

Instituto Superior de Psicologia Aplicada



A DEPRESSÃO À PROVA NO BDI-II: ESTUDO DAS CARACTERÍSTICAS
PSICOMÉTRICAS DO INSTRUMENTO NUMA AMOSTRA DE ESTUDANTES
UNIVERSITÁRIOS

Ana Rita Aires Nunes
10126

Tese submetida como requisito parcial para obtenção do grau de

Mestre em Psicologia

Especialidade em Psicologia Clínica

2009

Instituto Superior de Psicologia Aplicada

A DEPRESSÃO À PROVA NO BDI-II: ESTUDO DAS CARACTERÍSTICAS
PSICOMÉTRICAS DO INSTRUMENTO NUMA AMOSTRA DE ESTUDANTES
UNIVERSITÁRIOS

Ana Rita Aires Nunes

Dissertação orientada pela Professora Doutora Ana Cristina Quelhas

Tese submetida como requisito parcial para a obtenção do grau de

Mestre em Psicologia

Especialidade em Psicologia Clínica

2009

Dissertação de Mestrado realizada sob a orientação da Professora Doutora Ana Cristina Quelhas, apresentada no Instituto Superior de Psicologia Aplicada para a obtenção de grau de Mestre na especialidade em Psicologia Clínica conforme o despacho da DGES, nº 19673/2006 publicado em Diário da República 2ª série de 26 de Setembro, 2006.

A realização do presente trabalho não teria sido possível sem o apoio e dedicação de algumas pessoas, às quais desde já expresso o meu profundo agradecimento:

Aos indivíduos que participaram nesta investigação, aos professores que permitiram a recolha dos dados e ao ISPA, sem os quais não poderia ter realizado este trabalho.

Ao Dr. Csongor Juhos pela possibilidade de continuarmos a trabalhar lado a lado, pelo apoio, orientação e confiança que em mim depositou.

À Professora Doutora Ana Cristina Quelhas por me aceitar no seu seminário e pela orientação.

Ao Professor Doutor João Maroco pela enorme disponibilidade, pelas grandes lições de estatística e conselhos.

A toda a minha família, em particular aos meus pais e à minha irmã. Quase tudo se deve a eles!!! A vontade de trabalhar, o empenho e dedicação, a teimosia, a motivação...

Ao Zé por ter estado sempre do meu lado nestes anos tão difíceis, e principalmente nesta recta final, tão complicada e turbulenta! Pela partilha diária de bons e maus momentos, pelo apoio, carinho, compreensão e motivação de todos os dias...

A todos os meus amigos e colegas de faculdade, em particular à Pipas, à Piu, à Sarita e à Claudita pela partilha de experiências, apoio e motivação.

À Madalena... a musa da minha vida!

RESUMO

O presente estudo providencia dados psicométricos do Beck Depression Inventory – Second Edition (BDI-II), uma das medidas de auto-descrição mais usadas a nível mundial na avaliação da severidade da depressão, numa amostra de estudantes universitários. Para tal, participaram 670 alunos do ISPA, que preencheram não só o BDI-II, mas também o State-Trait Anxiety Inventory (STAI) e o Inventário de Avaliação Clínica da Depressão (IACLIDE). Neste estudo, o BDI-II revelou ter uma boa consistência interna. No entanto, em relação à sensibilidade dos itens, constatou-se que quase metade deles tem dificuldades na discriminação entre indivíduos com scores elevados e indivíduos com scores baixos. No que diz respeito à validade, todos os dados convergiram para a sua comprovação. O BDI-II mostrou-se altamente correlacionado com o IACLIDE, evidenciando uma boa validade concorrente e convergente, e moderadamente correlacionado com o STAI, revelando desta forma uma boa validade discriminante. Através da análise factorial confirmatória e exploratória, não foi encontrada uma estrutura factorial comparável à reportada para a população académica norte americana, tendo-se revelado que o modelo factorial que melhor se adequa a esta população é constituído por quatro dimensões. Duma maneira geral, conclui-se que o BDI-II apresenta fiabilidade e validade adequadas, pelo que pode ser usado com confiança para a avaliação da depressão.

Palavras-chave: BDI-II; estudo psicométrico; fiabilidade; sensibilidade; validade

ABSTRACT

This study provides psychometric data of Beck Depression Inventory - Second Edition (BDI-II), one of the self-report measures most used worldwide in assessing the severity of depression, in a sample of university students. To do so, 670 students of ISPA participated, which not only completed the BDI-II, but also the State-Trait Anxiety Inventory (STAI) as well as the Inventory of Clinical Depression Evaluation (IACLIDE). In this study, the BDI-II appeared to have a good internal consistency. However, in relation to the items sensitivity, it was found that almost half of them showed difficulties in discriminating between individuals with high and low scores. With regard to validity, all data converged to confirm it. The BDI-II was highly correlated with the IACLIDE, showing a good concurrent and convergent validity, and moderately correlated with the STAI, thus revealing a good discriminant validity. Through confirmatory and exploratory factor analysis, a factor structure comparable to the one reported to academic population in the US was not found, and it was revealed that the factor model that best fits this population consists of four dimensions. In general, one can conclude that the BDI-II has adequate reliability and validity; therefore, it can be used with confidence in the assessment of depression.

Key-words: BDI-II; psychometric study; reliability; sensitivity; validity

ÍNDICE

Introdução.....	1
Psicometria.....	3
Origem e História da Psicometria	3
Modelos da Psicometria	7
Natureza e Uso dos Testes Psicológicos	8
Definição de Teste Psicológico	9
Tipos de Testes Psicológicos.....	11
Usos e Finalidades dos Testes Psicológicos.....	12
Questões Éticas.....	14
Qualidades dos Testes Psicológicos.....	16
Padronização/Normatização	16
Validade.....	18
Validade de construto.	20
Validade de critério.....	22
Validade de conteúdo.	23
Fiabilidade	24
Análise dos Itens	26
Depressão.....	29
Teoria Cognitiva de Beck.....	32
Teoria do Desamparo Aprendido e Suas Reformulações	34
Teoria das Redes Associativas de Bower.....	37
Outras Teorias Cognitivas da Depressão	38
Ansiedade e Depressão – de Mãos Dadas ou de Costas Viradas?	39
Avaliação da Depressão e Sua Importância	41

Psicometria e Depressão.....	45
Instrumentos Existentes.....	46
Qualidades Psicométricas do BDI-II.....	50
Método.....	53
Amostra.....	53
Instrumentos.....	56
BDI-II.....	57
IACLIDE.....	62
STAI.....	66
Procedimento.....	70
Resultados.....	73
Fiabilidade.....	73
Sensibilidade.....	74
Forma da Distribuição dos Itens.....	75
Probabilidade das Opções de Resposta em Função do Score Total.....	76
Validade.....	77
Validade de Critério.....	78
Validade Concorrente.....	79
Validade de Construto.....	79
Validade Factorial.....	80
Validade Convergente.....	84
Validade Discriminante.....	86
Dados Sócio-Demográficos do BDI-II na População Académica do ISPA.....	88
Discussão.....	89
Referências Bibliográficas.....	101
Anexos.....	105
Anexo A: Metodologias Estatísticas que Permitem Avaliar a Validade de Construto....	106
Anexo B: Metodologias Estatísticas que Permitem Estimar a Fiabilidade.....	112
Anexo C: Metodologias Estatísticas que Permitem Analisar a Qualidade dos Itens.....	117
Anexo D: Estatística Descrita Relativa à Amostra.....	124
Anexo E: Instrumento de Recolha de Dados.....	130
Anexo F: Análise da Consistência Interna.....	141
Anexo G: Análise da Sensibilidade dos Itens – Forma da Distribuição dos Itens.....	145

Anexo H: Análise da Sensibilidade dos Itens – Probabilidade das Opções de Resposta em Função do Score Total	147
Anexo I: Análise da Validade Concorrente	155
Anexo J: Análise Factorial Confirmatória do Modelo de Beck para a População Académica do ISPA.....	157
Anexo K: Análise Factorial Exploratória	168
Anexo L: Análise Factorial Confirmatória	175
Anexo M: Análise Factorial Confirmatória com Itens com Pesos Factoriais nos Factores Iguais ou Superiores a 0,50.....	184
Anexo N: Correlações Inter-Itens dos Factores	193
Anexo O: Consistência Interna dos Factores	198
Anexo P: Análise das Validades Convergente e Discriminante	207
Anexo Q: Dados Sócio-Demográficos	210
Anexo R: Análise da Sensibilidade dos Itens do IACLIDE	218

LISTA DE TABELAS E FIGURAS

Lista de Tabelas

Tabela 1 – Scores totais médios dos três inventários em função dos cursos	55
Tabela 2 – Resumo dos dados referentes à consistência interna do STAI nos vários grupos .	69
Tabela 3 – Coeficientes de fiabilidade para cada um dos delineamentos	114

Lista de Figuras

Figura 1 – Esquema da teoria reformulada do desamparo aprendido	35
Figura 2 – Cadeia de acontecimentos na teoria da depressão de desespero	36
Figura 3 – Modelo factorial da população académica do Ispa	84

INTRODUÇÃO

A avaliação psicológica é parte integrante do exercício profissional do psicólogo nos mais diversos contextos (Gonçalves, Simões, Almeida & Machado, 2006), desde o clínico, ao de prevenção e de investigação. Ribeiro (1999, p.77) considera mesmo que “a avaliação psicológica constitui uma matriz base de identidade para os psicólogos”, e que foi através dela que a psicologia se constituiu como ciência. Contudo, a psicologia tem sido desde sempre criticada pela subjectividade das suas medidas, tendo sabido, todavia, superar esta crítica. De facto, como em psicologia se assume sempre o erro de medida sempre que se faz a avaliação, “O que está em jogo não será subjectividade vs objectividade, mas sim fidelidade e validade da medida vs falta dela” (Ribeiro, 2008, p.86). Assim, pode-se considerar que, duma maneira geral, a grande maioria das técnicas de avaliação psicológica podem ser usadas com confiança (quando usadas apropriadamente), pois têm validade e fiabilidade estabelecidas (Ribeiro, 2008).

Ainda assim, esta prática encontra-se um pouco limitada no nosso país, não só por não existirem muitos instrumentos de avaliação psicológica de origem portuguesa, mas também pela ausência de instrumentos estrangeiros devidamente adaptados à nossa população (Gonçalves et al., 2006). Esta situação leva a que muitas vezes, erroneamente, se utilizem instrumentos não adaptados e se façam especulações acerca dos seus resultados, sem se saber se os critérios utilizados para fazer essas especulações são adequados à população em estudo e, inclusivamente, sem se saber ao certo se o instrumento que se está a usar avalia

exactamente aquilo que se pretende avaliar (o que é avaliado, válida e eficazmente, num país por determinado instrumento, poderá não o ser num outro país).

Esta questão é de extrema importância em qualquer contexto, mas principalmente no clínico, no qual se trabalha com populações fragilizadas (por exemplo, pessoas deprimidas) e para as quais a avaliação psicológica serve como base não só para um diagnóstico, mas também como base de orientação para um tratamento a seguir ou como forma de avaliar os resultados duma intervenção já iniciada. Em contexto de investigação, no qual se pretende colectar dados sobre hipóteses formuladas, a sua importância não é menor, na medida em que fazer especulações a partir de dados não válidos, torna a investigação inválida e inútil. Fica desta forma demonstrada a importância de se fazer uma boa avaliação, utilizando bons instrumentos de avaliação, que meçam válida e eficazmente aquilo que se pretende medir.

Perante este problema e estando a depressão entre as perturbações psiquiátricas mais prevalentes e recorrentes (Gotlib & Hammen, 2009), considerada pela OMS como a doença mais dispendiosa a nível mundial (Murray & Lopez, 1996, in Gotlib & Hammen, 2009), que provoca sérios prejuízos em diversas áreas da vida dos indivíduos e que muito frequentemente não é tratada (Kessler & Wang, 2009), o objectivo do presente estudo é o de estudar as características psicométricas dum instrumento de avaliação da depressão e sua severidade, o Beck Depression Inventory – Second Edition (BDI-II; Beck, Steer & Brown, 1996), que tem sido um dos mais usados a nível mundial, tem demonstrado fortes propriedades psicométricas e tem sido adaptado para vários contextos e populações (Nezu, Nezu, Friedman & Lee, 2009).

Propomo-nos a fazê-lo não só porque é nossa convicção que é um instrumento válido e eficaz na avaliação da depressão e sua severidade, que detém boas qualidades psicométricas (fiabilidade e validade), mas também porque é um instrumento relativamente recente (comparado com os instrumentos validados e adaptados para a população portuguesa para avaliação da depressão), amplamente utilizado, e, como tal, para estarmos em sintonia com o que se passa lá fora temos de usar medidas universais.

É de referir que este estudo está inserido no âmbito da psicometria, pelo que vamo-nos debruçar não só no tema da depressão, que, no fundo, acaba por ser um pouco secundário, mas também, e principalmente, na psicometria em si e nos seus princípios metodológicos.

Psicometria

A psicometria tem sido, desde sempre, dominada e influenciada por investigadores de orientação estatística. Situação que não é de todo incompreensível visto a estatística ser fundamental e imprescindível para a psicometria, na medida em que nesta o fenómeno psicológico, seu objecto de estudo, é representado através do número. Apesar disto, parece erróneo conceber a psicometria como um ramo da estatística. Tal como Pasquali (2003, p.11-12) refere:

A Psicometria é uma área que pretende estudar fenómenos psicológicos. Consequentemente, seu objecto específico de estudo são os fenómenos psicológicos e não conceitos, no caso, o número. O número, nesta ciência, é apenas o modo de representar estes fenómenos psicológicos. Assim, a Psicometria deve ser concebida como um ramo da Psicologia e que se caracteriza por expressar (observar) o fenómeno psicológico através do número, em vez da pura descrição verbal. Nem por isso, ela deixa de ter como ponto central de sua existência o fenómeno psicológico.

Pode-se então considerar que o desenvolvimento e estabelecimento da avaliação psicológica (psicometria) como prática científica teve na sua base a evolução dos procedimentos estatísticos, que andou a par e passo com a evolução das técnicas de avaliação psicológica, permitindo o aperfeiçoamento destas mesmas técnicas (Ribeiro, 1999).

Origem e História da Psicometria

Segundo Anastasi e Urbina (2000) as raízes da testagem psicológica remontam à antiguidade; no entanto, os desenvolvimentos e acontecimentos mais marcantes que moldaram a testagem psicológica contemporânea datam do século XIX.

Nesta altura, a psicologia estava dividida em várias tendências epistemológicas, de entre as quais se destacam, por um lado, a psicologia alemã da introspecção, cujo foco de interesse era a experiência subjectiva, e, por outro, o empirismo inglês e norte-americano

focalizado no comportamento, assim como a escola (psicofísica) de Leipzig interessada nos processos mentais. Estas duas orientações caracterizavam-se pelo uso de procedimentos diferentes: Na psicologia introspectiva dominavam os procedimentos descritivos; para os empiristas, o uso de procedimentos quantitativos era fundamental. Portanto, não é de estranhar que as origens da psicometria se encontrem na abordagem empirista da psicologia.

Segundo Pasquali (2003), duas orientações, independentes e com diferentes preocupações, deram origem à chamada psicometria clássica. Na primeira orientação, havia uma preocupação prática de carácter psicopedagógico e clínico, que surgiu do interesse pelo tratamento mais humano dos doentes mentais, e que levou à criação de critérios uniformes para identificar e classificar estes indivíduos. A segunda orientação, na qual havia uma preocupação mais teórica e experimentalista, "...visava mais o desenvolvimento da própria teoria psicométrica e era, sobretudo, perseguida por psicólogos de orientação estatística" (Pasquali, 2003, p.15).

Ao contrário dos psicólogos da primeira orientação, os psicólogos experimentais estavam mais interessados na descoberta de uniformidades/generalizações do comportamento humano e não tanto na mensuração de diferenças individuais – estas eram vistas como desvios ou erros, que tornavam as generalizações aproximadas, em vez de exactas. Além disto, os psicólogos experimentais demonstravam a necessidade de se controlar as condições de testagem, dando início à padronização dos procedimentos de testagem, um dos importantes elementos da psicometria.

É possível enumerar os psicólogos que influenciaram a história da avaliação psicológica. A este respeito, Pasquali (2003) faz uma esquematização da evolução da psicometria, baseando-se nos psicólogos expoentes em cada época/década e nas "técnicas" mais marcantes para a avaliação psicológica, "dividindo-a" em sete fases, a saber: (1) A década de Galton: 1880; (2) A década de Cattell: 1890; (3) A década de Binet: 1900; (4) A era dos testes de inteligência: 1910-1930; (5) A década da análise factorial: 1930; (6) A era da sistematização: 1940-1980; e, finalmente, (7) A era da Psicometria moderna: A partir de 1980 até ao presente.

Francis Galton pode ser visto como o pioneiro na avaliação psicológica, com os seus trabalhos de avaliação do intelecto (aptidões humanas) através da discriminação sensorial.

“Galton também foi pioneiro no uso de métodos de escala de avaliação e questionário, e também no uso de técnicas de associação livre (...). Uma outra contribuição de Galton é encontrada no desenvolvimento de métodos estatísticos para a análise dos dados sobre diferenças individuais” (Anastasi & Urbina, 2000, p.43).

James McKeen Cattel também ocupa uma posição de enorme relevância. Influenciado por Galton, Cattel também acreditava que era possível avaliar as aptidões humanas através da discriminação sensorial e do tempo de reacção. Além disto, é a ele que se deve o termo “teste mental”.

