrari Cardoso – Universidade Estadual Paulista, Bauru, São Paulo, Brasil

Juliana Oliveira Gomes – Universidade São Francisco, Itatiba, São Paulo, Brasil

Resumo

Os objetivos principais desta pesquisa foram avaliar evidências de validade convergente entre a Escala Baptista de Depressão (Versão Adulto) - EBADEP-A e o Inventário de Depressão de Beck – BDI-II, além de avaliar a estabilidade temporal da EBADEP-A por intermédio do teste e reteste em um período de um mês. Fizeram parte da pesquisa 173 universitários de uma amostra de conveniência com média de idade de 24,45 ($DP=8,45$), a maioria mulheres (87,9%). Após um mês, 65 participantes, a maioria mulheres (90,8%), com média de 21 anos ($DP=5,48$) responderam novamente os instrumentos. Os resultados demonstraram, de acordo com critérios internacionais, excelentes índices de correlação entre ambas as escalas, bem como no teste e reteste, demonstrando adequadas qualidades psicométrica da EBADEP-A também nesses quesitos, comprovando outros estudos já realizados. Esses resultados e as limitações do estudo também são discutidos.

Palavras-chave: Validade; Confiabilidade; Medidas estatísticas; Avaliação psicológica.

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: convergent validity and temporal stability

Abstract

The main purpose of this research was to evaluate evidence of convergent validity between the Escala Baptista de Depressão (Versão Adulto) – EBADEP-A and the Beck Depression Inventory II – BDI-II, and to assess the temporal stability of EBADEP-A through the test and re-test method in a one-month period. The present study included 173 college students with an average of 24.45 years of age ($SD=8.45$), mostly women. After a one month period, 65 participants, mostly women (90.8%), with an average of 21 years of age ($SD=5.48$) answered the instrument a second time. The results showed, according to international standards, high correlation indices between scales, and the test-retest as well, demonstrating adequate psychometric qualities of EBADEP-A also in these questions, confirming other previous studies. These results and the limitations of the study are also discussed.

Keywords: Validity; Reliability; Statistical measures; Psychological assessment.

Escala Baptista de Depresión (Versión Adulto) – EBADEP-A: validez convergente y estabilidad temporal

Resumen

Los objetivos principales de esta investigación fueron evaluar evidencias de validez convergente entre la Escala Baptista de Depresión (Versión Adulto) – EBADEP-A y el Inventario de Depresión de Beck – BDI-II, además de evaluar la estabilidad temporal de la EBADEP-A por medio del test-retest en un período de un mes. Participaron de la investigación 173 universitarios de una muestra de conveniencia con media de edad de 24,45 ($DP=8,45$), la mayoría mujeres (87,9%). Pasado un mes, 65 participantes, la mayoría mujeres (90,8%), con media de 21 años ($DP=5,48$) respondieron nuevamente a los instrumentos. Los resultados demostraron, de acuerdo con criterios internacionales, excelentes índices de correlación entre ambas las escalas, así como en el test-retest, demostrando adecuadas calidades psicométricas de la EBADEP-A también en esos puntos, comprobando otros estudios ya realizados. Esos resultados y las limitaciones del estudio también son discutidos.

Palabras clave: Validez; Confiabilidad; Medidas estadísticas; Evaluación psicológica.

A depressão tem sido investigada como uma condição bastante comum na população mundial, principalmente quando se consideram estudos epidemiológicos multicêntricos com amostragens significativamente representadas. Atualmente, ter um episódio de depressão durante a vida parece ser mais uma norma que exceção. O curso da depressão maior é extremamente variável e a sua detecção parece ser falha nos sistemas de saúde de muitos países, até mesmo aqueles considerados como desenvolvidos (Andrews, Poulton & Skoog, 2005; Patten, Bilsker & Goldner, 2008).

Os diversos instrumentos que avaliam/rastreiam sintomatologia depressiva podem ter características bem diferentes, variando desde a maneira com que

mensuram (escalas do tipo *Likert* x dicotômica), passando pelo número de descritores avaliados, chegando a diferenças realmente expressivas na porcentagem de itens presentes nas principais categorias de avaliação da doença. Como exemplo, Calil e Pires (1998) apresentam algumas comparações entre medidas de depressão, dentre elas a Escala de Avaliação de Depressão de Hamilton, Escala de Avaliação de Depressão de Montgomery-Asberg, Escala de Avaliação de Melancolia de Bech-Rafaelsen, Inventário de Depressão de Beck, Escala de Auto-avaliação de Depressão de Zung, Inventário de Auto-avaliação de Wakedield e a Escala de Avaliação de Depressão de Carrol. A porcentagem de itens que avaliam a categoria humor de cada escala variou de 8 a

30%, enquanto os sintomas vegetativos variaram de 18 a 35%, motores de 0 a 18%, sociais de 0 a 9%, cognitivos de 0 a 52% (a maior variação encontrada), ansiedade de 0 a 17% e, por último, de irritabilidade, de 0 a 8%.

Outro detalhe interessante se refere às questões culturais relativas à expressão de sintomas da depressão. Como citam Bernstein (2006) e Spielberger, Ritterband, Reheiser e Brunner (2003), a cultura pode ser fundamental na expressão de alguns sintomas específicos em detrimento de outros. Logo, conclui-se pela necessidade de adaptação das escalas internacionais e/ou criação de escalas nacionais com linguagem apropriada e maior quantidade de descritores, sem desconsiderar o fato de que alguns deles devem ser tidos como obrigatórios, por serem mais universais do que outros, como por exemplo, os sintomas afetivos (humor)

Como apontam DPhil, Sheldon e House (2008), as escalas de rastreamento para a depressão são simples, rápidas e constituem uma forma barata de detectar possíveis casos e, a partir daí, possibilitar intervenções adequadas, já que o tratamento pode variar bastante entre um paciente que apresenta pontuações altas na categoria somática e outro que apresenta as mesmas pontuações na categoria cognitiva. Entretanto, independentemente das várias medidas internacionais encontradas para avaliar depressão, as escalas de rastreamento devem ser utilizadas com cautela, já que geralmente não coletam ou avaliam informações suficientes para o diagnóstico, tais como duração dos sintomas, critérios de exclusão ou diagnóstico diferencial, áreas mais afetadas da vida, dentre outras informações relevantes.

O processo de avaliação psicológica não envolve apenas a medição e a interpretação de resultados isolados de instrumentos. Uma vez que o clínico precisa integrar as várias informações que possui, provindas de diversas fontes e métodos de mensuração, também faz parte do processo de avaliação a seleção do tipo de escala adequada e específica, e isso inclui o discernimento entre escolher um ou outro tipo de instrumento de medida (Klein, Dougherty & Olino, 2005). Contudo, conforme relatam DPhil e colaboradores (2008), a utilização crescente de escalas de rastreamento de depressão parece estar associada ao aumento modesto de depressão por parte dos clínicos, pelo menos em atenção primária e ambientes hospitalares nos Estados Unidos.

As diversas medidas disponíveis para a depressão possuem pontos fortes e fracos (Lutz, Stahl, Howard, Grissom & Joske, 2002). Em razão do seu aspecto multifatorial, o fenômeno denominado como depressão é heterogêneo, complexo e pode ser definido

de formas múltiplas, e os diferentes modelos explicativos se refletem nos instrumentos avaliativos, os quais, além de representarem diferentes perspectivas, variam em relação ao objetivo que se propõem e por qual ou quais modelos/manuais as medidas estão baseadas (Ruscio, Brown & Ruscio, 2009).

Nesse sentido, parece ser de suma importância o desenvolvimento de instrumentos de rastreamento capazes de discriminar, com a máxima acurácia, os possíveis casos na população. Instrumentos sensíveis e específicos podem ser de grande utilidade para essa finalidade, já que a sensibilidade é a capacidade da escala em discriminar os casos que realmente possuem sintomas clínicos encontrados na patologia, enquanto à especificidade compete a discriminação daqueles que não possuem a patologia. Como apontam Nelson, Portera e Leon (2006), além da sensibilidade e especificidade, as escalas que avaliam sintomas relacionados a diagnósticos também devem ter outras qualidades psicométricas, tais como evidências de validade e confiabilidade.

Uma das evidências de validade relacionadas atualmente ao construto (Urbina, 2007), mas também reconhecida como convergente (AERA, APA & NCME, 1999), avalia a associação entre duas medidas que mensuram o mesmo construto como uma das formas de verificar se novas medidas avaliam o que se propõem. Geralmente é utilizado como parâmetro um instrumento já consagrado na literatura científica, também conhecido como "padrão ouro" (*gold standard*), para parametrizar o novo instrumento desenvolvido. Um padrão internacionalmente reconhecido para estimar os níveis adequados de correlação entre duas medidas similares é aquele preconizado por Prieto e Muñiz (2000), citando que correlações minimamente adequadas entre dois testes similares variam entre 0,4 e 0,5, correlações consideradas como boas variam entre 0,5 e 0,6 e correlações excelentes são aquelas acima de 0,6.

Outra medida psicométrica importante para as escalas diz respeito à precisão (erro da medida), e, dentre as várias possibilidades de se mensurar tal índice, enfatiza-se a estabilidade temporal, isto é, o quanto o teste mantém a precisão por algum período de tempo. Nesse quesito, Prieto e Muñiz (2000) também relatam que correlações minimamente adequadas variam entre 0,65 e 0,75, correlações consideradas como boas, entre 0,75 e 0,8 e índices excelentes estariam acima de 0,8.

Em relação aos instrumentos de rastreamento de depressão considerados como sendo de padrão ouro, o Inventário de Depressão de Beck (BDI) aparece como um dos mais considerados e utilizados no meio científico e clínico (Dowd, 2004; Myers & Winters, 2002). Desde 2001, o BDI foi lançado no Brasil após

diversos estudos de adaptação e validação do inventário e, mais recentemente, o BDI-II, que possui mudanças em alguns itens para adaptação da escala para o DSM-IV, também pode ser utilizado, após estudos de evidências de validade e precisão (Cunha, 2001; Goreinstein, Pang, Argimon & Werlang, 2011).

Como citado anteriormente, as medidas de depressão podem variar muito em diversos parâmetros, e a escolha da medida mais adequada, de acordo com o objetivo do clínico ou pesquisador, também faz parte do processo da avaliação psicológica (Klein & cols., 2005). No entanto, no Brasil, são poucas as opções de escolha de instrumentos que avaliam a sintomatologia depressiva, e não foi encontrado na literatura até agora pesquisada o desenvolvimento de um instrumento construído no país com diversas evidências de validade, para utilização em adultos. É com esse intuito que foi planejada a Escala Baptista de Depressão (versão Adulto) – EBADEP-A, com um conjunto inicial de 75 itens e 28 descritores para avaliar a depressão (vide descrição do instrumento no método), baseada em manuais de psiquiatria e teorias psicológicas sobre a depressão. Assim, dos itens iniciais, pesquisas com a versão inicial com os 75 itens, de 32, e a definitiva, com 45 itens (Baptista, 2012) foram desenvolvidas ao longo de quatro anos.