Enquanto Cattel, dava preferência ao estudo das aptidões humanas através da discriminação sensorial e do tempo de reacção, pois segundo ele as funções simples do intelecto podiam ser medidas com maior precisão e exactidão do que as funções mais complexas, Binet, após vários anos de pesquisa e estudo sobre como medir a inteligência, concluiu que “a mensuração directa, ainda que imperfeita, das funções intelectuais complexas era a mais promissora” (Anastasi & Urbina, 2000, p.44-45). Foi neste contexto que foi criada a primeira escala de Binet-Simon.

Apesar de Pasquali (2003) considerar que a terceira fase de desenvolvimento da Psicometria se “deve” a Binet, o próprio autor refere que esta década deveria ser considerada a década de Spearman, na medida em que foi ele quem lançou os fundamentos da teoria da psicometria clássica através de inúmeras obras.

A era dos testes de inteligência foi influenciada por uma série de acontecimentos, nomeadamente o aparecimento do teste de inteligência de Binet-Simon e suas revisões (em particular a de Terman, que deu origem ao Stanford-Binet); o artigo de Spearman sobre o factor G; e, finalmente, o impacto que a Primeira Grande Guerra teve no desenvolvimento da testagem de grupo, devido à necessidade de selecção (rápida, eficiente e universal) de recrutas quanto ao nível intelectual.

Apesar do uso disseminado e indiscriminado dos testes de inteligência, depressa se percebeu que este tipo de instrumento tinha muitas limitações. Embora tenham sido criados para avaliar uma série de funções que determinariam o nível intelectual geral dos indivíduos, constatou-se que nem todas as funções importantes do intelecto estavam representadas, isto é,

apenas alguns dos aspectos da inteligência eram medidos. Constatou-se também que estes testes dependiam muito de factores culturais, o que contrapunha a ideia do factor geral universal de Spearman. Para além disso, também se constatou que existiam discrepâncias no desempenho dos indivíduos em diferentes partes dos testes¹.

Foi neste contexto que os trabalhos desenvolvidos por Kelley e Thurstone, conhecidos como análise factorial, impulsionaram a criação dos testes de aptidões múltiplas. Estes trabalhos demonstraram a existência de inúmeros factores (ou traços) da inteligência que seriam relativamente independentes e que alguns desses traços, como a compreensão verbal e o raciocínio numérico, estavam representados, de forma proporcionalmente diferente, nos tradicionais testes de inteligência (Anastasi & Urbina, 2000). O avanço na análise factorial proporcionou, então, o desenvolvimento das chamadas baterias de aptidões múltiplas, nas quais

Em lugar de um score total ou QI, é obtido um score separado para traços como compreensão verbal, aptidão numérica, visualização espacial, raciocínio aritmético e velocidade perceptual. Estas baterias proporcionam, portanto, um instrumento adequado para se fazer o tipo de análise intra-individual, ou diagnóstico diferencial, que os usuários de testes vinham tentando fazer há muitos anos com testes de inteligência, com resultados imperfeitos e muitas vezes erróneos (Anastasi & Urbina, 2000, p.47).

A era da sistematização foi basicamente dominada por: Trabalhos de síntese sobre os avanços da Psicometria, sobre a teoria clássica dos testes psicológicos, sobre a medida escalar, sobre os avanços na análise factorial, sobre a medida em personalidade e sobre uma teoria de inteligência; e por trabalhos de crítica sobre o uso das escalas de medida e sobre a própria teoria clássica dos testes (Pasquali, 2003).

A era da psicometria moderna só não é intitulada de era da teoria de resposta ao item (TRI), na medida em que, apesar de ser o modelo mais utilizado nos países mais avançados, ainda apresenta problemas por solucionar, não podendo portanto constituir-se como o modelo da psicometria moderna. Além disso, a TRI não apareceu para substituir a psicometria

¹ Por exemplo, nos testes de inteligência elaborados para a testagem em grupo, os itens estavam normalmente organizados em sub-testes com conteúdos homogéneos, o que permitiu verificar que os indivíduos poderiam ter um desempenho diferente, muito superior por exemplo, no sub-teste verbal em relação ao numérico – este facto levou a que erroneamente muitos utilizadores de testes fizessem extrapolações em relação a estas diferenças, prática que “não é recomendada, porque os testes de inteligência não foram planejados com o objectivo de se fazer uma análise diferencial de aptidões” (Anastasi & Urbina, 2000, p.47).

clássica na sua totalidade, mas unicamente algumas partes dela (Pasquali, 2003). Actualmente, a psicometria está a ser marcada por uma série de acontecimentos, dos quais se destaca: A sistematização da psicometria clássica, a pesquisa na TRI, a pesquisa em áreas paralelas da psicometria, os trabalhos da psicologia cognitiva na área das aptidões (através da análise dos componentes cognitivos), e a publicação de um número cada vez maior de trabalhos de psicometria (Pasquali, 2003).

Modelos da Psicometria

Actualmente, aceita-se que existem dois modelos em psicometria, a saber: A teoria clássica dos testes (TCT) e a já referida TRI. Fugindo um pouco ao âmbito do presente trabalho, que se prende, fundamentalmente, com as questões mais práticas da psicometria e não tanto com as questões teóricas de base, não nos vamos deter muito neste assunto, resumindo as principais características dos dois modelos com as palavras de Pasquali (2003, p.67) a este respeito:

A Psicometria procura explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas, tipicamente chamadas de itens. A TCT se preocupa em explicar o resultado final total, isto é, a soma das respostas dadas a uma série de itens, expressa no chamado score total (T). (...). A TRI, por outro lado, não está interessada no score total em um teste; ela se interessa especificamente por cada um dos (...) itens e quer saber qual é a probabilidade e quais são os factores que afectam esta probabilidade de cada item individualmente ser acertado ou errado (em testes de aptidão) ou de ser aceito ou rejeitado (em testes de preferência: Personalidade, interesses, atitudes). Desta forma, você vê que a TCT tem interesse em produzir testes de qualidade, enquanto a TRI se interessa por produzir tarefas (itens) de qualidade. No final, então, você tem ou testes válidos (TCT) ou itens válidos (TRI), itens com os quais você poderá construir tantos testes válidos quantos quiser ou o número de itens permitir. Você logo percebe, também, que a riqueza na avaliação psicológica, dentro do enfoque da TRI, consiste em se conseguir construir armazéns de itens válidos para avaliar os traços latentes, armazéns estes chamados de bancos de itens.²

² Para uma visão mais detalhada sobre os modelos da psicometria aconselhamos a leitura de Pasquali (2003, cap.4).

Natureza e Uso dos Testes Psicológicos

A forma como é feita a recolha de dados sobre os indivíduos, isto é, a forma como é efectuada a avaliação psicológica, pode variar, segundo Cronbach (1996), entre dois pólos: O psicométrico e o “impressionista”. A principal diferença entre estes dois pólos reside naquilo que são as máximas, por assim dizer, do método psicométrico, a saber: A padronização e a objectividade.

No entanto, é possível enumerar outras diferenças marcantes entre os dois métodos. No que diz respeito à interpretação dos dados, na testagem psicométrica o objectivo é medir os indivíduos – “A testagem psicométrica resume o desempenho a números. Seu ideal é expresso em dois famosos pronunciamentos antigos: Se uma coisa existe, ela existe em certa quantidade; se existe em certa quantidade, ela pode ser medida” (Cronbach, 1996, p.53); enquanto que na testagem impressionista o objectivo é caracterizar o indivíduo.

Na testagem psicométrica predominam as tarefas estruturadas [uma tarefa é estruturada “quando todas as pessoas a interpretam da mesma forma” (Cronbach, 1996, p.53)], os registos objectivos dos dados, o rigor na avaliação e na combinação de dados, e a ênfase na validação; enquanto que na testagem impressionista as tarefas poderão ser não estruturadas, e os dados, bem como a sua avaliação e validação, são mais subjectivos e passíveis de subjectividade.

Para além disso, na testagem psicométrica existe uma maior preocupação com o resultado (produto) do que propriamente com os meios (processos) através dos quais esse resultado é alcançado; enquanto que na testagem impressionista tudo conta, sendo que muitas vezes o processo que o indivíduo faz para chegar ao produto é mais importante do que o próprio produto. Para além do que “o examinador psicométrico prefere um método uniforme, impessoal, ao passo que o impressionista prefere flexibilidade” (Cronbach, 1996, p.54).

Finalmente, na testagem psicométrica as interpretações são baseadas em grupos de referência estatisticamente definidos, e não em interpretações subjectivas e individualizadas, como acontece na testagem impressionista.

Apesar desta diferenciação entre os métodos de avaliação psicológica, na grande maioria das vezes, a aplicação de testes psicológicos alterna entre os dois pólos, isto é, pode basear-se tanto no pólo psicométrico para alguns aspectos da avaliação, como no impressionista para outros.

Na mesma linha de pensamento, Ribeiro (2008) refere que em psicologia (e na saúde) os investigadores podem diferenciar-se pelas técnicas de investigação que usam: Qualitativas vs. quantitativas. As primeiras permitem um estudo profundo e detalhado, na medida em que não existem categorias pré-determinadas que constriam a investigação; enquanto que as segundas “requerem o uso de medidas e métodos padronizados convertíveis em números, de tal modo que não permitem a expressão da variedade de perspectivas e experiências das pessoas dado que as opções de resposta são limitadas à partida” (Ribeiro, 2008, p.28).

Qualquer um dos métodos apresenta potencialidades e limitações, e, apesar de serem habitualmente vistos como opostos e mutuamente exclusivos, eles podem ser mais profícuos se usados complementarmente, na medida em que, como refere Ribeiro (2008, p.75), “A investigação qualitativa ajuda a explicar, a compreender o quê e o porquê, que a investigação quantitativa ajuda, com a força dos números, a identificar. Esta força dos números (...) não ajuda no entanto a compreensão dos pequenos detalhes que podem explicar os resultados”. Desta forma, existindo complementaridade enriquece-se a investigação.

Definição de Teste Psicológico

São várias as técnicas que se podem utilizar na avaliação psicológica, mas, numa maneira geral, pode-se dizer que todos os instrumentos de avaliação psicológica são testes, apesar de muitas vezes não utilizarem esse nome para se definirem (Ribeiro, 1999). Assim, segundo Ribeiro (1999; 2008), os testes podem ser designados de diferentes modos, que têm um significado técnico subjacente, nomeadamente: Inventários, questionários, testes, escalas, entrevistas, observação naturalista, listas de verificação e listas de classificação.

Ribeiro (2008, p.80) caracteriza “os testes como um tipo de medidas não intrusivas e que exigem ao avaliado um papel activo para fornecer uma resposta às questões que são colocadas, ou directamente ou como um problema”. Para Murphy e Davishofer (1998, in Ribeiro, 1999), um teste psicológico é uma amostra de comportamento, recolhida em

condições padronizadas, no qual existem regras para a avaliação das respostas e para a produção de informações quantitativas acerca dessa amostra de comportamento. Para Anastasi e Urbina (2000, p.18-19):

Um teste psicológico é essencialmente uma medida objectiva e padronizada de uma amostra de comportamento. Os testes psicológicos são como os testes de qualquer outra ciência, na medida em que são feitas observações sobre uma amostra pequena, mas cuidadosamente escolhida do comportamento de um indivíduo. (...). O facto de o teste cobrir ou não de forma adequada o comportamento que está sendo considerado depende, obviamente, do número e natureza dos itens da amostra.

Para que um teste psicológico se considere padronizado é necessário que se controlem as condições da testagem, isto é, que exista uniformidade dos procedimentos de testagem, desde a aplicação do teste (tendo em conta os materiais usados, as instruções dadas, os limites de tempo, os exemplos cedidos, a forma como se responde às questões colocadas pelos testandos, enfim, tudo aquilo que possa influenciar a situação de testagem), à sua cotação e análise, para que os scores obtidos por diferentes indivíduos, em diferentes momentos e lugares, sejam comparáveis.

Outra questão importante que se refere à padronização é a do estabelecimento de normas. Para tal, é necessário que o teste tenha sido aplicado a uma amostra grande e representativa da população para a qual foi criado – amostra de padronização ou amostra padrão, que serve para estabelecer as normas de comparação, que representam o desempenho normal ou médio. Segundo Anastasi e Urbina (2000, p.20) “Essas normas não só indicam o desempenho médio como também a frequência relativa de graus variados de desvios acima e abaixo da média. Assim, podemos avaliar diferentes graus de superioridade e inferioridade”.

Em relação à objectividade é dispensável uma grande explicação, na medida em que esta está implícita quando se considera a existência de uniformidade dos procedimentos de testagem, ou seja, o teste é objectivo na medida em que não depende de julgamentos subjectivos do examinador, está livre da sua influência – é isto que idealmente se deseja, é este o objectivo quando se constrói um teste.

Quando na literatura é mencionado o uso de testes psicológicos, é muito frequente aparecerem dois termos, que são de extrema importância para a psicometria, e são eles: A

fiabilidade e a validade. A fiabilidade (fidedignidade ou fidelidade) diz respeito à capacidade do teste medir de forma repetida e consistente (Hill & Hill, 2008), ou seja, indica se os scores obtidos pelos mesmos indivíduos, quando retestados (testados noutras ocasiões) com o mesmo teste ou com conjuntos equivalentes de itens, são idênticos (Ribeiro, 2008). Já a validade diz basicamente se o teste mede aquilo que pretende medir ou, segundo Anastasi e Urbina (2000, p.23), “seria mais preciso definir a validade como o quanto nós sabemos o que o teste mede”. Apesar da sua importância, não nos vamos alongar em relação a estes conceitos, pois eles serão detalhadamente abordados mais à frente, no que diz respeito à qualidade dos testes psicológicos.

Tipos de Testes Psicológicos

Os testes psicológicos podem ser de vários tipos. Quanto aos aspectos comportamentais que abrangem podem ser: Testes de realização, observações comportamentais e medidas de auto-registo para Murphy e Davidshofer (1998, in Ribeiro, 2008); testes de realização máxima (testes de capacidade) e testes de realização típica ou habitual (testes de personalidade, interesses, hábitos, etc.), que se subdividem em observação do comportamento e técnicas de auto-preenchimento, para Cronbach (1996); testes de habilidades e testes de personalidade para Anastasi e Urbina (2000).

Quanto à forma de aplicação, os testes podem ser classificados como técnicas de avaliação individual ou colectiva (Anastasi & Urbina, 2000; Ribeiro, 1999); e podem ser de papel e lápis (Ribeiro, 1999) ou de aplicação por computador (Anastasi & Urbina, 2000). Quanto à objectividade que lhes é inerente, segundo Sundberg (1977, in Ribeiro, 1999) os testes podem ser classificados como objectivos (técnicas psicométricas) ou não objectivos (técnicas comportamentais e projectivas).

Os testes psicológicos podem também diferenciar-se quanto ao modo como se interpretam os dados: Em relação com normas, com a teoria ou com critérios (Ribeiro, 2008). No primeiro caso, os resultados são numéricos e a nota obtida é comparada com a nota obtida por um grupo de referência (normativo) – típico das técnicas psicométricas. No segundo caso, as respostas são qualitativas e analisadas com base na teoria subjacente ao teste (que permite a categorização do indivíduo) – típico dos testes projectivos e da avaliação piagetiana. No último caso, as respostas são comparadas com um critério definido à priori (o que se

espera do indivíduo) – típico da avaliação feita na intervenção por objectivos (que se faz antes e depois do indivíduo fazer determinada aprendizagem).

Diferenciam-se ainda quanto ao tipo de preenchimento que requerem, podendo ser preenchidos pelo próprio indivíduo (auto-avaliação), perante o testador ou na sua ausência, ou preenchidos pelo testador (hetero-avaliação) – que pode acontecer em várias situações, nomeadamente no registo directo das respostas do indivíduo ou no registo indirecto das respostas ou sinais do indivíduo, na observação estruturada ou não estruturada (Ribeiro, 2008).

Segundo Ribeiro (1999), existem ainda testes de cultura equilibrada (técnicas de avaliação psicológica nas quais são minimizados os efeitos culturais) e testes livres de cultura (técnicas de avaliação psicológica nas quais são anulados os efeitos culturais).

Usos e Finalidades dos Testes Psicológicos

É importante ter em mente que nenhum teste fornece dados por si próprio e será inútil quando mal aplicado e/ou mal interpretado (Ribeiro, 2008). Ou seja, para serem efectivos os testes psicológicos deverão ser usados apropriadamente. Se usados por indivíduos sem escrúpulos ou por indivíduos com boas intenções, mas pouco informados (com falta de formação ou com formação inadequada), tornam-se inúteis, podendo provocar desconforto e uma resposta negativa no indivíduo testado – a este respeito Ribeiro (2008, p.80) refere que “utilizar este material sem treino é tão disparatado como utilizar análises sanguíneas, ou radiografias sem treino”. No entanto, é de referir que a formação em psicologia não qualifica automaticamente o profissional a utilizar todos e quaisquer tipos de testes – existem testes para os quais não é necessária nenhuma formação específica e que podem ser administrados e interpretados por pessoas responsáveis, e existem testes para os quais, para além da formação em psicologia, é “obrigatória” uma formação específica para o seu uso (Cronbach, 1996).

Cronbach (1996) distingue quatro usos específicos dos testes, dependendo dos seus objectivos/funções: Classificação; promoção do auto-entendimento; avaliação e modificação de tratamentos; e, finalmente, investigação científica. Esta diversidade de objectivos/funções, patente na passagem que se segue, alerta-nos para o facto de que não existe um único teste que se adequa igualmente bem para todos os fins (Cronbach, 1996).