Independentemente da versão utilizada, diversas evidências de validade de construto (Urbina, 2007), também denominadas de conteúdo, construto e critério (AERA & cols., 1999), foram encontradas para o instrumento, que também já foi nomeada anteriormente como Inventário de Depressão Brasileiro (INDEBRA) e Escala de Depressão (EDEP), mas que será denominada neste estudo como EBADEP-A. Dentre algumas dessas pesquisas, Dias (2008), com uma amostra de 100 pacientes de um ambulatório com doenças inflamatórias intestinais, e seus acompanhantes encontraram diferenças de média na EBADEP-A para ambos os grupos, além de correlacioná-la com o *Inflammatory Bowel Disease Questionnaire (IBDQ)*, que avalia qualidade de vida em pacientes com doenças intestinais inflamatórias e a Bateria Fatorial de Personalidade (BFP), encontrando correlação negativa de 0,71 com o IBDQ. Quanto às correlações com as escalas da BFP, somente a relação com a Escala de Abertura não apresentou significância, tendo sido apresentadas correlações significativas com as escalas de Extroversão ($r=-0,32$), Socialização ($r=-0,23$), Realização ($r=-0,15$) e Neuroticismo (0,48), sendo esta última a que avalia comportamentos depressivos.

Baptista, Souza e Alves (2008) avaliaram a associação entre a EBADEP-A e o BDI, tendo como participantes da pesquisa 157 universitários de uma

cidade de Minas Gerais, e os resultados indicaram alta correlação significativa de 0,75 entre os instrumentos. A replicação desse estudo foi feita por Baptista, Carneiro e Sisto (2010), com uma amostra de 165 estudantes universitários de uma cidade do interior de São Paulo, obtendo-se como resultado correlação significativa de 0,68. Souza (2010) avaliou dados de estrutura interna da EBADEP-A, pela da Teoria de Resposta ao Item (TRI), além de análises de fidedignidade, sensibilidade e especificidade da escala. Foi utilizada uma amostra de 691 universitários, 40 indivíduos diagnosticados como depressivos pelos critérios do DSM-IV, por intermédio de entrevista diagnóstica de Transtornos do Eixo I (*Structured Clinical Interview for DSM Disorders – SCID-CV*) e 40 controles não-depressivos, confirmados também pela SCID-CV. Os principais resultados demonstraram diferenciação significativa entre os grupos critério; sensibilidade de 0,98 e especificidade de 0,88 para a escala; precisão de 0,97. A análise do DIF (*Differential Item Functioning*) demonstrou 14 itens favorecendo aos homens e nove favorecendo às mulheres, sugerindo equidade na tendencialidade dos itens.

Apesar de os resultados com a EBADEP-A serem bastante promissores, nem sempre instrumentos que avaliam o mesmo construto possuem correlações adequadas entre eles. Por exemplo, Lutz e colaboradores (2002) intentaram avaliar a correlação entre a escala Hamilton de Depressão (HAM-D), uma escala heteroaplicativa também considerada como padrão ouro, com as categorias de depressão de outras duas escalas menos utilizadas, denominadas de FOCUS e COMPASS PC DS, sendo a primeira um *checklist* com 34 itens que avaliam depressão e ansiedade e a segunda uma escala autoaplicativa que avalia severidade de problemas de saúde em pacientes de atenção primária, também contendo uma subescala de depressão. As amostras adotadas no estudo foram pequenas ($N=87$ e $N=80$) e as correlações entre a HAM-D e as subescalas das duas outras medidas foram de -0,42 e -0,43, sendo inversas nas pontuações, demonstrando que nem sempre medidas similares possuem correlações consideradas como adequadas, talvez pelas diferenças características de cada teste.

Por sua vez, os estudos com o BDI-II demonstraram diversas evidências de validade para o instrumento no Brasil, confiabilidade, sensibilidade e especificidade, além das análises fatoriais que geraram evidências de construto. Por exemplo, a confiabilidade apresentou variação de 0,78 a 0,83 em diferentes amostras, enquanto a sensibilidade foi de 0,70 e especificidade 0,84. Correlações com alguns outros instrumentos similares variaram de 0,67 a 0,89 com o *Self-Report Questionnaire (SRQ)*, de 0,75 com o

Montgomery-Asberg Depression Rating Scale (MADRS) em amostra de adultos e 0,66 com a *Hamilton Anxiety Scale* (HAM-A). Por último, a estabilidade temporal do BDI-II, por intermédio do coeficiente de correlação intra-classe, foi de 0,89 em 60 estudantes, com intervalo de 15 dias (Goreinstein & cols., 2011).

Isso posto, o objetivo do presente estudo foi analisar, com uma amostra de conveniência, a prevalência de sintomatologia depressiva em universitários, além de buscar evidência de validade convergente entre a EBADEP-A e o BDI-II, e analisar a estabilidade temporal da EBADEP-A. Análises secundárias de diferença de médias em relação a sexo, estado civil, faixa etária sobre o grupo que afirmou já ter ou não passado por avaliação psicológica/psiquiátrica e com os grupos que afirmaram ou não estar passando por algum tratamento para depressão também foram realizadas.

Método

Participantes

Foram participantes deste estudo 173 estudantes universitários de uma amostra por conveniência, com idade variando entre 17 e 61 anos, média de 24,45 ($DP=8,45$), com maioria do sexo feminino ($N=152$; 87,9%). A distribuição quanto ao estado civil foi de 81,5% solteiros, 15% casados e o restante distribuído em outras categorias (desquitados/divorciados, etc.). Em relação a ter passado por alguma avaliação psicológica/psiquiátrica, 7,5% responderam que sim e afirmaram terem tido diagnóstico de depressão na ocasião; 15% responderam que passaram por avaliação, mas sem diagnóstico e 14,5% passaram por avaliação e tiveram como resultado outro diagnóstico. Da amostra, 6,4% relataram que atualmente passaram por algum tratamento (psicológico e/ou medicamentoso) para depressão e 45,1% afirmaram ter casos de depressão na família. Após um mês da realização da pesquisa, os instrumentos foram aplicados novamente em uma amostra de 65 estudantes, com idade média de 21 anos ($DP=5,48$), variando entre 17 e 42 anos, a maioria mulheres ($N=59$; 90,8%). Cabe ressaltar que, em ambas as amostras, 31 participantes (equivalente a 17,9% para o primeiro grupo e 47,7% para o segundo) não responderam a idade.

Instrumentos

Inicialmente foi aplicado um questionário de identificação contendo informações sociodemográficas, tais como sexo e idade, além de dados sobre histórico de avaliação psicológica ou psiquiátrica e atual tratamento para depressão, para a caracterização da amostra, além de mais dois instrumentos padronizados, como se segue.

Escala Baptista de Depressão (Versão Adulto) – EBADEP-A (Baptista, 2012) – foi construída a partir de indicadores provenientes de manuais diagnósticos tais como o DSM-IV-TR (APA, 2002), CID 10 (OMS, 1993), além de descritores baseados na Teoria Cognitiva de Beck (Beck, Rush, Shaw & Emery, 1997) e Teoria Comportamental da Depressão (Ferster, Culbertson & Boren, 1977). Trata-se de um instrumento de autoaplicação para avaliação de sintomatologia depressiva, com 45 itens divididos em duas frases por item, em formato *Likert* de quatro pontos. A pontuação varia de zero a três, com mínima de zero e a máxima de 135, de forma que a menor pontuação indicaria menor sintomatologia, além de poder ser utilizada em uma ampla faixa etária, que varia de 17 a 81 anos de idade. Após estudos de evidências de validade de conteúdo, critério e construto, além de precisão, sensibilidade, especificidade e normatização em uma amostra de 1.676 participantes de grupos como depressivos, pacientes psiquiátricos, universitários, hospitalizados por doença física, dentre outros, os resultados demonstraram bons índices psicométricos. Segundo o modelo *Rasch*, os índices médios de ajuste *infit* e *outfit* (1,05 e 1,06) se mostraram adequados; a escala possibilitou discriminar os diferentes grupos-critério amostral; apresentou confiabilidade de 0,95 pelo alfa de Cronbach, além de sensibilidade de 77,5% e especificidade de 87,5%. Em termos de distribuição de itens e descritores em categorias de depressão, a EBADEP-A possui 18% dos itens na categoria social (incapacidade, queda de produtividade, dependência, esquivas sociais, etc.), 20% em humor (humor deprimido, choro, anedonia, autoestima rebaixada, etc.), 33% em cognitivos (desamparo, desesperança, indecisão, ideação suicida, etc.), 18% em vegetativos/somáticos (fadiga, alteração de apetite, sono, etc.), 4,5% em motores (agitação/lentidão), 4,5% em irritabilidade e 2% em ansiedade.

Inventário de Depressão de Beck – BDI-II (Goreinstein & cols., 2011) – versão revisada do BDI, com a troca de quatro itens para adaptação ao DSM-IV (agitação, desvalorização, dificuldade de concentração e falta de energia) e modificação de dois outros (apetite e sono). O instrumento é autoaplicativo, possui 21 itens, variando de quatro a sete opções de resposta (de 0 a 63 pontos, apontando mais sintomas depressivos quanto maior a pontuação) e é aconselhável desde adolescentes a partir de 13 anos de idade. O BDI-II possui diversas evidências de validade, além de estudos de confiabilidade, sensibilidade, especificidade e normatização. O modelo fatorial rotacionado demonstrou a existência de duas dimensões, sendo a

primeira denominada de Cognitivo-Afetivo e a segunda de Somato-Afetivo.

Procedimentos

Após a aprovação do projeto pelo Comitê de Ética e o contato com a direção e docentes para a aplicação dos instrumentos, os mesmos foram distribuídos aos discentes, na seguinte sequência: Termo de Consentimento Livre e Esclarecido (em duas vias), Questionário de Identificação (QI), EBADEP-A e BDI-II. Após passado um mês do primeiro encontro, os instrumentos foram novamente aplicados, na mesma ordem, sendo registrado um tempo médio de 40 minutos para o término da pesquisa pelos universitários.

Resultados

Em relação à estatística descritiva relativa à primeira aplicação, a pontuação média da EBADEP-A foi de 26,87 ($DP=23,43$), variando de zero a 109,

considerando 173 participantes, enquanto para o BDI foi registrada pontuação média de 7,6 ($DP=7$), com mínimo de zero e máximo de 34 pontos. No que se refere à segunda aplicação, realizada após um mês, ressalta-se que nem todos os participantes estavam presentes para responder à pesquisa, talvez pela coincidência entre a ocasião de realização dos procedimentos e o período imediatamente após as provas. Dessa forma, 65 protocolos foram respondidos. Nesse segundo momento, foi registrada pontuação mínima de zero e máxima de 109 pontos para a EBADEP-A, com média de 32,42 ($DP=22,56$), enquanto o BDI-II apresentou variação de zero a 29 pontos e média de 7,67 ($DP=6,02$).

Baseado nas pontuações normativas dos instrumentos, pode-se obter a distribuição em quanto à classificação da avaliação da sintomatologia depressiva por ambos os instrumentos, ou seja, sem sintomatologia (ou sintomatologia mínima), sintomatologia leve, moderada e severa (ou grave). Tais dados são apresentados na Tabela 1.

Tabela 1. Distribuição dos níveis de depressão por instrumento na amostra

Níveis de depressão	EBADEP-A		EBADEP-A		BDI-II		BDI-II	
	1ª aplicação		2ª aplicação		1ª aplicação		2ª aplicação	
	f	%	f	%	f	%	f	%
Sem sintomas	156	90,2	58	89,2	141	81,5	57	87,7
Leve	8	4,6	3	4,6	23	13,3	5	7,7
Moderada	9	5,2	4	6,2	5	2,9	1	1,5
Severa	-	-	-	-	4	2,3	2	3,1
Total	173	100	65	100	173	100	65	100

Como pode ser observado na Tabela 1, tanto na primeira quanto na segunda aplicação, a maior porcentagem de participantes acabou sendo enquadrado na categoria sem sintomas de depressão, havendo uma pequena variação entre a primeira e segunda aplicações da EBADEP-A em questão de porcentagens. Já no BDI-II, observou-se também mudanças, havendo leve diminuição de participantes com sintomatologia severa.