Tradicionalmente, a função dos testes psicológicos era a de medir diferenças entre os indivíduos ou entre as reacções do mesmo indivíduo em diferentes circunstâncias. Um dos primeiros problemas que estimulou o desenvolvimento dos testes psicológicos foi a identificação das pessoas mentalmente retardadas. Até hoje, a detecção das deficiências intelectuais permanece uma aplicação importante de certos tipos de testes psicológicos. Usos clínicos relacionados de testes incluem o exame de pessoas com transtornos emocionais graves e outros tipos de problemas comportamentais. Um forte ímpeto para o desenvolvimento inicial dos testes também foi proporcionado pelas necessidades de avaliação que surgiram na educação. Este foi o caso dos famosos testes Binet, que introduziram a testagem da inteligência. No momento, as escolas estão entre os principais usuários de testes. A classificação das crianças com referência à sua habilidade de beneficiar de diferentes tipos de instrução escolar, a identificação de aprendizes notavelmente lentos ou rápidos, o aconselhamento educacional e ocupacional no ensino médio e na universidade, e a selecção de candidatos para as escolas profissionais estão entre os muitos usos educacionais dos testes.

A selecção e a classificação de pessoal nas indústrias representa uma outra aplicação importante da testagem psicológica. (...) Uma aplicação bastante relacionada da testagem psicológica é encontrada na selecção e classificação de pessoal militar. (...)

O uso de testes no aconselhamento individual foi-se ampliando gradualmente, de uma orientação estreitamente definida referente a planos educacionais e vocacionais para um envolvimento com todos os aspectos da vida da pessoa. O bem-estar emocional e as relações interpessoais efectivas tornaram-se objectivos cada vez mais proeminentes do aconselhamento. Também existe uma ênfase crescente no uso de testes para aumentar o auto-entendimento e o desenvolvimento pessoal. Nesta estrutura, os scores de teste são parte das informações dadas ao indivíduo que o ajudam em seus processos de tomada de decisão.

Está claramente evidente que os testes psicológicos são empregados actualmente na solução de uma variedade ampla de problemas práticos. Mas nós não devemos perder de vista o facto de que esses testes também estão atendendo a funções importantes na pesquisa básica. (...) os testes psicológicos oferecem instrumentos padronizados para a investigação de problemas variados, como as mudanças desenvolvimentais no indivíduo durante o ciclo vital, a eficácia relativa de diferentes procedimentos educacionais, os resultados da psicoterapia, o impacto de programas comunitários e a influência das variáveis ambientais sobre o desempenho humano (Anastasi & Urbina, 2000, p.17-18).

A este respeito, Ribeiro (1999) diz-nos que qualquer tipo de avaliação tem como objectivo produzir informação relevante e adequada para a tomada de decisões, pois “nunca se avalia só para avaliar, principalmente em psicologia: A seguir à avaliação algo deve ocorrer” (p.78), pelo que na avaliação psicológica as decisões resultantes podem ser de quatro tipos, à semelhança do referido por Cronbach (1996) quanto aos objectivos/funções dos testes: De selecção (quando se pretende seleccionar um indivíduo para alguma função); de monitorização da intervenção (quando se pretende avaliar o impacto duma intervenção, em

termos dos resultados esperados); de investigação (quando se pretende colectar dados sobre hipóteses formuladas); e de diagnóstico (quando se pretende determinar uma acção a seguir).

Questões Éticas

Desde há algum tempo, mais propriamente desde a década de 50, quando foi criado o código formal de ética para os psicólogos (constituído por uma série de padrões éticos) pela APA, tem havido uma maior preocupação com as questões éticas na prática da psicologia em geral e no uso dos testes psicológicos em particular (Anastasi & Urbina, 2000). Assim, como refere Ribeiro (2008), o código de ética obriga todos os psicólogos a estarem sujeitos a certas exigências acerca da selecção, administração, classificação, interpretação e do uso das técnicas de avaliação psicológica. Para além destas exigências, o código de ética obriga a não utilização de técnicas de avaliação desadequadas, por serem obsoletas ou por não existirem dados relativos às propriedades psico e clinimétricas (ver adiante).

Para além do já referido e entre outros aspectos, focar-se-ão aqueles que parecem ser mais pertinentes para esta preocupação ética, a saber: A qualificação e competência do profissional que usa testes; a protecção da privacidade/os direitos do indivíduo testado; a confidencialidade dos dados; e a comunicação dos resultados dos testes.

Em relação à qualificação e competência do profissional que usa os testes, o código de ética diz que os psicólogos deverão “oferecer apenas os serviços e usar apenas as técnicas para os quais estão qualificados por formação, treinamento e experiência” (APA, 1992, p.1599, citado por Anastasi & Urbina, 2000, p.434). Se assim for, não se faz mais do que proteger o indivíduo testado do uso inadequado dos testes, na medida em que um profissional competente selecciona os testes mais apropriados para o seu objectivo, assim como para o indivíduo testado, tendo em conta as qualidades formais dos testes e as condições específicas que afectam o desempenho dos indivíduos testados; e finalmente, um profissional competente tira conclusões e faz recomendações baseadas não só no score do teste, mas também baseadas na história de vida dos indivíduos.

No que diz respeito à protecção da privacidade, esta liga-se essencialmente ao uso de testes de personalidade (nos quais os indivíduos têm que partilhar os seus sentimentos, pensamentos e pormenores da sua vida), apesar de se aplicar a qualquer outro tipo de teste.

Independentemente do tipo de teste, existem dois princípios no que toca à protecção da privacidade (Anastasi & Urbina, 2000): A relevância da testagem, isto é, todas as informações que são colectadas pelo teste deverão ser relevantes para o objectivo da testagem; e o consentimento informado, princípio que obriga a que o profissional informe os indivíduos testados sobre o objectivo da testagem, o tipo de informações que são colectadas e sobre como vão ser usados os resultados dos testes.

Para além disto, num processo de investigação, deverá ser obtido o consentimento informado dos indivíduos testados (exceptuando quando a investigação é anónima, quando se faz observação naturalista ou investigação de arquivos), de maneira a que eles percebam não só a natureza da investigação, mas também que são livres de participar ou não, e de abandonar a investigação, e quais as consequências da sua decisão (Ribeiro, 2008).

Actualmente é cada vez mais usual os indivíduos testados terem acesso aos resultados dos seus testes, e até mesmo terem liberdade de comentar e esclarecer algumas informações que derivam da sua análise. Independentemente deste avanço, ninguém poderá ter acesso a estes dados sem o conhecimento e consentimento do indivíduo testado, exceptuando os casos em que a própria lei permita ou ordene.

No seguimento desta questão, coloca-se uma outra de grande importância que se refere à forma como são comunicados os resultados dos testes. É necessário ter presente que a informação a ser dada deve ser significativa e útil para quem a recebe, conter interpretações apropriadas e levar em conta as características dos receptores – é muito diferente falar-se para um profissional competente no uso de testes, para o próprio indivíduo testado, para os pais de menores, para os professores, para os psiquiatras, para os tribunais, entre outros. Finalmente, deve-se ter especial atenção ao impacto emocional e social que os resultados dos testes poderão ter, seja no indivíduo testado, seja nos seus pais, nos seus professores ou outros.

Findo este ponto, inicia-se de seguida aquele que é fulcral neste trabalho e que diz respeito aos princípios técnicos e metodológicos da psicometria (Anastasi & Urbina, 2000).

Qualidades dos Testes Psicológicos

Tal como qualquer medida, os testes psicológicos devem ser dotados de alguns critérios, ou qualidades métricas, assim como os seus itens constituintes (Ribeiro, 2008). Relativamente às propriedades psicométricas dos testes, estas são de dois tipos: A fiabilidade e a validade, que devem ser adequadas para que a medida obtida pelos testes possa ser usada com confiança (Hill & Hill, 2008). Em relação aos itens, estes podem ser vistos como testes e, como tal, os procedimentos utilizados para validar os testes são também adequados para sua validação (Murphy & Davidshofer, 1998, in Ribeiro, 2008).

Descrever-se-ão de seguida os critérios ou qualidades métricas que os testes deverão ter para serem considerados uma boa medida, referenciando não só a questão da fiabilidade, validade e análise dos itens, bem como a questão da padronização/normatização.

Padronização/Normatização

A padronização ou normatização diz respeito à necessidade de se controlar/uniformizar as condições da testagem desde a aplicação dos testes à sua cotação e análise. Para Cronbach e outros autores (Cronbach, 1996; Cronbach, 1996, in Pasquali, 2003) existe uma clara distinção entre estes dois termos, sendo que consideram que a padronização se refere à uniformidade de aplicação dos testes e que a normatização se prende com a uniformidade de interpretação dos resultados dos testes. Apesar dos termos serem usados indistintamente, considera-se, tal como Cronbach, que ambos se referem a questões diferentes e, portanto, vão ser tratados separadamente.

Sendo que a padronização se refere à uniformidade das condições da aplicação dos testes psicológicos, com a qual se pretende garantir uma boa qualidade dos dados, e, conseqüentemente, um uso adequado e legítimo dos mesmos, tem que se ter em conta tanto o material de testagem como o próprio ambiente em que este é usado. No entanto, é de referir que isto só é válido para testes de boa qualidade – se uma má aplicação de um teste válido torna os seus dados inválidos, uma boa aplicação de um teste não válido, também não serve de nada, pois os dados resultantes serão também eles inválidos.

Em relação ao material da testagem, existem duas condições a ter em conta (Pasquali, 2003): A qualidade do teste (o teste tem que ser válido e preciso); e a pertinência do mesmo (o teste tem que ser relevante para o objectivo da testagem e tem que se adequar ao indivíduo testado). No que diz respeito ao ambiente da testagem, Pasquali (2003) refere que deverão ser tidas em consideração as seguintes questões:

- *Os procedimentos de aplicação* – a aplicação do teste deve seguir em pormenor as instruções e recomendações do autor, e a qualidade do ambiente físico e psicológico deve ser garantida para que o indivíduo esteja em óptimas condições de desempenho;
- *O controlo dos vieses do examinador* – deve-se tentar minimizar a influência do modo de ser e de estar do examinador, pois pode afectar o desempenho dos indivíduos e, consequentemente, os resultados dos testes;
- *O direito dos indivíduos testados* – deve ser obtido o consentimento informado dos indivíduos testados, informá-los sobre o tipo de dados que são recolhidos e como podem ser usados, e, caso os scores sejam usados para tomar decisões, deve-se dar a conhecer aos indivíduos os seus scores e sua interpretação;
- *As normas de divulgação dos resultados (sigilo profissional)* – só tem direito aos resultados dos testes o próprio indivíduo testado ou o solicitante da testagem, e os resultados devem estar em condições de segurança (inacessíveis e sem identificação dos indivíduos).

A normatização dos testes psicológicos diz respeito aos seus padrões de interpretação dos resultados, através dos quais um resultado obtém sentido. Segundo Pasquali (2003, p.239), “Uma tal norma permite situar o score de um sujeito, permitindo: (1) determinar a posição que o sujeito ocupa no traço medido pelo teste que produziu o tal score e, (2) comparar o score deste sujeito com o score de qualquer outro sujeito”.

É de referir que as normas representam o desempenho da amostra de padronização no teste, tal como referido anteriormente, e podem basear-se em vários critérios de referência, a saber: No nível de desenvolvimento humano (normas de desenvolvimento), num grupo padrão constituído pela amostra representativa da população para a qual o teste foi construído (normas intra-grupo) e num critério externo (normas referentes a critério).

Validade

A validade é, indubitavelmente, um parâmetro fundamental e indispensável na psicometria e na avaliação da qualidade dos testes psicológicos, na medida em que diz se o teste mede aquilo que pretende medir, e não qualquer outra coisa parecida ou diferente, e quão bem ele o faz (Ribeiro, 2008), isto é, diz se há ou não congruência entre o traço medido e a sua representação física/comportamental (itens). Ou seja,

...ao se medirem os comportamentos (itens), que são a representação do traço latente, está-se medindo o próprio traço latente. Tal suposição é justificada se a representação comportamental for legítima. Esta legitimação somente é possível se existe uma teoria prévia do traço que fundamente que tal representação comportamental constitui uma hipótese dedutível desta teoria. A validade do teste (este constituindo a hipótese), então, será estabelecida pela testagem empírica da verificação da hipótese (Pasquali, 2003, p.162).

Portanto, diz-se que um teste tem validade (não se pode dizer que é válido, pois o que existe são graus de validade) quando é uma medida da variável latente que se pretende medir, sendo esta "...uma variável que não pode ser observada nem medida directamente mas que pode ser definida a partir de um conjunto de outras variáveis (possíveis de serem observadas ou medidas) que medem qualquer coisa em comum (nomeadamente, a variável latente)" (Hill & Hill, 2008, p.135). Às variáveis passíveis de serem observadas ou medidas, Hill e Hill (2008) chamam variáveis componentes (pois compõem a variável latente), variáveis que podem ser medidas através de itens, pelo que deverá haver pelo menos um item para cada uma delas.

Pode-se dizer que este parâmetro tem a sua "história" dividida em três fases, nas quais predomina cada um dos tipos de validade conhecidos actualmente: Validade de conteúdo, de critério e de construto – a conhecida validade dos três C's. Alguns autores, como Hill e Hill (2008), referem-se a estes tipos de validade como validade de conteúdo, validade prática e validade teórica, respectivamente. Murphy e Davidshofer (1998, in Ribeiro, 2008), por exemplo, referem que a validade tem quatro faces: Validade de conteúdo, de construto, preditiva e concorrente, tendendo as duas últimas a ser agrupadas na validade de critério.

Voltando à "história" do parâmetro validade, o predomínio da validade de conteúdo fez-se sentir entre 1900 e 1950 (Pasquali, 2003), período durante o qual se considerava um teste válido quando o seu conteúdo correspondia ao domínio de conteúdo que o teste pretendia avaliar.

Entre 1950 e 1970 houve o predomínio da validade de critério (Pasquali, 2003). Nesta fase, um teste era considerado válido quando conseguia prever outros comportamentos, que não os amostrados, ou os mesmos comportamentos numa situação futura ou noutra condição. “Não interessava saber por que o teste predizia, bastava mostrar que de facto ele o fazia e isto era o critério de sua validade” (Pasquali, 2003, p.160).

Desde 1970 até ao presente tem-se vindo a sentir a predominância da validade de construto (ou dos traços latentes) na validação dos testes psicológicos. Segundo Anastasi e Urbina (2000, p.108), “Construtos são categorias amplas, derivadas das características comuns compartilhadas por variáveis comportamentais directamente observáveis. Porém, elas são entidades teóricas, não observáveis directamente”. Sendo que é através da validade de construto que se especifica o que o teste mede, isto é, que se verifica a legitimidade da representação comportamental dos traços latentes, esta passou a ser considerada por muitos a validade fundamental, que, de certa forma, contém as outras duas validades, ou seja, “os procedimentos de validação preditiva e de conteúdo estão entre as muitas fontes de informação que contribuem para a definição e o entendimento dos construtos avaliados por um teste” (Anastasi & Urbina, 2000, p.108).

Segundo Anastasi e Urbina (2000), a selecção dos procedimentos de validação depende, logicamente, do uso que vai ser dado ao teste, sendo que o mesmo teste, quando usado para objectivos diferenciados, pode e deve ser validado de forma diferente. A este respeito Ribeiro (2008, p.94) refere que

A validade deve ser determinada com referência ao uso particular para que o teste foi desenvolvido. Os processos de determinação da validade abrangem a determinação da relação entre a realização no teste e outros factos observáveis ligados às características do comportamento em consideração. Estabelecer a validade de um teste é sempre um procedimento que depende do julgamento do investigador perante o valor da evidência que conseguiu produzir. Os juízos para estabelecer a validade de uma medida devem ser baseados em evidência empírica sólida. Não é o teste, propriamente, que é validado, mas sim as inferências e as conclusões a que se chegam com base na nota do teste. Ou seja, a validade não é uma propriedade do teste mas sim uma função do que a nota do teste mede.

É ainda de referir que a validade de qualquer teste que meça sentimentos, satisfação, atitudes, crenças, opiniões, etc., pode ser fortemente influenciada por uma tendência que os

indivíduos podem ter para dar respostas socialmente desejáveis, que “é especialmente forte quando os itens se referem a assuntos íntimos ou embaraçosos – ou quando a resposta verdadeira ameaça a auto-confiança do respondente” (Hill & Hill, 2008, p.153).

Validade de construto.

A validade de construto é considerada a validade nobre de qualquer teste ou medida (Ribeiro, 2008); é aquilo que permite dizer que o teste mede um determinado construto teórico ou traço latente; é “a forma mais fundamental de validade dos instrumentos psicológicos e com toda a razão, dado que ela constitui a maneira directa de verificar a hipótese da representação comportamental dos traços latentes...” (Pasquali, 2003, p.164). O que está aqui em jogo não é “...descobrir o construto a partir de uma representação existente (teste), mas sim descobrir se a representação (teste) constitui uma representação legítima, adequada do construto” (Pasquali, 2003, p.165), que se quer medir.

Pelo facto de os construtos serem dimensões não observáveis, ou seja, serem sumários abstractos de fenómenos que ocorrem com regularidade e de estarem relacionados com acontecimentos observáveis concretos (Murphy e Davidshofer, 1998, in Ribeiro, 2008), a determinação da validade de construto é muito “complexa e decisiva para se considerar uma medida como séria” (Ribeiro, 2008, p.95).