A análise de correlação foi realizada pelo teste de Pearson entre as duas escalas, bem como entre a primeira e a segunda aplicação, utilizando-se nível de significância mínimo de 5%. A correlação entre a EBADEP-A e o BDI-II foi de 0,69 ($p \leq 0,000$) na amostra total. Já a correlação entre a EBADEP-A e o BDI-II dos 65 participantes da segunda aplicação foi de 0,86 ($p \leq 0,000$) e a correlação entre a primeira e a segunda aplicações da EBADEP-A, isto é, com um mês de intervalo, foi de 0,80 ($p \leq 0,000$).

Por último, análises de diferença de médias foram efetuadas, a fim de se atender aos objetivos secundários do presente artigo. Foi realizada comparação entre os

11 participantes que afirmaram estar passando por tratamento para a depressão com o resto do grupo ($N=159$), isso porque, mesmo havendo uma grande diferença entre os grupos, os 11 indivíduos tiveram média maior na pontuação da EBADEP-A (média ranqueada de 35,27 x 26,64). A análise do teste *t* demonstrou não haver diferença significativa na pontuação da EBADEP-A ($t=-0,797$; $p=0,443$).

A segunda análise foi efetuada comparando-se as quatro possibilidades de respostas em relação à avaliação psiquiátrica, ou seja, "não ter passado por avaliação psiquiátrica/psicológica ($N=107$); ter passado por avaliação, mas sem diagnóstico ($N=26$); ter passado por avaliação mas com outro diagnóstico que não depressão ($N=25$) e ter passado por avaliação com diagnóstico de depressão ($N=13$). O resultado da prova da ANOVA não demonstrou diferença estatisticamente significativa entre os grupos ($F[3,167]=1,56$; $p=0,200$), indicando a média ranqueada de pontuação subindo conforme as categorias descritas anteriormente (24,29 x 28,12 x 33,24 x 34,77).

Ao se compararem os grupos dois a dois, não houve diferença entre o que afirmou nunca ter passado por avaliação (média ranqueada de 24,29) e o que relatou haver passado com diagnóstico de depressão ($M=34,77$; $t=-1,19$; $p=0,257$), e entre o grupo que não recebeu diagnóstico na avaliação ($M=28,12$) e o que passou por avaliação, porém com outro diagnóstico ($M=33,24$; $t=-6,47$; $p=0,521$). Não foram constatadas também diferenças em relação à presença de histórico familiar de depressão ($t=0,191$; $p=0,849$) para a EBADEP-A.

Por fim, foram realizadas análises de diferença de médias também para as variáveis relativas a sexo, estado civil e idade, sendo que, em relação a essa última, a amostra foi dividida em função de Quartis. Não foram encontradas diferenças estatisticamente significativas em relação a essas variáveis, tanto para o primeiro grupo quanto para o segundo, em nenhum dos instrumentos.

Discussão e considerações finais

Os objetivos principais deste estudo foram avaliar a evidência de validade convergente entre a EBADEP-A e o BDI-II e de verificar a estabilidade temporal da EBADEP-A. Como objetivos secundários foram avaliadas possíveis diferenças nos escores da EBADEP-A nos participantes que relataram estar passando por tratamento atual de depressão, bem como a existência de diferenças entre gênero, faixas etárias, estado civil e entre os grupos que relataram ter ou não passado por avaliação psiquiátrica/psicológica, sem diagnóstico de transtorno ou com diagnóstico, sendo o mesmo de depressão ou outros transtornos.

Um primeiro resultado, que não dependeu diretamente dos objetivos do trabalho, ou seja, um resultado colateral, esteve relacionado com a porcentagem de pessoas que apresentou sintomatologia de depressão por ambos os instrumentos (EBADEP-A e o BDI-II). Sabe-se que o Episódio Depressivo Maior (EDM) e/ou a sintomatologia de depressão são fenômenos muito comuns na população mundial. A Organização Mundial de Saúde (WHO, 2010) vem afirmando que esse fenômeno se configura em um dos principais problemas de saúde mental na população, atingindo mais de 120 milhões de pessoas no mundo, além do que, como afirma Stewart (2008), a depressão não é reconhecida como deveria ser por profissionais da saúde e, por isso, é pouco tratada. Ainda, é importante afirmar que instrumentos de rastreamento, como é o caso da EBADEP-A, podem ser importantes ferramentas na detecção da depressão, já que, como afirma Stewart (2008), a utilização de instrumentos de rastreamento está diretamente relacionada com o

aumento do seu reconhecimento na população, tendo como consequência o uso de tratamentos adequados nas pessoas que convivem com o problema.

Diversos são os estudos que demonstram o quão prevalente é a depressão, tendo como um dos exemplos um levantamento de estudos epidemiológicos internacionais e nacionais, de base populacional, realizado por Vercaro, Uchoa e Lima-Costa (2002), em que os autores compilam mais de 45 *surveys* na Europa, América do Sul, do Norte e Central, Ásia e Oceania. Como resultados, foi demonstrado que a prevalência de episódios depressivos nos últimos 30 dias variou entre 2,2% e 5,4%; no último ano, de 3,5% a 10,3% e, durante a vida, de 3,7% a 15,7%. Não levando em consideração o diagnóstico de Episódio Depressivo Maior (EDM), mas a sintomatologia depressiva, o presente estudo obteve resultados muito similares, como por exemplo, o estudo de Quintero, Garcia, Jiménez e Ortiz (2004), que também avaliaram níveis de sintomatologia depressiva em 218 universitários da Colômbia, utilizando, como o estudo atual, dois instrumentos de depressão (escala de Zung e o BDI). Como desfecho, apresentaram uma porcentagem de 75,2% e 69,7% de participantes sem depressão (Zung e BDI); 18,3% e 22,9% com sintomatologia leve; 5,5% e 6% para moderada e, 0,9% e 1,4% para severa. Já Andrade e Argimon (2006) encontraram, em uma amostra de conveniência de 706 adolescentes do Rio Grande do Sul, utilizando o BDI, 71,4% da amostra sem sintomatologia, 18,4% com sintomas leves, 9,1% moderado e 0,3% com sintomatologia grave.

No presente estudo, alguma diferenciação foi encontrada na discriminação de participantes com ou sem sintomatologia depressiva (nos diversos níveis), a partir do nível leve, já que a quantidade de participantes sem sintomatologia foi idêntica para ambas escalas. Essa variação de detecção nos níveis pode estar justamente relacionada com o as características de cada medida. Como apontam Calil e Pires (1998), a porcentagem de itens em cada categoria dos instrumentos de depressão (ex.: afetivo, cognitivo, somático) pode ser bem diferente avaliando o mesmo construto, mas de forma particular.

Complementando o raciocínio, Smar (2003), ao realizar uma descrição pormenorizada de cinco escalas que avaliam sintomatologia depressiva, encontrou várias diferenças entre as mesmas, no que diz respeito ao conteúdo dos itens e dimensões, número de itens, formas de correção, formato de resposta, método e tempo de administração e propriedades psicométricas. Retomando a questão, diferença na detecção de porcentagens nos vários níveis de depressão nos dois instrumentos utilizados nesta pesquisa pode também

estar ligada, além das variáveis já mencionadas, a quanto cada escala possui maior capacidade em discriminar sintomatologia mais branda ou severa. Baptista (2012), ao realizar um procedimento estatístico de transferência de normas entre a EBADEP-A e o BDI, pela Teoria de Resposta ao Item (TRI), já havia se reportado a essa questão ao encontrar que os dois instrumentos possuíam itens que avaliavam o mesmo fenômeno, porém variavam quanto à dificuldade/facilidade para discriminar o fenômeno, expresso pelo mapa de itens do modelo *Rasch*, indicando que uma das escalas mensurava com mais acurácia indivíduos com sintomas mais leves, diferentemente da outra, de acordo com o endosso que a mostra efetuava ao responder os itens das escalas.

Com relação à evidência de validade convergente entre a EBADEP-A e o BDI-II, o resultado encontrado pode ser considerado, segundo os critérios de Prieto e Muñiz (2000), como sendo excelente, corroborando outros achados que compararam a EBADEP-A com o BDI e a CES-D, efetuados por Baptista e colaboradores (2008, 2010) e Baptista, Souza, Gomes, Alves e Carneiro (2012). Todas as correlações encontradas foram consideradas excelentes, corroborando as adequadas qualidades psicométricas da EBADEP-A também nesse quesito, apesar do baixo número amostral da presente pesquisa. A evidências de validade baseada na comparação com outro instrumento, nesse caso também relatada como convergente, pode servir para avaliar se um instrumento recém-construído mensura o mesmo construto quando comparado a um instrumento já bem estabelecido em relação às propriedades psicométricas e utilização pela comunidade científica e clínica.

Nesse caso, o BDI-II, bem como o BDI, podem ser considerados como os dois dos testes mais utilizados mundialmente e estudados psicometricamente, como afirmam Vanheule, Desmet, Groenvynck, Rosseel e Fontaine (2008) e Whisman, Perez e Ramel (2000). Como resultado, a existência de correlações altas e significativas entre dois instrumentos que avaliam o mesmo construto permite afirmar a existência de similaridade entre eles, já que a validade convergente também pode ser definida como o grau em que duas medidas designadas para medir um construto são relacionadas (AERA & cols., 1999).

Referente à estabilidade temporal, também conhecida como confiabilidade pela técnica do teste e re-teste, Martins (2006) afirma que essa técnica avalia a constância dos resultados aplicados, nesse caso, com um intervalo de tempo entre as medidas. Esse tipo de propriedade psicométrica depende ainda da estabilidade do fenômeno e do espaço de tempo em que são realizadas as medidas. No caso, uma medida

fidedigna (outra denominação de confiabilidade) tende a oferecer consistência e precisão na medida em que pode garantir estabilidade entre diferentes mensurações com o mesmo sujeito, pela alta correlação.

Um tempo muito pequeno entre as aplicações (ex.: um ou dois dias) pode não ser adequado devido ao efeito da memória ou, em casos de medidas específicas (ex.: inteligência) do fator aprendizagem, em que o avaliando, ao realizar a prova, pode aprender com ela, aumentando a chance de uma pontuação maior na próxima aplicação da mesma prova. Quando o tempo é muito grande (ex.: dois meses), pode haver também uma subavaliação da estabilidade, principalmente quando o construto avaliado é um transtorno, já que, por exemplo, em um período de um mês, com a ingestão de medicamentos, realização de psicoterapia ou mesmo remissão espontânea, a sintomatologia depressiva pode diminuir consideravelmente. Como afirma Martins (2006), o pesquisador deve optar por um tempo maior no teste e reteste, pois isso tende a uma maior segurança no resultado do índice de correlação.

Nesse sentido, o resultado do teste e reteste da EBADEP-A, em um intervalo de um mês, gerou um índice de correlação bastante adequado e muito próximo daquele referenciado por Goreinstein e colaboradores (2011) em estudo de estabilidade temporal com o BDI-II, em um período de 15 dias. O índice encontrado nesse estudo com a EBADEP-A, pelo critério de Prieto e Muñiz (2000) pode ser considerado como excelente, já que está acima de 0,8, de acordo com os autores.

As análises secundárias apresentaram uma ausência de discriminação em relação a características da amostra como sexo, idade e estado civil. Quanto ao gênero, a presença majoritária de mulheres, em ambos os grupos, podem ter influenciado nas análises de diferença de média, embora seja uma característica comum em diversos estudos nos quais o grupo de mulheres também se destaca (Baptista & cols., 2008, 2010; Dias, 2008; Souza, 2010; dentre outros). Em relação à idade, analisada por meio da divisão por faixas etárias, a ausência de informações de 31 participantes também pode ter influenciado nas análises.