Em psicologia a definição de um construto faz-se da seguinte forma: Identificam-se os comportamentos que se relacionam com o construto a medir; de seguida, identificam-se outros construtos e define-se a relação que estabelecem com o construto a medir; finalmente, identificam-se os comportamentos que se relacionam com os outros construtos e, tendo em conta a relação que se estabelece entre os construtos, determina-se em que condições os comportamentos se relacionam com o construto que se pretende medir (Ribeiro, 2008).

Cronbach (1996) refere que na busca pela validade de construto procura-se basicamente confirmar a convergência de indicadores que se correlacionam, bem como a divergência de indicadores que não se correlacionam com o construto em causa – é a chamada validade convergente e discriminante de que tantos autores falam.

Segundo Hill e Hill (2008) a validade de construto (ou teórica como lhe chamam) é de três tipos: Convergente, divergente (ou discriminante) e factorial. Para que uma medida da variável latente tenha validade convergente deverá correlacionar-se com outras medidas da mesma variável latente, com as quais está teoricamente relacionada. Para que haja validade discriminante, uma medida da variável latente não deverá estar correlacionada com medidas de outras variáveis latentes, com as quais não se relaciona teoricamente. Finalmente, na validade factorial estabelece-se o coeficiente de correlação entre o(s) factor(es) e as medidas da variável latente, que indica a validade factorial de cada medida da variável que se obtém com o instrumento.

É de referir que segundo Herdman, Fox-Rushby e Badia (1998, in Ribeiro, 2008, p.96) “A validade convergente refere-se à extensão em que a correlação do instrumento com instrumentos que medem o mesmo construto é maior do que a correlação com os que medem construtos diferentes”. Relativamente à validade divergente, os mesmos autores afirmam que esta se refere “...à extensão em que a correlação do instrumento com instrumentos que medem diferentes construtos é menor do que a correlação com os que medem o mesmo construto” (Herdman, Fox-Rushby & Badia, 1998, in Ribeiro, 2008, p.96).

Existem inúmeros métodos e/ou técnicas, que permitem avaliar a validade de construto (ver Anexo A). Apesar de existir alguma divergência ao nível da nomenclatura usada pelos diferentes autores, existe uma grande equivalência entre as várias técnicas que referem. Como é óbvio, todas as técnicas apresentam limitações, no entanto,

Nem por isso se justifica o simples abandono das mesmas. Primeiramente, porque em ciência empírica nada existe de perfeito e isento de erro e, em segundo lugar, a consciência destas dificuldades deve servir para melhorar e não abandonar as técnicas. Aliás, é recomendável o uso de mais de uma das técnicas (...) para demonstrar a validade de construto do teste, dado que a convergência de resultados das várias técnicas constitui garantia para a validade do instrumento (Pasquali, 2003, p.184-185).

Resumindo, como Anastasi e Urbina (2000, p.117) referem "Qualquer dado que lance luz sobre a natureza do traço sob consideração e as condições que afectam seu desenvolvimento e suas manifestações representa uma evidência apropriada para a sua validação”.

Validade de critério.

Sendo a validade de critério do teste vista como o seu grau de eficácia para predizer um comportamento particular de um indivíduo, este comportamento é o critério através do qual a medida obtida pelo teste é avaliada (Anastasi & Urbina, 2000; Pasquali, 2003; Ribeiro, 2008). É de referir que, como é óbvio, o comportamento do indivíduo deverá ser avaliado com técnicas independentes ao próprio teste que se está a validar.

Considera-se que a validade de critério se pode diferenciar em validade preditiva e validade concorrente (ou simultânea), tendo como base as relações temporais que se estabelecem entre a colecta dos dados do teste e os do critério (Anastasi & Urbina, 2000; Cronbach, 1996; Hill & Hill, 2008; Pasquali, 2003; Ribeiro, 2008). Assim, quando a recolha dos dados do teste e do critério ocorre simultaneamente, fala-se de validade concorrente; quando a recolha dos dados de critério é efectuada algum tempo depois da recolha dos dados do teste, trata-se de validade preditiva. O que é relevante não é propriamente a questão temporal, mas sim o uso de critérios válidos e/ou adequados, que sejam medidos de forma válida e independente do teste que está a ser validado (Pasquali, 2003).

Além disto, a distinção entre validade preditiva e validade concorrente está intimamente ligada aos próprios objectivos de testagem, pelo que se considera que “A validação concorrente é relevante para os testes empregados para o diagnóstico do status existente, e não para a predição de futuros resultados” (Anastasi & Urbina, 2000, p.112). Cronbach (1996) refere ainda que faz sentido fazer um estudo concorrente quando se propõe um teste como substituto dum outro, que se apresente com características fracas (como ser dispendioso, por exemplo), esperando-se que o teste que se está a propor como substituto indique o mesmo que o outro teste indicaria a respeito dos indivíduos.

Existe uma série de critérios que são habitualmente utilizados, dependendo, obviamente, do tipo de teste que se está a validar, e são segundo Pasquali (2003): O desempenho académico (para a validação de testes de inteligência, por exemplo); o desempenho no treino de funções específicas; o desempenho profissional; o diagnóstico psiquiátrico (para a validação de testes de personalidade, nos quais se meçam perturbações psiquiátricas); o diagnóstico subjectivo; e outros testes disponíveis.

No que diz respeito ao último critério referido, que é de extrema importância para o presente trabalho, os dados resultantes de outro teste (que se considere válido e que meça o mesmo construto teórico que o teste que se está a estudar, validar e/ou adaptar) servem de critério para validar o “novo” teste. Se assim for, “Espera-se neste caso que a correlação do novo teste seja elevada, de pelo menos 0,75” (Pasquali, 2003, p.187), sendo que a esta correlação, entre o score dum teste e o score de uma medida de critério, se chama coeficiente de validade. No entanto, segundo Anastasi e Urbina (2000, p.118), “...essas correlações devem ser moderadamente altas, mas não altas demais. Se o novo teste apresenta uma correlação muito alta com um teste já existente, sem vantagens extras como brevidade ou facilidade de aplicação, o novo teste representa uma duplicação desnecessária”.

Aqui surge uma questão importante, que se liga ao objectivo deste trabalho: Para quê criar, ou neste caso estudar as características psicométricas dum outro teste se já existe um no mercado que mede válida e eficazmente aquilo que se pretende medir? A resposta a esta questão prende-se basicamente a questões relacionadas com o uso de medidas universais (como o BDI-II) e de economia – economia de tempo e de dinheiro, isto é, caso o novo teste não se apresente como mais simples, mais breve e menos dispendioso, não faz sentido algum criar ou validar e/ou adaptar um novo teste; contrariamente, se existirem vantagens claras sobre o teste antigo, então um novo teste deve “surgir” – entre outras questões.

Validade de conteúdo.

A validade de conteúdo não é calculada (Hill & Hill, 2008); é uma validade de carácter mais teórico, como refere Ribeiro (2008), sendo que se considera que um teste tem validade de conteúdo adequada se ele for considerado uma amostra representativa do domínio de comportamento que está a ser medido, na qual todos os aspectos importantes desse domínio de comportamento estão representados nos itens de forma proporcionalmente correcta (Anastasi & Urbina, 2000), ou seja, há validade de conteúdo se “os atributos inclusos no teste são uma amostra representativa dos que se pretendem medir” (Ribeiro, 2008, p.94).

É de referir que a validade de conteúdo “...depende da relevância das respostas de teste do indivíduo para a área de comportamento em consideração, e não da aparente relevância do conteúdo dos itens” (Anastasi & Urbina, 2000, p.108). A este respeito convém distinguir a validade de conteúdo da validade aparente – a validade aparente, apesar de ser

uma característica desejável nos testes, pois garante, de certa forma, a adesão aos mesmos, “...não se refere àquilo que o teste realmente mede, mas àquilo que ele superficialmente parece medir” (Anastasi & Urbina, 2000, p.110).

O processo de validação de conteúdo está presente desde o início, isto é, desde que se inicia a construção do teste. Pasquali (2003, p.189) refere mesmo que “A validade de conteúdo de um teste é praticamente garantida pela técnica de construção do mesmo”, e que esta técnica se baseia nos seguintes passos: (1) definição do domínio cognitivo; (2) definição do universo de conteúdo; (3) definição da representatividade de conteúdo; (4) elaboração da tabela de especificação, relacionando os conteúdos com os processos cognitivos, assim como a importância relativa de cada um deles; (5) construção do teste; (6) análise teórica dos itens; e, finalmente, (7) análise empírica dos itens. É de referir que a análise teórica dos itens se baseia na verificação da compreensão dos itens (análise semântica) e da pertinência dos mesmos (análise de juízes). Já a análise empírica dos itens visa determinar os seus graus de dificuldade e de discriminação, entre outros parâmetros, assunto vai ser discutido mais à frente no tópico sobre a análise dos itens.

Murphy e Davidshofer (1998, in Ribeiro, 2008) propõe um método para avaliar a validade de conteúdo que consiste, numa primeira fase, em descrever o conteúdo do domínio de comportamento a avaliar (ou construto); de seguida determinar as áreas do conteúdo do domínio de comportamento que cada item avalia; e, finalmente, comparar os dados dos dois primeiros passos (estrutura do teste vs. estrutura do conteúdo do domínio de comportamento). Consistindo este último passo num julgamento, a forma mais adequada para o fazer é através da análise de juízes especialistas no domínio de comportamento em questão – havendo concordância entre os juízes de que os itens avaliam de facto tal domínio de comportamento, fica demonstrada a adequabilidade da validade de conteúdo.

Fiabilidade

A fiabilidade, tal como referido anteriormente, diz respeito à capacidade do teste medir de forma repetida e consistente (Hill & Hill, 2008), ou seja, indica se os scores obtidos pelos mesmos indivíduos quando retestados (testados noutras ocasiões) com o mesmo teste ou com conjuntos equivalentes de itens, são idênticos, assumindo-se um certo erro (Ribeiro, 2008). A seguinte passagem de Pasquali (2003, p.192) ilustra bem o que isto significa:

A fidedignidade ou a precisão de um teste diz respeito à característica que ele deve possuir, a saber, a de medir sem erros... Medir sem erros significa que o mesmo teste, medindo os mesmos sujeitos em ocasiões diferentes, ou testes equivalentes, medindo os mesmos sujeitos na mesma ocasião, produzem resultados idênticos, isto é, a correlação entre estas duas medidas deve ser de 1. Entretanto, como o erro está sempre presente em qualquer medida, esta correlação se afasta tanto do 1 quanto maior for o erro cometido na medida. A análise da precisão de um instrumento psicológico quer mostrar precisamente o quanto ele se afasta do ideal de correlação 1, determinando um coeficiente que, quanto mais próximo de 1, menos erro o teste comete ao ser utilizado.

Outra possível interpretação do significado da fiabilidade dum teste é que ela nos "...indica a extensão em que as diferenças individuais na realização do teste são atribuíveis a diferenças «verdadeiras» ou reais, das características avaliadas” (Ribeiro, 2008, p.92-93).

Segundo Pasquali (2003), existem três tipos de delineamentos experimentais e dois tipos de análises estatísticas para se estimar o coeficiente de fiabilidade. Em relação aos tipos de análises estatísticas que são usadas na estimação do coeficiente de fiabilidade, apesar de existirem muitos índices de fiabilidade, elas resumem-se a duas: A correlação e a técnica alfa (ver Anexo B). Os delineamentos experimentais são os seguintes:

- *Uma amostra de indivíduos, um único teste e uma única situação de testagem* – os dados são analisados no que se refere à consistência interna dos itens através das seguintes análises estatísticas: Análise das duas metades (ou split-half) ou análise pelas técnicas alfa;
- *Uma amostra de indivíduos, dois testes e uma única situação de testagem* – a análise dos dados passa pela correlação entre os scores obtidos nas formas paralelas do teste e/ou conjuntos equivalentes de itens;
- *Uma amostra de indivíduos, um único teste e duas situações de testagem* – analisam-se os dados com base na correlação entre os scores obtidos nas duas situações de testagem (teste-reteste).

Convém ainda referir que existe uma outra forma de estimativa da fiabilidade, e é ela a fiabilidade inter-avaliador ou de cotador (Anastasi & Urbina, 2000; Ribeiro, 2008). Aqui, a fiabilidade é estimada através da correlação entre os scores de dois ou mais avaliadores ou juízes, que são convidados a avaliar o mesmo teste.

Torna-se necessário neste momento fazer um reparo em relação a um aspecto com alguma relevância. Tem-se referido até agora que a fiabilidade se estima, e não que se mede, mas ainda não se justificou porquê. É impossível fazer uma medição precisa da fiabilidade, pois as amostras que são usadas para este fim nunca são amostras perfeitas, logo, os coeficientes de fiabilidade variam de amostra para amostra, pelo que tem-se sempre que tratar estes dados como estimativas (Hill & Hill, 2008).

Existem algumas variáveis, ou factores, que poderão influenciar a fiabilidade – para além dos factores internos do próprio teste, como as características dos itens, existem factores externos ao conteúdo do mesmo, como a variabilidade da amostra e o comprimento do próprio teste (Pasquali, 2003). No que se refere à amostra, sabe-se que quanto maior e mais variável for, maior será o coeficiente de correlação e logo o de fiabilidade. Da mesma forma, o comprimento do teste também tem implicações na fiabilidade – “quanto maior número de itens tiver o teste, maior será seu índice de precisão, pois o erro tende a zero quando o número se aproxima do infinito, segundo o famoso teorema de Bernoulli” (Pasquali, 2003, p.222).

Finalmente, resta acrescentar que enquanto que é possível um teste ter uma boa fiabilidade e pouca validade, é impossível, sem uma fiabilidade adequada, um teste ter uma boa validade (Hill & Hill, 2008).

Análise dos Itens

A análise dos itens constitui-se uma das importantes formas de se avaliar a qualidade dos testes psicológicos na medida em que, sendo os itens a representação comportamental do traço latente/construto teórico que está a ser medido, a sua análise leva ao conhecimento do próprio traço latente. A análise da adequabilidade dos itens permite então verificar se a sua contribuição é suficiente para definir a variável latente em questão (Hill & Hill, 2008).

Segundo Pasquali (2003), existem duas formas de analisar os itens: Teórica e empiricamente. A análise teórica é feita por juízes e é de dois tipos: Análise semântica, na qual os juízes são os próprios indivíduos que constituem a amostra para a qual se cria o teste, e com a qual se pretende estabelecer a dificuldade de compreensão dos itens [esta não se deve “constituir em factor complicador na resposta dos indivíduos, dado que não se quer medir a

compreensão deles (a não ser, obviamente, que o teste queira medir precisamente isso), mas sim a magnitude do atributo a que os itens se referem” (Pasquali, 2003, p.107)]; e análise de conteúdo, na qual os juízes são peritos na área do construto que está a ser medido, e com a qual se pretende demonstrar que os itens são pertinentes para o construto em questão, isto é, se eles são uma representação adequada do construto/traço latente.

A análise empírica (ou estatística) baseia-se na avaliação de uma série de parâmetros que os itens devem apresentar para se constituírem adequadamente como uma representação comportamental daquilo que o teste pretende medir. Os parâmetros em questão são, segundo Pasquali (2003): A unidimensionalidade, a dificuldade, a discriminação (ou sensibilidade), os vieses e/ou tendências de resposta, a validade e a precisão. A avaliação destes parâmetros é realizada com base nos dados da amostra para a qual se cria o teste, através de uma série de metodologias estatísticas, que vão desde uma análise geométrica (gráfica)³ a uma análise algébrica (Pasquali, 2003), na qual o presente estudo se vai centrar (ver Anexo C).

A unidimensionalidade prende-se com o facto dos itens medirem um mesmo e único construto teórico/traço latente, e é, segundo Pasquali (2003), uma condição necessária para análises posteriores dos itens, como a análise da sua dificuldade e da sua discriminação.

A dificuldade dos itens define-se normalmente em termos de percentagem (proporção) de indivíduos que responde de forma acertada aos itens – no caso de testes de aptidão em que as respostas são do tipo certo-errado (Anastasi & Urbina, 2000; Pasquali, 2003; Ribeiro, 2008). Quando nos deparamos com outro tipo de teste, como os testes de personalidade, em que as respostas são do tipo aceitar ou escolher determinada opção de resposta, a dificuldade dos itens é definida em termos do próprio traço latente medido, sendo proporcional à quantidade de traço latente necessária para que os itens sejam aceites – quanto mais traço latente for necessário para que um item seja escolhido mais difícil ele é (Pasquali, 2003).

O parâmetro discriminação (ou sensibilidade) refere-se à capacidade dos itens para diferenciar os indivíduos no que toca ao traço latente que o teste mede, ou seja, os itens têm

³ A análise gráfica dos itens baseia-se no modelo da TCT e é habitualmente utilizada com testes de aptidão em que as respostas são, normalmente, do tipo certo-errado. Através desta análise recolhem-se informações sobre os seguintes parâmetros: Dificuldade, discriminação e “chute” – para mais informações sobre esta análise ver Pasquali (2003).

capacidade de discriminação quando diferenciam os indivíduos que obtêm scores elevados dos que obtêm scores baixos (Anastasi & Urbina, 2000; Pasquali, 2003; Ribeiro, 2008).

A validade dos itens refere-se ao facto deles medirem aquilo que pretendem medir. Para a TCT comprova-se através da correlação item-critério, que deverá ser elevada para que se considere um item válido, e este critério deverá ser um teste paralelo ao teste que está a ser validado e que contém o próprio item em questão (Pasquali, 2003) – para Maroco (2005) esta validade é a validade convergente do instrumento. Para a TRI existe validade dos itens quando eles representam de forma adequada o traço latente, e esta verifica-se através do seu peso factorial, que quanto maior, maior será a validade do próprio item, pois maior a sua representatividade no factor (Pasquali, 2003) – a esta validade vários autores (e. g. Anastasi & Urbina, 2000; Hill & Hill, 2008) chamam validade factorial do instrumento.

Finalmente, uma última questão em relação aos itens deve ser tida em conta – as respostas enviesadas. Esta é uma questão importante, na medida em que mesmo quando a qualidade dos itens é boa, ela pode ser posta em causa por factores alheios a eles, factores esses que se relacionam com os próprios indivíduos testados. Para Pasquali (2003) as respostas enviesadas podem ser causadas por três factores: Culturais/dependentes do nível sócio-económico; respostas aleatórias e, por último, respostas estereotipadas.

Antes de se passar à abordagem da depressão e para concluir, um bom instrumento de avaliação psicológica deverá ter propriedades psicométricas rigorosas que, habitualmente, se reportam a normas. No entanto, é de referir que em contexto clínico são as características individuais de cada um, o que há de único nos indivíduos, que são privilegiadas. Assim sendo, alguns instrumentos que satisfazem condições psicométricas não são satisfatórios na tomada de decisões clínicas. Por este motivo surgiu em contexto de saúde o conceito de “climétrica”, querendo com ele designar “a ciência de quantificação dos fenómenos clínicos com particular atenção para a validação das variáveis de resultado” (Wijisman, Hekster, Keyser, Renier & Meinardi, 1991, citados por Ribeiro, 2008, p.99-100). Para além de propriedades psico e clinimétricas, um teste deverá ter também outras características, principalmente em contexto de saúde (em que os indivíduos estão, de certa forma, fragilizados), entre elas: Apropriabilidade, economia, adequabilidade, interpretabilidade, responsividade, aceitabilidade e utilidade (Ribeiro, 2008).

Depressão

É normal ao longo da vida de qualquer ser humano existirem períodos de tristeza, de desânimo e de abatimento físico e psicológico. Estes períodos, que são intrínsecos à experiência humana, ajudam os indivíduos, de certa forma, a equilibrar a sua vida, o seu bem-estar emocional e físico. No fundo, em certas situações, como quando o humor triste é uma reacção a um acontecimento negativo ou desagradável, estes períodos de tristeza podem ser adaptativos. No entanto, quando dominam, de forma permanente, o funcionamento psicológico dos indivíduos, provocando mal-estar significativo e prejuízos no funcionamento social, profissional, ou noutros domínios da vida dos indivíduos, passa-se da normalidade para a psicopatologia, ou seja, passa-se para a esfera da depressão propriamente dita.

Duma maneira geral, considera-se que se está no domínio da depressão sempre que ao humor depressivo ou à perda e/ou diminuição de prazer ou do interesse nas actividades habituais, se associam quatro ou mais dos seguintes sintomas, sintomas esses que devem persistir durante a maior parte do dia, quase todos os dias, durante pelo menos duas semanas, e “que representam uma alteração do funcionamento prévio” (American Psychiatric Association, 2002, p.356) do indivíduo:

- Alterações no apetite ou peso – perda ou ganho significativo de peso e/ou diminuição ou aumento de apetite;
- Alterações no sono – insónia ou hipersónia;
- Alterações na actividade psicomotora – agitação ou lentificação psicomotora;
- Perda de energia ou fadiga;
- Sentimentos de desvalorização pessoal ou culpa excessiva ou inapropriada;
- Diminuição da capacidade de pensamento, concentração ou decisão;
- Pensamentos recorrentes acerca da morte, ideação suicida ou tentativas de suicídio.

A junção destes sintomas promove então um mal-estar geral significativo no indivíduo deprimido, para além de o incapacitarem, de forma mais ou menos acentuada, de funcionar de forma eficaz nas várias áreas da sua vida – que abrangem “cognição, comportamento, afecto e

funcionamento físico” (Hammen, 1997, in Cláudio, 2004, p.67), ou seja, que o incapacitam de funcionar de forma eficiente tanto a nível pessoal, social ou profissional.

A depressão consiste então num “...afecto doloroso específico composto de tristeza, de mal-estar, de dor moral, de perda de prazer com tudo o que anteriormente era considerado agradável e, finalmente, por uma agressividade mais ou menos expressa contra si (culpabilidade) ou contra os outros (irritabilidade)” (Braconnier, 2007, p.161).

Existem inúmeras teorias sobre a depressão. Das mais remotas, destaca-se a teoria dos quatro humores de Hipócrates, que durante muito tempo serviu de base para a descrição dos sintomas clínicos da depressão, na altura denominada melancolia: “Ânimo entristecido, sentimento de um abismo infinito, extinção do desejo e da fala, impressão de hebetismo, seguida de exaltação, além de atracção irresistível pela morte, pelas ruínas, pela nostalgia e pelo luto” (Roudinesco & Plon, 2000, p.499).

Na teoria psicanalítica, Freud postula, no seu texto “Luto e Melancolia” (1917, citado por Roudinesco & Plon, 2000), que no luto o indivíduo consegue desligar-se de forma gradual do objecto perdido, enquanto que na melancolia (forma patológica do luto) o indivíduo não consegue fazer o luto do objecto perdido, pois culpabiliza-se pela sua morte, negando-a e ficando “possuído pelo morto ou pela doença que originou a sua morte. Em suma, o ego identifica-se com o objecto perdido, a ponto de ele mesmo se perder no desespero infinito de um nada irremediável” (Freud, 1917, in Roudinesco & Plon, 2000, p.499).

Ainda na teoria psicanalítica, Matos (2001a) refere que não é a perda dos objectos, das pessoas propriamente ditas (o que caracteriza o luto), que caracteriza a depressão, mas sim a perda do amor dos objectos significativos. A depressão gera-se na “relação de amor não correspondido” (Matos, 2001b, p.497) com o objecto, que é culpabilizante, desamante e inferiorizante (desnarcisante), na qual o indivíduo idealiza o objecto, projecta a sua bondade e introjecta tanto a maldade do objecto como a culpa que este lhe projecta. O que resulta desta relação é o abatimento, os sentimentos de culpa e inferioridade, e a baixa auto-estima.

Poderia-se continuar a citar inúmeros autores tanto da perspectiva psicanalítica, como de outras perspectivas, mas o presente estudo vai-se focar na perspectiva cognitiva da depressão. Duma maneira geral, as teorias cognitivas da depressão sugerem que os

pensamentos, as inferências, atitudes e interpretações, bem como a forma como os indivíduos vêem e/ou interpretam e recordam os acontecimentos determinam a sua resposta emocional aos mesmos; assim sendo, as cognições têm um papel crucial no quanto as pessoas são afectadas por experiências negativas, e determinam se estas experiências serão seguidas de uma rápida recuperação ou de episódios recorrentes de depressão (Joormann, 2009). Estas teorias partem então do pressuposto que a investigação do conteúdo da cognição e da natureza dos processos cognitivos é crucial para o entendimento da génese e manutenção da depressão, assumindo que estas se devem à interacção entre uma vulnerabilidade psicológica e um factor precipitante (acontecimento negativo) inesperado (Joormann, 2009).

O estudo da interacção entre cognição e emoção na depressão tem já uma longa história. Inicialmente focou-se na demonstração de que o conteúdo dos pensamentos dos indivíduos deprimidos difere do dos não deprimidos, e de que os indivíduos deprimidos exibem deficits cognitivos e um processamento disfuncional do material emocional (Joormann, 2009). Estes estudos iniciais deram provas cabais e consistentes de que a depressão é caracterizada por pensamentos automáticos negativos sobre o self, o futuro e o mundo (Ingram, Miranda & Segal, 1998; Mathews & MacLeod, 2005; in Joormann, 2009).

Mais recentemente, o estudo da interacção entre cognição e emoção na depressão tem sublinhado o papel que os processos cognitivos desempenham na regulação das emoções e do humor (e.g., Larsen, 2000, in Joormann, 2009), tendo a depressão sido conceptualizada como uma perturbação de desregulação emocional. Desta forma, tem-se sugerido que não é tanto uma resposta desadequada a um acontecimento negativo que caracteriza os indivíduos vulneráveis à depressão, mas mais uma incapacidade para regular a duração e a intensidade do afecto negativo resultante desse acontecimento (e.g., Teasdale, 1988, in Joormann, 2009).

Nestes estudos mais recentes, tem-se demonstrado, por exemplo, que os indivíduos que respondem a acontecimentos negativos e a estados de humor negativos de forma desadequada, como com ruminação, tendem a viver episódios de depressão recorrentes (e.g. Nolen-Hoeksema, 2000; Nolen-Hoeksema, Wisco & Lyubomirsky, in press; in Joormann, 2009). Estes indivíduos mantêm o seu estado de humor negativo (Nolen-Hoeksema et al., in press, in Joormann, 2009), na medida em que este se associa à activação de representações congruentes com o estado de humor na memória de trabalho (e.g. Siemer, 2005, in Joormann, 2009), relacionando-se, desta forma, a pensamentos negativos frequentes, a uma atenção

selectiva para estímulos negativos, e a uma maior acessibilidade a memórias negativas (Blaney, 1986; Mathews & MacLeod, 2005; in Joormann, 2009). Assim, a ocorrência frequente de ruminação não intencional (juntamente com os pensamentos e as memórias negativas incontrolláveis), tem sido considerada como responsável pelo aumento do risco de aparecimento da depressão, pela sua manutenção e pelo aumento do risco da sua recorrência (Joormann, 2009).

Após esta breve exposição sobre o estudo da interação entre cognição e emoção, passa-se para uma breve exposição das teorias cognitivas da depressão. Joormann (2009), na sua revisão de literatura acerca dos modelos cognitivos da depressão, diferencia as teorias que se baseiam nos esquemas e no estilo atributivo; as que se baseiam nos recursos disponíveis e na interferência afectiva; e as que se focam na inibição e no controlo mental. No primeiro conjunto de teorias encontram-se as que, no fundo, têm guiado o estudo da cognição na depressão (Joormann, 2009): A teoria cognitiva de Beck, a teoria do desamparo aprendido de Seligman (e subsequentes reformulações) e a teoria das redes associativas de Bower. No segundo conjunto estão a hipótese dos recursos disponíveis de Ellis e Ashbrook e a hipótese da interferência afectiva de Siegle, Ingram e Matt. Finalmente, têm-se as teorias, como a de Hertel, que se baseiam na inibição e no controlo mental.

Teoria Cognitiva de Beck

O modelo cognitivo da depressão de Beck sofreu ao longo dos tempos inúmeras alterações (1963, 1964, in por Cláudio, 2004). Na sua base estão observações clínicas sistemáticas e intervenções experimentais (Beck, 1963, 1964, 1967, in Beck, Rush, Shaw & Emery, 1979), que lhe permitiram passar duma visão simplista, na qual a relação entre cognição e emoção era do tipo causa-efeito, para uma visão mais complexa, na qual a cognição não é a causa da emoção (perturbações psicológicas), mas sim um dos mecanismos que interactivam neste processo, a par de outros, como biológicos, psicológicos e sociais (Power & Dalgleish, 2008).

Existe neste modelo uma visão dimensional da depressão, na qual há “o extremar de aspectos cognitivos, comportamentais e emocionais normais” (Cláudio, 2004, p.69), ou seja, não se considera a existência de descontinuidade entre o normal e o patológico, o que existe é “...um contínuo, com variabilidade de severidade e de persistência de sintomas, entre o humor triste não patológico e a depressão major” (Cláudio, 2004, p.70). Resumindo,

...nesta perspectiva a depressão seria caracterizada por uma presença persistente do humor negativo, da perda de prazer, de uma inibição comportamental e de alterações fisiológicas. Os sintomas relacionados com estes aspectos poderiam apresentar-se de forma diferente, quer na severidade, quer na evolução, o que indicaria a existência de um intervalo entre sintomatologia grave e ausência de sintomas. Os sintomas da depressão teriam implicações negativas em todas as esferas de vida dos sujeitos e poderiam estar associados a uma manutenção da depressão e também às recorrências que se observam (Clark, Beck & Alford, 1999, in Cláudio, 2004, p.70).

Beck (1976, in Joormann, 2009) considera que as representações existentes na memória, ou esquemas, levam os indivíduos a filtrar os estímulos do meio ambiente, de tal forma que a sua atenção é direccionada para informações que são congruentes com esses esquemas. O que se passa com os indivíduos deprimidos é que os seus esquemas são disfuncionais e negativos (Beck et al., 1979), e incluem temas de perda, separação, fracasso, desvalorização pessoal, e rejeição; consequentemente, estes indivíduos exibem um enviesamento sistemático no processamento da informação que é relevante para estes temas, ou seja, direccionam selectivamente a sua atenção para estímulos negativos no ambiente que os rodeia, e interpretam estímulos neutros e ambíguos de uma forma congruente com os esquemas existentes (Joormann, 2009).

Quando os esquemas disfuncionais são accionados por acontecimentos negativos, são geradas cognições negativas específicas que tomam a forma de pensamentos automáticos que se baseiam em visões pessimistas acerca de si, do mundo, e do futuro (Joormann, 2009) – a chamada tríade cognitiva. Estes pensamentos negativos automáticos são, no caso dos indivíduos deprimidos, dificilmente descartados e são, inclusivamente, suportados por erros cognitivos (Power & Dalgleish, 2008) – como inferências arbitrárias, abstracção selectiva, generalização abusiva, magnificação e minimização, personalização, e pensamento dicotómico e absolutista (Beck et al., 1979), que mais não fazem se não manter e reforçar a crença da validade dos pensamentos negativos, independentemente de existirem provas contrárias (Beck, 1967, in Beck et al., 1979). Como Beck et al. (1979, p.13) referem,

In milder depressions the patient is generally able to view his negative thoughts with some objectivity. As depression worsens, his thinking becomes increasingly dominated by negative ideas, although there may be no logical connection between actual situations and his negative interpretations. As the prepotent idiosyncratic schemas lead to distortions of reality and consequently to systematic errors in the depressed person's thinking, he is less able to entertain the notion that this negative interpretations are erroneous. In the more severe states of depression, the patient's thinking may become completely dominated by the idiosyncratic schema: He is completely preoccupied with perserverative, repetitive negative thoughts and may find it enormously difficult to concentrate on external stimuli (...) or engage in voluntary mental activities (...). In such instances we infer that the idiosyncratic cognitive organization has become autonomous. The depressive cognitive organization may become so independent of external stimulation that the individual is unresponsive to changes in his environment.

Finalmente, o modelo presume que os esquemas negativos e disfuncionais, bem como os enviesamentos, podem permanecer latentes, significando que estão menos acessíveis e que não influenciam a cognição e afecto, desde que não sejam activados por um acontecimento negativo; uma vez activados, iniciam e mantêm o estado de depressão através dum ciclo vicioso de aumento de pensamentos e afectos negativos, representando, desta forma, factores estáveis de vulnerabilidade para novas depressões (Joormann, 2009). Resumindo, na concepção de Beck o esquema disfuncional do indivíduo deprimido é dominado por cognições negativas (crenças e/ou suposições) sobre si, o mundo e o futuro, e o encadeamento destas cognições (pensamentos negativos automáticos) representa o factor de vulnerabilidade cognitiva à depressão (Braconnier, 2007).

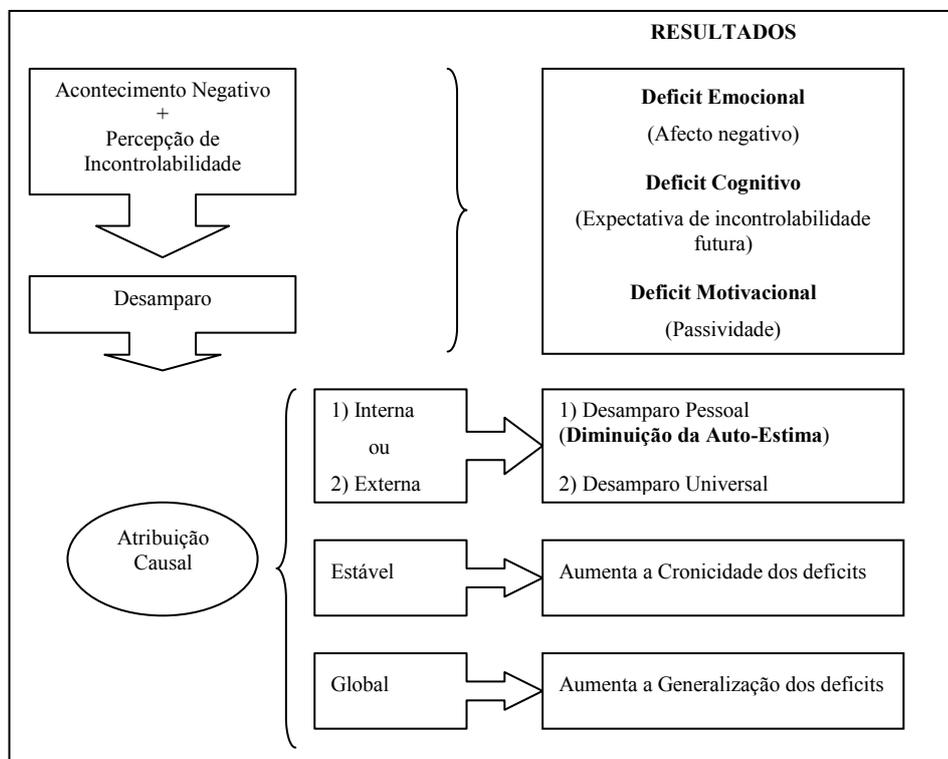
Teoria do Desamparo Aprendido e Suas Reformulações

Tal como a teoria cognitiva de Beck, a teoria do desamparo aprendido de Seligman (1975, in Power & Dalgleish, 2008) também sofreu algumas modificações. Na sua base está o conceito de desamparo aprendido, que diz que as expectativas acerca da incontrolabilidade sobre os eventos resulta em episódios de depressão (Joormann, 2009), ou seja, são as expectativas de que os resultados de determinada situação não dependem dos esforços do indivíduo, as principais causas do desamparo aprendido e dos deficits motivacionais, cognitivos e emocionais, que estão patentes na depressão (Abramson, Seligman & Teasdale,

1978). Numa tentativa de ultrapassar algumas críticas [como a não consideração do deficit ao nível da auto-estima e da tristeza, típicos da perturbação depressiva, entre outras críticas (Power & Dalgleish, 2008)], Abramson et al. (1978) reformularam a teoria inicial.