No que se refere aos participantes que afirmaram estar em tratamento para a depressão, bem como dos que afirmaram ter recebido diagnóstico de depressão ou algum outro transtorno psiquiátrico, ou, ainda, os que relataram possuir histórico familiar de depressão, tais dados não estão de acordo com o descrito por Baptista (2012), a partir do que pode-se demonstrar uma capacidade da EBADEP-A em discriminar diversos grupos critério em sua pontuação, tais como

os não depressivos dos universitários; os acompanhantes dos pacientes hospitalizados dos próprios pacientes; universitários de pacientes psiquiátricos e, por último, pacientes depressivos (confirmados pela SCID) em relação aos outros grupos, tais como pacientes psiquiátricos, pacientes hospitalizados (hospital geral), universitários, acompanhantes de pacientes hospitalizados (hospital geral) e não-depressivos (confirmados pela SCID).

Os resultados deste estudo, de maneira geral, apontaram características psicométricas adequadas da EBADEP-A, recomendadas pelos critérios internacionais nos quesitos validade convergente e confiabilidade teste e reteste. Martínez, García, Martínez, Mendoza e Ferrer (2007) apontaram que, por exemplo, na Espanha, os estudos demonstram inadequação conceitual de diversas escalas de depressão, ausência de normas de aplicação e, em alguns casos, insuficientes propriedades psicométricas das escalas, inclusive sobre evidências de validade de construto e confiabilidade teste e reteste. Os resultados do estudo atual vêm corroborar outros estudos sobre evidências de validade de construto, critério, confiabilidade e sensibilidade/especificidade da EBADEP-A, como demonstram os estudos de Baptista e colaboradores (2008, 2010), Dias (2008), Souza (2010), dentre outros.

É importante que diversas medidas do mesmo construto sejam disponibilizadas para clínicos e pesquisadores a fim de ampliar as possibilidades do uso desses instrumentos em diferentes contextos, de acordo com a característica de cada escala. Como afirmam Klein e colaboradores (2005), a seleção dos instrumentos com foco em suas especificidades é um dos exercícios importantes do avaliador, possibilitando maior acurácia em todo o processo da avaliação psicológica. Apesar dos resultados encontrados na presente pesquisa terem sido adequados, alguns limites deste estudo devem ser considerados. O principal deles se refere ao número amostral, cuja diferença provavelmente ocorreu em virtude de uma coincidência com o período pós-prova dos universitários. Como afirmaram Martins (2006) e Prieto e Muñiz (2000), os estudos de correlação possuem maior credibilidade quando também possuem grupos mais numerosos, além de outras tantas características metodológicas.

Outras limitações envolveram as variáveis referentes à caracterização da amostra, tanto em relação ao gênero, com maioria feminina, quanto a faixas etárias, das quais não foram obtidas diferenças em relação aos escores nos instrumentos e ao estado civil, no qual presença majoritária de participantes casados também pode ter sido um agravante

influenciador nas análises e merece maiores investigações em estudos semelhantes. Dessa forma, estudos futuros devem ser realizados com amostras mais numerosas, além de se poder também pensar em utilizar outros instrumentos de depressão para novas evidências de validade que não os já testados. Outra sugestão se refere a estudos de teste e reteste com pacientes comprovadamente diagnosticados com depressão, a fim de avaliar se o índice de correlação seria muito divergente do encontrado neste estudo.

Referências

- American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME] (1999). *Standards for educational and psychological testing*. Nova Iorque: American Educational Research Association.
- American Psychiatric Association [APA] (2002). *DSM-IV-TR: manual diagnóstico e estatístico de transtornos mentais*. Porto Alegre: Artmed.
- Andrade, T. M. R., & Argimon, I. I. L. (2006). Sintomas depressivos e o uso de substâncias psicoativas durante a vida em adolescentes. *Revista Brasileira de Terapias Cognitivas*, 2(1), 65-104.
- Andrews, G., Poulton, R., & Skoog, I. (2005). Lifetime risk of depression: restricted to a minority or waiting for most? *British Journal of Psychiatry*, 187, 495-496.
- Baptista, M. N. (2012). *Manual técnico da Escala Baptista de Depressão em Adultos (EBADEP-A)*. São Paulo: Vetor.
- Baptista, M. N., Carneiro, A. M., & Sisto, F. F. (2010). Estudo psicométrico de escalas de depressão (EDEP e BDI) e o Inventário de Percepção de Suporte Familiar – IPSF. *Psicologia em Pesquisa*, 4(1), 65-73.
- Baptista, M. N., Souza, M. S. & Alves, G. A. S. (2008). Evidências de Validade entre a Escala de Depressão (EDEP), o BDI e o Inventário de Percepção de Suporte Familiar (IPSF). *Psico-USF*, 13(2), 211-220.
- Baptista, M. N., Souza, M. S., Gomes, J. O., Alves, G. A. S., & Carneiro, A. M. (2012). Validade convergente e comparação de itens entre a EDEP e CES-D. *Psicologia: Teoria e Prática*, 14(1), 140-152.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1997). *Terapia cognitiva da depressão*. Porto Alegre: Artes Médicas.