Nesta reformulação, introduzem a teoria da atribuição de Weiner, pelo que, apesar do desamparo continuar a depender da percepção e expectativa de incontrolabilidade, os efeitos subsequentes são agora vistos como resultantes do tipo e da importância da própria situação, juntamente com a explicação que o indivíduo dá como causa da situação (Power & Dalgleish, 2008). Segundo Abramson et al. (1978) os deficits motivacionais, cognitivos e afectivos típicos da depressão dependem duma combinação particular de atribuições causais feita no seguimento dum acontecimento negativo, ou seja, existe um estilo atributivo típico dos indivíduos deprimidos ou vulneráveis à depressão: Estes tendem a avaliar as causas dos acontecimentos negativos como sendo internas, estáveis e globais; contrariamente às causas de acontecimentos positivos que tendem a ser vistas como externas, instáveis e específicas (Abramson et al., 1978). Na figura 1 encontra-se uma síntese esquemática desta teoria.

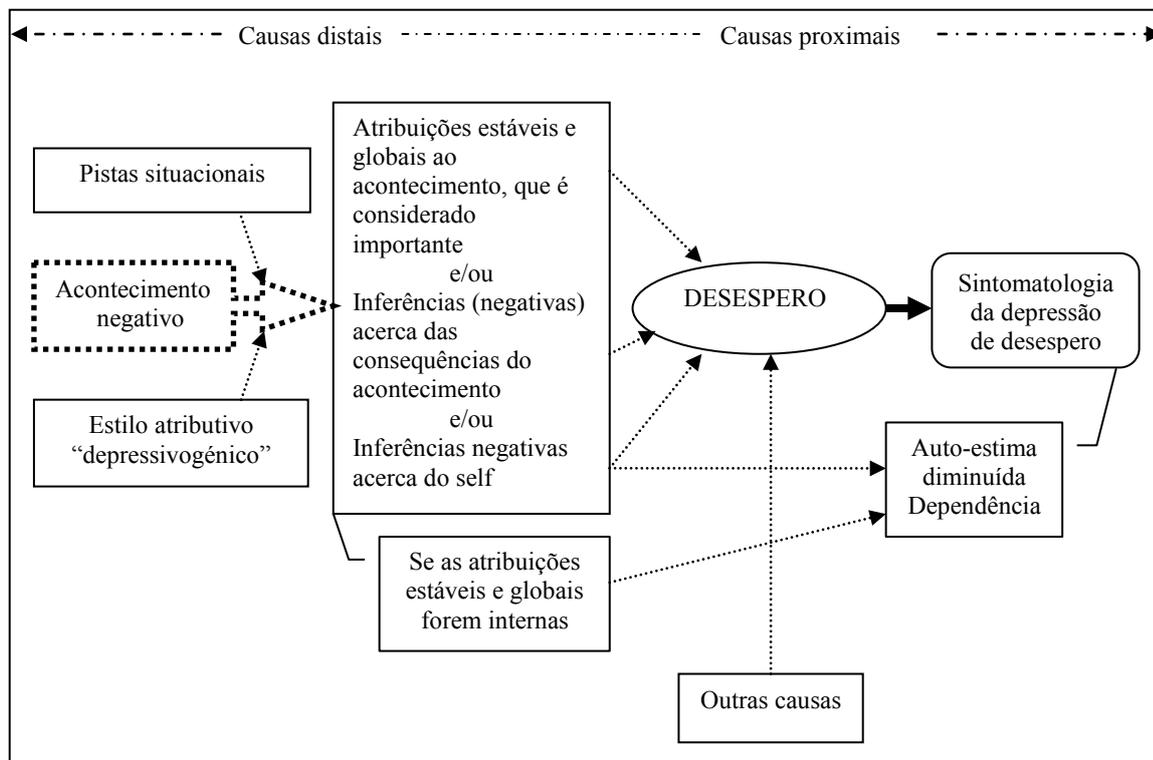
Figura 1 – Esquema da teoria reformulada do desamparo aprendido⁴



⁴ Adaptado de Power e Dalgleish (2008).

Volvidos aproximadamente 10 anos, a teoria sofre uma nova reformulação, sendo denominada de teoria da depressão de desespero (Abramson, Metalsky & Alloy, 1989). Nesta, propôs-se que o desespero [expectativa de que resultados altamente desejados não vão acontecer ou de que resultados altamente adversos são certos (Joormann, 2009)] é causa proximal e suficiente para a ocorrência de depressão. O desespero é uma consequência de atribuições causais estáveis e globais a acontecimentos negativos que são valorizados, e/ou de inferências negativas acerca das consequências desses acontecimentos, bem como acerca do self, que têm de ser vistas como importantes, irremediáveis, amplas e inalteráveis (Abramson et al., 1989). As atribuições a estes acontecimentos a causas internas leva ainda a uma baixa auto-estima e a sentimentos de desvalorização pessoal, que reforçam a sintomatologia da depressão (Joormann, 2009). Abramson et al. (1989) referem que a sua teoria não se aplica a todas as formas de depressão; pelo contrário, ela pode representar um subtipo importante de perturbação depressiva (Alloy, Abramson, Walshaw & Neeren, 2006, in Joormann, 2009). A representação esquemática da Figura 2 facilita a compreensão desta teoria.

Figura 2 – Cadeia de acontecimentos na teoria da depressão de desespero⁵



⁵ Adaptado de Abramson et al. (1989).

Teoria das Redes Associativas de Bower

Bower (1981, in Joormann, 2009) afirma que as redes associativas, que contêm numerosas ramificações (cada uma com representações semânticas específicas que podem ser activadas por estímulos do ambiente), levam a enviesamentos cognitivos nos indivíduos deprimidos. A activação de qualquer uma destas ramificações provoca uma activação parcial, de todas as outras ramificações dentro da rede associativa através de um processo de activação alastrada; assim sendo, as representações das ramificações parcialmente activadas requerem uma menor activação para ocorrer do que as representações não activadas, o que resulta num processo de vantagem para estímulos que estão relacionados com estas representações parcialmente activas (Joorman, 2009). O que acontece no indivíduo deprimido é que ele “activa «núcleos emotivos» negativos a nível da memória, criando um círculo vicioso em que quanto mais está deprimido mais pensamentos negativos tem” (Braconnier, 2007, p.254). Assim como Beck, Bower também afirma que as redes associativas são estáveis e que o enviesamento da atenção dos indivíduos deprimidos tende a permanecer para além do episódio de depressão (Joormann, 2009).

Como refere Braconnier (2007, p.254), qualquer uma das três teorias referidas “pressupõe que o indivíduo responde mais às suas representações internas de uma situação do que à situação em si”. No entanto, enquanto que na teoria da depressão de desespero se fazem inferências acerca do conteúdo da cognição dos deprimidos, na teoria cognitiva de Beck e na teoria das redes associativas de Bower fazem-se inferências acerca dos próprios processos cognitivos (Joormann, 2009). Estas últimas (a teoria cognitiva de Beck e teoria das redes associativas de Bower) aproximam-se, pois ambas afirmam que os indivíduos deprimidos se caracterizam por enviesamentos cognitivos em todos os aspectos do processamento da informação (percepção, atenção, memória e pensamento), que facilitam o processamento de informação com valência negativa; e ambas consideram que os enviesamentos permanecem latentes, o que caracteriza o funcionamento dos indivíduos que, ainda que não deprimidos no momento, estão vulneráveis a viver episódios de depressão (Joormann, 2009).

Outras Teorias Cognitivas da Depressão

Na revisão de literatura acerca dos modelos cognitivos da depressão, Joormann (2009) diferencia as teorias que se baseiam nos esquemas e no estilo atributivo já abordadas, das que se baseiam nos recursos disponíveis e na interferência afectiva e das que se focam na inibição e no controlo mental. A hipótese dos recursos disponíveis postula que existe um limite de quantidade de recursos disponíveis para as operações cognitivas, e que a depressão inibe, ou reduz funcionalmente, estes recursos; assim, tendo a sua capacidade cognitiva reduzida, os indivíduos deprimidos têm dificuldades ao nível da memória e em iniciar outros processos cognitivos que exijam algum esforço (e.g. Ellis & Ashbrook, 1988, in Joormann, 2009). De forma semelhante, a hipótese da interferência afectiva postula que, estando os indivíduos deprimidos preocupados com o processamento de informação emocional, a sua performance não é prejudicada em tarefas nas quais necessitem de processar aspectos emocionais dos estímulos, mas sofre prejuízos se tiverem que ignorar os aspectos emocionais e responder a outros aspectos da informação (Siegle, Ingram & Matt, 2002, in Joormann, 2009).

Hertel (2004, in Joormann, 2009) propôs que a depressão se caracteriza por um reduzido controlo cognitivo que prejudica a inibição de respostas automáticas (ruminação e pensamentos negativos sobre o self). Ultrapassar estas respostas e dirigir a atenção para o que é necessário e relevante numa tarefa (função dos processos inibitórios e do controlo cognitivo), é então algo difícil quando se apresenta disfunções nos processos inibitórios, como é o caso dos indivíduos deprimidos (Joormann, 2009). Desta forma, os indivíduos que exibem um deficit ao nível do controlo cognitivo estão mais vulneráveis à ruminação e a viver episódios de depressão. Segundo Joormann (2009), parece que o controlo cognitivo e a inibição conseguem explicar a maioria dos deficits cognitivos na depressão e o processamento preferencial de informação relevante à depressão; além disso, o controlo cognitivo parece ter também um papel importante na regulação emocional numa forma mais abrangente (e.g., Joormann, Yoon & Zetsche, 2007, in Joormann, 2009).

É de referir que há evidências empíricas recentes que apoiam consistentemente as referidas teorias cognitivas da depressão, assim como evidências contrárias ou pouco consistentes (Joormann, 2009). Existem autores que se opõem categoricamente a estas teorias, chegando mesmo a afirmar que elas (especificamente a teoria cognitiva de Beck, a teoria do desamparo aprendido e a teoria das redes associativas de Bower) têm uma base

conceptual inadequada, pois assentam em teorias da ciência cognitiva que são discutíveis, pelo que não se podem constituir como teorias da depressão adequadas (Champion & Power, 1995; Power & Champion, 1988). Outros referem que as teorias cognitivas da depressão, modelos simplistas da realidade, que assentam numa visão unidireccional da relação entre cognição e emoção, com ênfase excessiva no intraindividual, eram aceitáveis num estágio inicial de desenvolvimento da teoria da depressão; no entanto, consideram que depressa surgiu a necessidade de serem substituídas por modelos mais complexos, onde se incluíssem não só cognições e factores precipitantes, mas também outras variáveis, como suporte social, capacidades de coping, entre outras, no fundo que incluíssem factores interpessoais e sociais (Hammen, 2006).

As referidas perspectivas que se opõem às teorias cognitivas da depressão constituem-se naquilo que foi denominado por teorias sócio-cognitivas da depressão, que consideram que a vulnerabilidade às perturbações psicológicas não se localiza unicamente em factores internos do indivíduo ou em factores externos a ele, enfatizando a existência duma interacção complexa entre os dois, isto é, entre o interno e o externo (Power & Dalglish, 2008). Interacção complexa existe também entre depressão e ansiedade, que se aborda de seguida.

Ansiedade e Depressão – de Mãos Dadas ou de Costas Viradas?

Tanto a depressão como a ansiedade são perturbações mentais muito frequentes na população normal, havendo estudos epidemiológicos que revelam que “cerca de 15% da população em geral sofre quer de depressão quer de ansiedade durante um período de um ano, o que coloca estes quadros clínicos entre as doenças mais correntes” (Boyd & Weissman, 1981, in Montgomery, 2000, p.13).

A ligação entre as duas perturbações tem originado inúmeros estudos ao longo dos tempos. Esta ligação é inquestionável, pois é difícil encontrar um indivíduo deprimido que não apresente sintomas de ansiedade, apesar de continuar a haver tentativas de separação entre as duas perturbações, separação que torna inevitavelmente difícil o diagnóstico dos indivíduos que apresentam estados mistos de ansiedade e depressão (Burke, Regier &

Christie, 1988; Weissman, Leaf, Tischler, Blazer, Karno, Livingston & Florio, 1988; Angst & Dobler-Mikola, 1985; in Montgomery, 2000).

A discussão sobre a existência dum continuum nas perturbações mentais ou a existência de inúmeras perturbações diferentes, fraccionadas, tem uma longa história (Montgomery, 2000). O que desde sempre justificou a separação entre ansiedade e depressão em dois estados psicopatológicos distintos foi o facto destas perturbações se diferenciarem “não só na sintomatologia, mas também no curso da doença e na resposta ao tratamento” (Montgomery, 2000, p.15).

No entanto, têm-se acumulado evidências que questionam a validade desta separação, na medida em que em vários estudos epidemiológicos se constatou haver uma clara sobreposição entre os dois estados psicopatológicos, sendo que muitos indivíduos com critérios de diagnóstico para a depressão também apresentam critérios de diagnóstico para a ansiedade (Angst & Dobler-Mikola, 1985; Breier, Charney & Heninger, 1984; in Montgomery, 2000). Verifica-se mesmo que a maioria dos indivíduos que recorre aos cuidados de saúde primários “sofre daquilo que parece ser um quadro misto de ansiedade e depressão e apenas um pequeno número se apresenta com ansiedade simples ou depressão simples” (Montgomery, 2000, p.15).

Além disso, mesmo nos estudos, como o de Mountjoy e Roth (1982, in Montgomery, 2000, p.17), “que dizem demonstrar uma separação entre os dois quadros, é referida sintomatologia bastante sobreponível na depressão e na ansiedade”. Aliás, pode-se mesmo constatar que a grande maioria dos sintomas representados nas escalas de ansiedade, são úteis para a avaliação da evolução e gravidade da depressão, facto que demonstra que a depressão inclui uma componente de ansiedade. Graziani (2005) refere inclusivamente que a ansiedade só se estabeleceu como perturbação psicológica no final do séc. XIX, pelo facto de em quase todas as doenças mentais poder aparecer uma componente de ansiedade.

Finalmente, um outro factor poderá indicar a existência duma perturbação subjacente comum entre ansiedade e depressão – a estabilidade do diagnóstico de depressão e a relativa instabilidade do diagnóstico de ansiedade (Montgomery, 2000). A verificação de que a estabilidade do diagnóstico de ansiedade (avaliada com a metodologia teste-reteste) só se deu em 10% dos casos, ao contrário do que aconteceu com os diagnósticos de depressão, que

tenderam a manter-se ao longo do tempo; e a verificação de que 24% dos diagnósticos de ansiedade se converteram em diagnósticos de depressão e 14% deles se converteram em estados mistos de ansiedade e depressão (Angst, Vollrath, Merikangas & Ernst, 1990, in Montgomery, 2000), pode levar à consideração que “em alguns casos, a ansiedade é uma forma de depressão ou é uma característica prodrômica de uma perturbação depressiva em evolução” (Montgomery, 2000, p.18).

Desta forma, Montgomery (2000) reconhece que para fins de investigação poderá ser proveitoso adoptar uma visão fraccionada das perturbações, principalmente quando se trata de avaliar aspectos específicos de cada uma delas; no entanto, na prática clínica, esta separação pode não ser tão proveitosa, por tudo o que já foi referido e que pode ser reforçado com o facto de até mesmo a terapêutica antidepressiva ser mais eficaz que a ansiolítica no tratamento de estados de ansiedade. É por todos estes motivos que a maioria dos médicos dos cuidados de saúde primários vêem a depressão e a ansiedade como uma só doença (Montgomery, 2000).

Avaliação da Depressão e Sua Importância

A depressão é uma perturbação mental grave, muito frequente, que provoca prejuízos nas mais diversas áreas da vida dos indivíduos, e que pode mesmo pôr a vida em perigo. Desta forma, a importância da sua avaliação clínica (diagnóstico) e do seu tratamento é inquestionável, e reforçada pelos dados que se seguem.

Há 35 anos atrás, estimava-se que 12% da população adulta já tinha tido ou teria, em algum momento da sua vida, um episódio de depressão que necessitasse de tratamento psicológico e/ou psiquiátrico (Schuyler & Katz, 1973, in Beck et al., 1979). No final da década de 70, e apesar dos avanços nos tratamentos farmacológicos para a depressão, não existiam evidências de que a prevalência da mesma tivesse diminuído, além disso, a taxa de suicídio (um dos índices da prevalência da depressão) tinha aumentado e não diminuído (Beck et al., 1979). Ainda na década de 70, Secunda, Katz, Friedman e Schuyler (1973, in Beck et al., 1979) revelaram, no seu relatório “Os Distúrbios Depressivos”, que a depressão era responsável por 75% das hospitalizações psiquiátricas e que num dado ano 15% dos adultos,

com idades entre os 18 e os 74 anos, podiam sofrer sintomas depressivos significativos. No mesmo relatório, os autores referiram que os custos com a depressão se situavam entre os 300 e os 900 milhões de dólares por ano.

Os dados mais recentes da American Psychiatric Association (2002) não são animadores: O risco de prevalência na população adulta para a perturbação depressiva grave tem variado entre os 10 e os 25% nas mulheres e entre os 5 e os 12% nos homens; a prevalência pontual da mesma perturbação tem variado entre 5 a 9% para as mulheres e entre 2 a 3% para os homens (taxas que não se relacionam com a raça, educação, nível salarial ou estado civil); e finalmente, esta perturbação está associada a uma elevada taxa de mortalidade, sendo que a taxa de suicídio pode ir até aos 15% dos casos. A este respeito Braconnier (2007) refere que entre um terço e metade dos suicídios e tentativas de suicídio (com consequências médicas graves) podem ser atribuídas a esta perturbação, sendo que “o seu risco de morte por suicídio é trinta vezes superior ao da população geral” (p.160).

Mais recentemente, numa revisão de literatura acerca da epidemiologia da depressão, Kessler e Wang (2009) revelaram dados alarmantes. Estima-se que 16,6% da população adulta já teve (ou terá), em algum momento da sua vida, um episódio de depressão (Kessler, Berglund, et al., 2005a, in Kessler & Wang, 2009), o que sugere que mais de 30 milhões de adultos americanos já tiveram esta perturbação (Haro, Arbabzadeh-Bouchez, et al., 2006, in Kessler & Wang, 2009). Para além disto, Kessler e Wang (2009) concluem que entre metade a dois terços da população que nunca teve um episódio de depressão, o terá em algum momento da sua vida. Outro dado preocupante é que mais de 80% da população que já teve um episódio de depressão tem episódios recorrentes (Kessler, Berglund, et al., 2003, in Kessler & Wang, 2009).

As consequências desta perturbação são devastadoras e têm impacto nas mais diversas áreas da vida do indivíduo. Segundo Kessler e Wang (2009) os prejuízos causados pela depressão (seja a nível de resultados escolares, das relações familiares ou da produtividade no trabalho) são comparáveis aos causados por doenças físicas graves. Por exemplo, a depressão é a segunda doença (a primeira são as dores crónicas de costas e de pescoço) a causar maior absentismo no trabalho – calculado em 386,6 milhões de dias por ano (Merikangas, Ames, et al., 2007, in Kessler & Wang, 2009). Este impacto a nível profissional não se repercute unicamente ao nível do absentismo, mas também ao da produtividade/performance dos

indivíduos – a junção dos dois factores representa um prejuízo que ultrapassa os 36 biliões de dólares por ano (Kessler, Akiskal, et al., 2006, in Kessler & Wang, 2009), prejuízo que para os autores está subestimado, pois não inclui outras repercussões possíveis que a depressão pode ter nos locais de trabalho (como acidentes, impacto nos colegas, etc.).

Finalmente, outro dado preocupante é que apenas um terço da população que necessita de tratamento psicológico e/ou psiquiátrico o faz no mesmo ano em que se inicia o episódio de depressão (Kessler, Olfson, et al., 1998; Olfson, Kessler, et al., 1998; Wang, Berglund, et al., 2005; in Kessler & Wang, 2009). No entanto, sabe-se que a maioria dos indivíduos que sofre de depressão (não tratada/ainda não diagnosticada), recorre com frequência aos cuidados de saúde primários com queixas físicas vagas (Kessler & Wang, 2009), pelo que se torna extremamente importante a detecção precoce destes casos, não só para um tratamento eficaz, mas também para reduzir os custos que esta situação acarreta para os serviços de saúde.

É por tudo o que foi referido que se considera que as perturbações depressivas estão entre as mais comuns e as mais debilitantes e que têm sido avaliadas como uma das mais dispendiosas a nível mundial (Murray & Lopez, 1996, in Gotlib & Hammen, 2009), tendo como referência a taxa de mortalidade, o grau de incapacidade e os custos económicos que lhe estão associados (Murray & Lopez, 1996, in Hammen & Mazure, 2003).

Desta forma, é importante não só saber identificar a depressão, mas também avaliar a sua gravidade (que no fundo não é mais do que quantificá-la), pois a sua gravidade relaciona-se directamente “com o grau de incapacidade que o indivíduo sente para a vida em geral ou, especificamente, no seu trabalho e ambientes social e familiar” (Vaz-Serra, 1994, p.15). Duma maneira geral, a forma mais habitual de avaliação das perturbações mentais, e da depressão, é através da recolha da história clínica do indivíduo (entrevista clínica), que segundo Vaz-Serra (1994, p.15-16) é um processo repleto de riscos, na medida em que

A pesquisa psicopatológica está dependente da preparação teórica e do treino do clínico, da boa ou má colaboração do doente e ainda dos sintomas que são mais valorizados em dada época ou por determinada teoria. Ao analisarmos uma história clínica podemos apreciar o que lá está mas ficamos *sem saber o que lá não está*, isto é, se o observador pesquisou os sintomas de uma forma imparcial ou se ligou apenas à presença de alguns. Por outro lado uma história clínica não nos ajuda a *medir a depressão*, isto é, não nos dá indicadores objectivos que nos revelem a gravidade do doente e a possibilidade de comparação com outros enfermos.

É ainda de referir que a história clínica do indivíduo pode também conter dados que resultem de informações dadas por pessoas significativas para o indivíduo, como os familiares e/ou amigos. No entanto, as limitações referidas anteriormente também se aplicam a estas situações. Assim sendo, se se quiser ter uma maior precisão na avaliação clínica torna-se necessário recorrer a outro método de avaliação. É neste sentido que as escalas de avaliação das perturbações mentais, nomeadamente as de avaliação da depressão, se mostram úteis. Como refere Vaz-Serra (1994, p.16),

Em termos de precisão, uma escala para medir a intensidade de um quadro depressivo tem a vantagem de permitir um registo uniforme e padronizado do doente. Um grupo determinado de sintomas *está sempre lá* e os que lá faltam provavelmente têm uma razão explicativa lógica, de natureza estatística, *para lá não estarem*. Os sintomas são assinalados independentemente de quaisquer pressupostos teóricos interpretativos. Se for passada em períodos regulares de tempo fica-se a conhecer o que vai acontecendo à gravidade da depressão e a forma como os sintomas vão evoluindo. Deste modo pode servir, por exemplo, para avaliar a eficácia de determinada terapêutica. Além disso torna-se um meio fácil de identificar casos de depressão em populações de risco ou, simplesmente, de efectuar estudos de natureza epidemiológica. A presença de certos sintomas realçados pela escala pode igualmente constituir o ponto de partida para uma intervenção psicoterapêutica.

No entanto, é importante ter presente que mesmo o uso de instrumentos psicológicos, como as escalas de avaliação da depressão, tem limitações inerentes, nomeadamente a possibilidade de os indivíduos darem respostas enviesadas, respostas essas que podem ser de vários tipos: A tendência de resposta, que reflecte uma propensão do indivíduo para concordar ou discordar com todas as questões colocadas; a tendência central, na qual o indivíduo mostra relutância em dar respostas extremadas; e a tendência para dar respostas socialmente desejáveis, de forma a transparecer uma boa imagem de si (Vaz-Serra, 1994). Apesar desta limitação, não deixa de parecer que o uso de escalas de avaliação da depressão se constitui uma boa forma de avaliação deste quadro psicopatológico e da sua gravidade.

Psicometria e Depressão

Antes de se fazer referência a alguns instrumentos psicológicos de avaliação da depressão, é de referir que estes instrumentos se apresentam normalmente como escalas, que são de dois tipos: De auto-avaliação, nas quais é o próprio indivíduo que responde, que se avalia; e de hetero-avaliação, que são escalas que são preenchidas pelo próprio clínico que avalia o indivíduo (Vaz-Serra, 1994). Ambas apresentam vantagens e desvantagens.

De entre as vantagens das escalas de auto-avaliação destacam-se as seguintes: Permitem assinalar como o indivíduo se vê e como se sente (que muitas vezes pode não coincidir com o que transparece ao clínico), e não como é visto por estranhos (evitando pressupostos que o clínico possa ter); poupam tempo ao clínico, pois é o indivíduo que responde e as escalas podem ser preenchidas em qualquer altura; podem ser preenchidas em casa e devolvidas via correio [metodologia criticada por muitos autores, nomeadamente Cronbach (1996)], o que possibilita a avaliação da evolução da depressão e ultrapassar problemas de distância; e, finalmente, permitem a investigação em grupo, pois as escalas podem ser respondidas por muitos indivíduos ao mesmo tempo (Vaz-Serra, 1994). As principais desvantagens destas escalas prendem-se com o facto de não poderem ser preenchidas por indivíduos com pouca literacia, demasiado perturbados (por exemplo, com demarcada agitação ou lentidão psicomotora) ou com manifestações de depressão psicótica (Vaz-Serra, 1994). Estas escalas exigem ainda que as questões não sejam ambíguas e que sejam de fácil compreensão para todos e quaisquer indivíduos. Além disso, estão sujeitas a eventuais enviesamentos nas respostas dos indivíduos (que poderão não ser sinceros e fazer transparecer algo que não existe ou não é verdadeiro).

As escalas de hetero-avaliação ultrapassam as desvantagens referidas para as de auto-avaliação e, além disso, “a fiabilidade da sua aplicação aumenta com o treino e os conhecimentos do observador” (Vaz-Serra, 1994, p.26). No entanto, também apresentam desvantagens, como o facto do clínico ter de estar totalmente isento de pressupostos em relação ao indivíduo quando faz a observação, e saber o significado exacto de cada sintoma e

das pontuações associadas a ele para que os seus registos sejam de confiança (Vaz-Serra, 1994).

Instrumentos Existentes

Numa revisão de literatura acerca da história da avaliação da depressão, Santor, Gregus e Welch (2006, in Nezu et al., 2009) constataram que desde 1918 foram desenvolvidos mais de 280 instrumentos de avaliação da severidade da depressão. Descrevem-se de seguida os mais representativos e significativos a nível internacional e nacional, seja pela sua ampla utilização ou pelo seu recente aparecimento na literatura.

De entre os primeiros instrumentos de avaliação da depressão está a sub-escala de depressão do Minnesota Multiphasic Personality Inventory (MMPI) de Hathaway e McKinley (1943, in Vaz-Serra, 1994), que ao longo dos tempos revelou não ser uma boa medida da depressão por estar mais relacionada com aspectos da personalidade do que propriamente com os estados depressivos (Snaith et al., 1971, in Vaz-Serra, 1994). A tendência é que deixe de ser usada, pois, segundo Vaz-Serra (1994), cada vez mais se prefere usar escalas que avaliem apenas uma perturbação psicológica e não várias entidades clínicas.

Uma das primeiras escalas de auto-avaliação da depressão foi o Beck Depression Inventory (BDI) de Beck, Ward, Mendelson, Mock e Erbaugh (1961, in Vaz-Serra, 1994), constituído por itens referentes a aspectos cognitivos, somáticos, afectivos, interpessoais e de comportamentos observáveis, com o intuito de avaliar toda a sintomatologia depressiva. Alguns anos após a sua construção, Rehm (1989, in Vaz-Serra, 1994) evidenciou que esta escala apresentava correlações positivas, moderadas a elevadas, com instrumentos de avaliação da ansiedade, o que levou a equacionar se a escala avaliaria tanto depressão como ansiedade ou se apenas reflectia o facto, clinicamente comprovado, de a depressão ser frequentemente acompanhada de ansiedade. Esta versão da escala encontra-se aferida e adaptada para a população portuguesa (Vaz-Serra & Abreu, 1973, in Vaz-Serra, 1994).

Recentemente, o BDI sofreu uma série de modificações que resultaram no BDI-II (Beck et al., 1996), que, como referem Nezu et al. (2009), é um dos instrumentos de auto-avaliação da depressão mais usados a nível mundial, tanto em população clínica como não clínica, e que será abordado em pormenor aquando dos instrumentos usados. É no entanto de referir que o BDI-II tem demonstrado fortes propriedades psicométricas e tem sido adaptado para vários contextos e populações (Nezu et al., 2009).

Em 1965 apareceu a Escala de Auto-Avaliação da Depressão de Zung, outro dos instrumentos de auto-avaliação da depressão muito popular (Nezu et al., 2009), constituída por itens que abrangiam os mesmos aspectos que o BDI (Vaz-Serra, 1994), com o objectivo de avaliar, de forma rápida, sintomas depressivos ao nível da cognição, dos afectos e do comportamento (Nezu et al., 2009). Apesar de ter uma boa capacidade discriminativa entre deprimidos e não deprimidos, foram-lhe atribuídas algumas limitações, como o facto de se correlacionar significativamente com instrumentos de avaliação da ansiedade (Rehm, 1989, in Vaz-Serra, 1994). As suas propriedades psicométricas têm sido avaliadas em inúmeras populações, e têm sido efectuadas análises factoriais que demonstram que a escala reflecte três factores; contudo, estes factores têm sido interpretados de forma diferente (Sakamoto, Kijima, Tomoda & Kambara, 1998; Kivela & Pakkala, 1987; in Nezu et al., 2009).

Em 1971 foi criado o Inventário Wakefield de Depressão de Snaith, Ahmed, Mehta e Hamilton, que não é mais do que uma versão reduzida e modificada da escala de Zung (Vaz-Serra, 1994), e foi criado o Profile of Moods States de McNair, Lorr e Droppleman, que, assim como o BDI, também se encontra aferido e validado para a população portuguesa (Azevedo, Silva & Dias, 1991, in Vaz-Serra, 1994).

A Escala de Depressão do Centro de Estudos Epidemiológicos (CES-D) foi criada, em 1977, com o objectivo de avaliar o nível de sintomatologia depressiva actual, com ênfase na componente afectiva (Gonçalves & Fagulha, 2006). Foi construída com base noutros instrumentos já validados, como o BDI, de maneira a estarem representados os principais sintomas da depressão (Nezu et al., 2009). A CES-D tem sido amplamente utilizada em diversos contextos e as suas propriedades psicométricas têm sido provadas (Gonçalves & Fagulha, 2006; Nezu et al., 2009). Esta escala foi recentemente validada para a população portuguesa (Gonçalves & Fagulha, 2004, in Gonçalves & Fagulha, 2006).

Já na década de 80, Derogatis (1982, in Canavarro, 2007) criou o Brief Symptom Inventory (BSI), que avalia sintomas psicopatológicos, que se agrupam em nove dimensões sintomatológicas, entre as quais se encontra a dimensão depressão, e em três índices globais. A análise das pontuações das várias dimensões indica o tipo de sintomatologia que perturba o indivíduo, enquanto que a dos índices indica o nível de sintomatologia psicopatológica que o indivíduo apresenta. Segundo Canavarro (2007), o BSI é um bom indicador de sintomas psicopatológicos e um bom discriminador entre indivíduos emocionalmente perturbados e não perturbados, e a sua consistência interna e estrutura factorial são indicadores de que é um instrumento conceptualmente sólido. Este inventário encontra-se adaptado para a população portuguesa (Canavarro, 1999, in Canavarro, 2007).

Em 1983 foi criada a Escala de Ansiedade e Depressão Hospitalar (HADS), com o objectivo de identificar perturbações depressivas e ansiosas em populações clínicas não psiquiátricas (Zigmond & Snaith, 1983, in Nezu et al., 2009). Apesar de incluir o termo “hospitalar” no seu nome, vários estudos têm demonstrado que a escala é válida em contextos comunitário e de cuidados de saúde primários, para além de se ter mostrado útil em contexto clínico psiquiátrico e psicológico (Pais-Ribeiro, Silva, Ferreira, Martins, Meneses & Baltar, 2007). Bjelland, Dahl, Haug e Neckelmann (2002, in Nezu et al., 2009) numa revisão de literatura sobre 747 estudos realizados com esta escala desde 1997, demonstraram que esta detinha boas qualidades psicométricas e concluíram que identificava eficazmente casos de depressão em diversas populações clínicas e não clínicas. Esta escala continua a ser uma das mais usadas e mais úteis em diferentes contextos e condições, em diversos grupos etários e em diferentes países, línguas e culturas (Pais-Ribeiro et al., 2007). É de referir que foi muito recentemente validada para a população portuguesa (Pais-Ribeiro et al., 2007).

A partir de meados dos anos 90, surgiu uma série de instrumentos de avaliação da depressão, de entre os quais se destacam os seguintes:

- A Cardiac Depression Scale (Hare & Davis, 1996, in Nezu et al., 2009), desenvolvida para avaliar o nível de depressão em doentes cardíacos, nos quais se regista uma elevada prevalência e incidência de perturbações depressivas;
- O Inventory of Depressive Symptomatology que foi inicialmente construído com o intuito de superar os instrumentos existentes para avaliar a depressão (Rush, Giles, Schlessler, Fulton, Weissenburger & Burns, 1986, in Nezu et al., 2009), e que foi

revisto uma década depois com o objectivo de reflectir os critérios de diagnóstico do Diagnostic and Statitiscal Manual of Mental Disorders – Fourth Edition (DSM-IV; Rush, Gullion, Basco, Jarret & Trivedi, 1996, in Nezu et al., 2009), sendo uma medida fiável da severidade dos sintomas e da alteração dos mesmos;

- O Harvard Department of Psychiatry/National Depression Screening Day é um questionário de auto-avaliação criado para ser usado especificamente naquele instituto dos EUA, que pretende ser um guia para a referência aos serviços de saúde mental (Baer et al., 2000, in Nezu et al., 2009), e que tem sido usado na aferição da depressão em populações clínicas e psiquiátricas (Nezu et al., 2009);
- A Patient Health Questionnaire Depression Scale (Kroenke, Sptitzer & Williams, 2001, in Nezu et al., 2009) é uma escala destinada a avaliar a severidade da depressão, constituída por nove itens que derivam directamente dos nove sintomas de depressão major descritos no DSM-IV;
- A Depression in the Medically III-18 e a Depression in the Medically III-20 são escalas, com ênfase na componente cognitiva, especificamente desenvolvidas para avaliar os níveis de depressão em populações clínicas doentes.