- Bernstein, K. S. (2006). Clinical assessment and management of depression. *Medsurg Nursing*, 15(6), 333-342.
- Calil, H. M., & Pires, M. L. N. (1998). Aspectos gerais das escalas de avaliação de depressão. *Revista de Psiquiatria Clínica, Edição Especial*, 25(5), 240-244.
- Cunha, J. (2001). *Manual em português das escalas Beck*. São Paulo: Casa do Psicólogo.
- Dias, C. C. (2008). *Evidências de validade convergente e discriminante da Escala de Depressão (EDEP)* (Dissertação de Mestrado). Universidade São Francisco, Itatiba, São Paulo, Brasil.
- Dowd, T. (2004). Depression: theory, assessment, and new directions in practice. *International Journal of Clinical and Health Psychology*, 4(2), 413-423.
- DPhil, S. G., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal*, 178(8), 997-1003.
- Ferster, C. B., Culbertson, S., & Boren, C. P. (1977). Depressão clínica. Em C. B. Ferster, S. Culbertson & C. P. Boren (Orgs.), *Princípios do comportamento* (pp.699-725, M. I. R. Silva, M. A. C. Rodrigues & M. B. L. Pardo, trad.). São Paulo: Hucitec.
- Goreinstein, C., Pang, W. Y., Argimon, I. L., & Werlang, B. S. G. (2011). *Manual do Inventário de Depressão de Beck – BDI-II*. São Paulo: Casa do Psicólogo.
- Klein, D. N., Dougherty, L. R., & Olinio, T. M. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 412-432.
- Lutz, W., Stahl, S. M., Howard, K. I., Grissom, G. R., & Joske, R. (2002). Some relationships among assessments of depression. *Journal of Clinical Psychology*, 58(12), 1545-1553.
- Martínez, M. J. C., García, J. C., Martínez, M. R., Mendoza, C. L. M., & Ferrer, A. R. (2007). Revisión estructurada de las escalas de depresión en personas mayores. *International Journal of Clinical and Health Psychology*, 7(3), 823-846.
- Martins, G. A. (2006). Sobre confiabilidade e validade. *RBGN*, 8(20), 1-12.
- Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales. II: Scales for Internalizing Disorders. *Journal of American Academy of Child and Adolescent Psychiatry*, 41(6), 634-659.
- Nelson, J. C., Portera, L., & Leon, A. C. (2006). Assessment of outcome in depression. *Journal of Psychopharmacology*, 20(4), 47-53.
- Organização Mundial de Saúde. [OMS]. (1993). *Classificação de transtornos mentais e de comportamento da CID-10: descrições clínicas e diretrizes diagnósticas*. Porto Alegre: Artes Médicas.
- Patten, S. B., Bilsker, D., & Goldner, E. (2008). The evolving understanding of major depression epidemiology: implications for practice and policy. *The Canadian Journal of Psychiatry*, 53(10), 689-695.
- Prieto, G., & Muñiz, J. (2000). *Um modelo para avaliar la calidad de tests usados na Espanha*. Recuperado: 25 de Março de 2011. Disponível: <http://www.cop.es/index.php?page=evaluar-calidad>.
- Quintero, M. A., Garcia, C. C., Jiménez, V. L. G., & Ortiz, T. M. L. (2004). Caracterización de la depresión en jóvenes universitarios. *Universitas Psychologica Bogotá (Colombia)*, 3(1), 17-26.
- Ruscio, J., Brown, T. A., & Ruscio, A. M. (2009). A taxometric investigation of DSM-IV major depression in a large outpatient sample: interpretable structural results depend on the mode of assessment. *Assessment*, 16(2), 127-144.
- Smar, K. L. (2003). Measures of depression and depressive symptoms. *Arthritis & Rheumatism*, 49(5S), 134-146.
- Souza, M. S. (2010). *Evidências de validade e precisão para a Escala de Depressão de Baptista e Sisto (EDEP)* (Tese de Doutorado). Universidade São Francisco, Itatiba, São Paulo, Brasil.
- Spielberger, C. D., Ritterband, L. M., Reheiser, E. C., & Brunner, T. M. (2003). The nature and measurement of depression. *Revista Internacional de Psicología Clínica Y de la Salud*, 3(2), 209-234.
- Stewart, D. E. (2008). Battling depression. *Canadian Medical Association Journal*, 178(8), 1023-1024.
- Urbina, S. (2007). *Fundamentos da Testagem Psicológica*. Porto Alegre: Artmed.
- Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of Beck Depression Scale. *Assessment*, 15(2), 177-187.
- Vorcaro, C. M. R., Uchoa, E., & Lima-Costa, M. F. F. (2002). Prevalência e características associadas à depressão: revisão de estudos epidemiológicos com base populacional. *Jornal Brasileiro de Psiquiatria*, 51(3), 167-182.

Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Scale Inventory-second edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*(4), 545-551.

World Health Organization. [WHO]. (2010). *Depression*. Recuperado: 20 de janeiro de 2010. Disponível: www.who.int/topics/depression/en/.

Recebido em 21/08/2012
Reformulado em 25/09/2012
Aprovado em 08/10/2012

Nota:

Apoio financeiro. CNPq. CAPES.

Sobre os autores:

Makilim Nunes Baptista é psicólogo, doutor pelo Depto. de Psiquiatria e Psicologia Médica da Universidade Federal de São Paulo, docente do Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco/Itatiba e bolsista produtividade CNPq.

Hugo Ferrari Cardoso é psicólogo, mestre pelo programa de pós-graduação *Stricto Sensu* em Psicologia da Universidade São Francisco/Itatiba, doutorando do programa de pós-graduação em *Stricto Sensu* em Psicologia da Universidade São Francisco/Itatiba e docente da Universidade Estadual Paulista (UNESP-Bauru).

Juliana Oliveira Gomes é psicóloga, mestre pelo programa de pós-graduação *Stricto Sensu* em Psicologia da Universidade São Francisco/Itatiba, doutoranda do programa de pós-graduação em *Stricto Sensu* em Psicologia da Universidade São Francisco/Itatiba e bolsista CAPES.

Contato com os autores:

Programa de Pós Graduação *Stricto Sensu* em Psicologia.

Comissão de Pós Graduação. Universidade São Francisco.

Av. Alexandre Rodrigues Barbosa, 45 – Centro – CEP 13251-900 – Itatiba-SP, Brasil.

E-mail: makilim.baptista@saofrancisco.edu.br