As qualidades psicométricas de todos estes instrumentos têm sido demonstradas. No entanto, duma maneira geral, pode-se considerar que são necessários mais estudos, devido ao seu recente aparecimento. É de mencionar que não se tem conhecimento de que algum destes instrumentos tenha sido validado e/ou adaptado para a população portuguesa.

Até agora foram referidos instrumentos de auto-avaliação da depressão, isto é, nos quais é o próprio indivíduo que se avalia e que responde aos instrumentos, pelo que se passa de imediato a abordar alguns instrumentos de hetero-avaliação da depressão.

Hamilton foi um dos pioneiros na construção de escalas de hetero-avaliação, com a sua escala de depressão (Vaz-Serra, 1994). A Hamilton Rating Scale for Depression (Hamilton, 1960, in Nezu et al., 2009) foi originalmente construída para avaliar a severidade dos sintomas depressivos em pacientes previamente diagnosticados com uma perturbação depressiva, e é historicamente o instrumento de hetero-avaliação mais usado e muitas vezes visto como “golden standard”. Apesar da popularidade, a sua fiabilidade e validade têm sido postas em causa (Bagby, Ryder, Schuller & Marshall, 2004, in Nezu et al., 2009). Além disso, Bagby et al. (2004, in Nezu et al., 2009) referem que, como a escala foi construída há

algumas décadas, muitos dos sintomas que hoje são valorizados na avaliação da depressão não estão presentes neste instrumento, concluindo que a escala está conceptual e psicometricamente desajustada, pelo que deveria ser considerada uma revisão substancial ou o completo abandono da mesma.

Existem ainda outras escalas de hetero-avaliação, como a Escala de Depressão de Montgomery e Asberg, a Comprehensive Psychopathological Rating Scale, a Escala de Melancolia e a Escala de Depressão de Cronholm-Ottosson (Vaz-Serra, 1994).

Poderia-se continuar com esta lista, mas já está evidente a existência do grande número de instrumentos psicológicos que avaliam a depressão, o que dá conta da importância que tem sido dada a este quadro psicopatológico nos últimos tempos. No entanto, a nível nacional, o número de instrumentos psicológicos disponíveis para avaliar a depressão é mais limitado. Existe um instrumento português de auto-avaliação da depressão, o Inventário de Avaliação Clínica da Depressão (IACLIDE) criado por Vaz-Serra (1994), que se propõe medir a intensidade dos quadros clínicos depressivos, através de itens referentes a perturbações biológicas, cognitivas, interpessoais e de desempenho de tarefa, e que também será abordado em pormenor mais adiante.

Qualidades Psicométricas do BDI-II

Com tantos instrumentos que existem para avaliar a depressão, para quê estudar as características/qualidades psicométricas do BDI-II? Primeiro, como se pôde constatar, só existe um instrumento português válido, criado especificamente para a população portuguesa (o IACLIDE), o que à partida é um pouco limitador – se se quisesse usar um instrumento verdadeiramente adequado à população portuguesa não se teria outra opção se não usar o IACLIDE. A não ser que, como é óbvio, se tivessem disponíveis outros instrumentos, não portugueses, devidamente validados e/ou adaptados para a população portuguesa – o que remete para a segunda justificação. Ora,

Um dos problemas centrais que se nos deparam em Portugal é o da ausência de instrumentos devidamente adaptados à nossa população e aos seus sub-grupos específicos. Assim, defrontamo-nos, na prática e na

investigação, com uma clara dificuldade em avaliar construtos que aparecem referidos na literatura científica e que são essenciais para o trabalho do psicólogo. Esta dificuldade de ordem prática “estimula”, por vezes, utilizações claramente abusivas dos instrumentos de avaliação. Muitas vezes, os psicólogos utilizam uma estratégia preocupante para lidar com esta ausência de adaptações nacionais: Usam critérios provenientes de outros países (Gonçalves et al., 2006, p.14).

De facto, sabendo que existem em todo o mundo mais de 280 instrumentos de avaliação da depressão, não existem muitos instrumentos estrangeiros validados e/ou adaptados para a população portuguesa – só estão validados o BDI (a sua primeira versão), o Profile of Moods States, a CES-D, o BSI (que não mede exclusivamente depressão) e a HADS, facto ainda mais limitador, na medida em que se trata de instrumentos que podem, de certa forma, estar já ultrapassados. O mais recente destes instrumentos é a HADS, que, ainda assim, já tem 25 anos de existência; portanto, é de supor que muito provavelmente existirão outros instrumentos, mais actuais, que sejam tão ou mais válidos do que estes para avaliar a depressão.

Como consequência, a tendência é a de que, tal como referido por Gonçalves et al. (2006), se usem instrumentos que não estão devidamente validados para a nossa população, como é o caso do BDI-II, assumindo que os critérios oriundos de outros países se adequam à população em estudo, o que pode, em muitos casos, ser um erro. Para além do que, como Ribeiro (2008) refere, a ausência ou falta de instrumentos de avaliação adaptados para a nossa população não garante o rigor métrico das avaliações que são feitas no nosso país com instrumentos por nós criados, pois não temos forma de saber se chegaríamos ao mesmo resultado com outros instrumentos.

Para além disto, segundo Streiner e Norman (1989, in Vaz-Serra, 1994), o aparecimento de um instrumento novo justifica-se quando: Se está perante um novo construto para o qual não existe nenhum instrumento para o medir; se deseja substituir um instrumento já existente por um que ofereça vantagens significativas (que seja menos extenso, mais barato e menos invasivo); ou quando não se está satisfeito com os instrumentos existentes e se acredita que de alguma forma omitiram algum aspecto chave do construto.

Assim, para além de ser nossa convicção que o BDI-II se constitui um instrumento adequado para avaliar a depressão e que se pode constituir como uma boa alternativa ao

IACLIDE, pois apresenta vantagens significativas em termos de brevidade e custos, sendo também menos invasivo; o BDI-II pretende reflectir os critérios de diagnóstico do DSM-IV, enquanto que o IACLIDE pretende reflectir os critérios de diagnóstico do Diagnostic and Statitiscal Manual of Mental Disorders – Third Edition – Revised (DSM-III-R), ou seja, pode-se de certa forma considerar que se encontra um tanto ou quanto desactualizado.

O presente estudo propõe-se então a analisar as características psicométricas daquele que é considerado por muitos um dos instrumentos de avaliação da depressão mais usados em todo o mundo, inclusivamente no nosso país (ainda que sem estudos de validação e/ou adaptação para a população portuguesa), o BDI-II, com o intuito de dar um contributo para o processo de validação e/ou adaptação do instrumento para a população portuguesa. Para além de ser importante andar-se a par e passo com os desenvolvimentos a nível internacional, e para tal é necessário que se utilizem medidas universais (instrumentos amplamente conhecidos e utilizados no estrangeiro), crê-se também que este instrumento poderá ser mais profícuo que outros, nomeadamente o IACLIDE, em termos de características psicométricas e outras.

É de referir que este tipo de estudo é muito semelhante aos estudos de validação e/ou adaptação de instrumentos psicológicos. Assim, sendo que para que a adaptação dum instrumento seja adequada é necessário proceder-se de forma muito semelhante ao processo de construção inicial do mesmo, de tal forma que a tradução deverá ser submetida aos mesmos procedimentos psicométricos pelos quais a versão original do instrumento passou (Bradley, 1994, in Ribeiro, 2008), é isto que se pretende fazer no presente estudo. Propomos a fazê-lo para uma amostra de estudantes universitários, esperando que os nossos dados possam servir de referência para investigações que incidam em populações semelhantes à aqui examinada.

Desta forma, os dados colectados vão ser analisados pormenorizadamente com o intuito de verificar/comprovar as qualidades psicométricas do BDI-II e dos seus itens, no que diz respeito à sua fiabilidade (consistência interna), validade de critério (concorrente), validade de construto (convergente, discriminante e factorial) e à sensibilidade dos itens.

MÉTODO

Amostra

Segundo Tinsley e Tinsley (1987, in Ribeiro, 2008), uma amostra é considerada pobre quando é constituída por 100 indivíduos, razoável quando tem 200 indivíduos, boa quando constituída por 300 indivíduos, muito boa se tiver 500 indivíduos e excelente se tiver 1000 indivíduos. A este respeito Ribeiro (2008) recomenda que uma amostra tenha mais de 300 indivíduos. Desta forma, pode-se considerar que a amostra do presente estudo é muito boa, na medida em que é constituída por 670 indivíduos.

Da amostra total, 572 indivíduos eram do género feminino (85,4%) e 98 do género masculino (14,6%)⁶. Pressupõe-se que os indivíduos, alunos do primeiro, segundo, terceiro e quarto anos de Psicologia (das três vertentes do curso: Clínica, Educacional e Social), de Reabilitação e Inserção Social (RIS) e de Desenvolvimento Comunitário e Saúde Mental (DCSM) do ISPA, provêm dum estrato sócio-económico médio e médio-alto. Os indivíduos encontram-se na faixa etária dos 18 aos 48 anos, sendo que a média de idades é de 22,38 com um desvio-padrão de 4,40.

⁶ Toda a estatística descritiva relativa à amostra encontra-se em anexo (ver Anexo D).

Tal como referido, os indivíduos eram alunos do ISPA que frequentavam os três cursos da instituição. Assim, do curso de RIS participaram 41 indivíduos, do curso de DCSM participaram 23 indivíduos, e do curso de Psicologia houve 606 participantes. Da amostra total, sem diferenciação dos cursos, 125 indivíduos frequentavam o primeiro ano, 168 o segundo ano, 200 o terceiro ano e 177 frequentavam o quarto ano.

Em relação ao curso de RIS, dos 41 indivíduos 39 eram do género feminino e 2 eram do género masculino, que tinham idades compreendidas entre os 19 e os 27 anos ($x=21,98$ e $s=1,89$), e dos quais 9 eram do primeiro ano, 10 do segundo, 6 do terceiro e 16 do quarto ano. Do curso de DCSM participaram 23 indivíduos, dos quais 19 eram do género feminino e 4 do género masculino, com idades compreendidas entre os 18 e os 28 anos ($x=21,48$ e $s=2,21$), e em que 6 eram do primeiro ano, 7 do segundo, 7 do terceiro e 3 eram do quarto ano. Relativamente ao curso de Psicologia, participaram 606 indivíduos, sendo 514 do género feminino e 92 do género masculino, que se encontravam na faixa etária dos 18 aos 48 anos ($x=22,45$ e $s=4,57$), e dos quais 110 frequentavam o primeiro ano, 151 o segundo, 187 o terceiro e 158 frequentavam o quarto ano. Destes últimos, isto é, dos 158 indivíduos que frequentavam o quarto ano de Psicologia, 100 eram da área de Clínica, 38 da de Educacional e 20 da área de Social.

É de referir que a recolha de dados foi feita em todas as turmas de cada um dos anos referidos e para cada um dos cursos. Desta forma, para o curso de Psicologia, em cada um dos anos, os indivíduos poder-se-ão dividir por turnos, isto é, manhã, tarde e noite, havendo, desta forma, 338 indivíduos que frequentavam o turno da manhã, 144 o da tarde e 124 o da noite. No entanto, é de salientar que os indivíduos do quarto ano da área de Educacional e de Social só se podem dividir por dois turnos: O da manhã e o da noite. Em relação aos outros dois cursos, RIS e DCSM, não poderá ser feita esta divisão, na medida em que para cada ano existe unicamente uma turma.

É importante salientar que dos 670 indivíduos, 560 foram convidados a preencher um questionário constituído por três inventários de avaliação da depressão e da ansiedade – o BDI-II (Beck et al., 1996) e o IACLIDE (Vaz-Serra, 1994), que avaliam a severidade de depressão; e o State-Trait Anxiety Inventory – Form Y (STAI; Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983), que avalia o nível de ansiedade; e 110 foram convidados a

preencher um questionário constituído unicamente por dois dos inventários (o BDI-II e o STAI) – os alunos do primeiro ano de Psicologia.

Tal situação deveu-se ao facto dos alunos do primeiro ano de Psicologia terem sido alvo dum trabalho de investigação, anterior ao actual, no qual preencheram um questionário constituído pelo BDI-II e pelo STAI. Neste trabalho, os indivíduos estiveram envolvidos em três situações de recolha de dados. Na medida em que era possível aproveitar esses dados, pois foram recolhidos nas mesmas condições experimentais, optou-se por não se fazer uma nova recolha de dados com estes indivíduos, para evitar eventuais recusas e enviesamentos dos dados (por aprendizagem da situação), apesar de se ter perdido os dados relativos ao IACLIDE. No entanto, sendo a amostra razoavelmente grande, este facto parece não colocar em causa os resultados do presente estudo.

Em relação aos dados resultantes dos três inventários, os valores médios do score total de cada um deles, bem como o seu desvio-padrão, podem ser resumidos na Tabela 1, que se apresenta de seguida:

Tabela 1 – Scores totais médios dos três inventários em função dos cursos

		BDI-II	IACLIDE	STAI
RIS	<i>N</i>	41	41	41
	<i>x</i>	9,24	10,85	37,42
	<i>s</i>	7,12	9,15	9,10
DCSM	<i>N</i>	23	23	23
	<i>x</i>	11,57	12,78	38,83
	<i>s</i>	5,80	8,37	8,11
PSICOLOGIA	<i>N</i>	606	495	606
	<i>x</i>	9,41	12,04	38,31
	<i>s</i>	7,69	11,75	11,07
TOTAL	<i>N</i>	670	559	670
	<i>x</i>	9,47	11,98	38,28
	<i>s</i>	7,60	11,45	10,86

Perante estes dados é possível constatar que apesar de não existirem diferenças significativas entre os alunos dos três cursos, os alunos que frequentam o curso de DCSM

tenderam a obter resultados mais elevados do que os restantes em qualquer um dos instrumentos de avaliação da depressão e da ansiedade. No entanto, os alunos de Psicologia tenderam a estar mais próximos dos alunos de DCSM nos resultados do IACLIDE e do STAI. Os alunos de RIS foram aqueles que obtiveram resultados mais baixos em qualquer um dos instrumentos.

Instrumentos

Para o presente estudo das características psicométricas do BDI-II foi utilizado, tal como referido, um instrumento constituído por três inventários (ver Anexo E). Um dos inventários era o BDI-II (instrumento em estudo); os outros eram o IACLIDE, que foi utilizado com o intuito de analisar a validade concorrente e convergente do BDI-II (ambos têm o mesmo construto teórico de base – depressão), e o STAI, que foi utilizado com o intuito de analisar a validade discriminante do BDI-II (avaliam construtos teoricamente diferentes – depressão e ansiedade, mas que aparecem repetidamente como estando altamente correlacionados).

É de referir que o instrumento foi estruturado para que os resultados do BDI-II não fossem influenciados pelos outros inventários; desta forma, estando dividido em três partes, apresentava a seguinte estrutura: Numa primeira página, onde se recolhiam dados sócio-demográficos como o género e a idade, fazia-se referência ao objectivo do estudo, bem como a questões relacionadas com a voluntariedade da participação e confidencialidade dos dados; de seguida, a primeira parte do questionário era, tal como referido, constituída pelo BDI-II, a segunda pelo STAI e a terceira pelo IACLIDE.

O facto de o STAI ter aparecido entre os dois inventários que avaliam o mesmo construto teórico, BDI-II e IACLIDE, justifica-se com a tentativa de que os dados do BDI-II não influenciassem os do IACLIDE, por serem em tudo muito semelhantes. No entanto, é difícil garantir que ainda assim não tenham havido implicações tanto no STAI como no

IACLIDE originadas pelo BDI-II. De resto, a única coisa que se conseguiu garantir, e a mais importante para o presente estudo, é que os resultados do BDI-II não fossem influenciados pelos resultados obtidos nos outros inventários. A este respeito é importante referir que em qualquer estudo das características psicométricas de instrumentos ou de validação e/ou adaptação dos mesmos, em que seja efectuada a validação concorrente, convergente e discriminante, a passagem do instrumento em estudo é habitualmente, se não mesmo sempre, feita no início, seguida da passagem dos restantes instrumentos.

BDI-II

O BDI-II é um instrumento de auto-descrição, constituído por 21 itens, que permite avaliar a severidade da depressão em adultos e adolescentes a partir dos treze anos de idade. Esta versão do inventário foi desenvolvida para aferir sintomas correspondentes aos critérios de diagnóstico das perturbações depressivas descritos no DSM-IV (Beck et al., 1996).

É importante referir que o BDI-II foi construído numa base de 35 anos de acumulação de dados psicométricos e experiência clínica com o BDI (Beck, et al., 1961, in Beck et al., 1996) e com o Amended Beck Depression Inventory (BDI-IA; Beck, et al., 1979, in Beck et al., 1996), que são os antecessores da versão utilizada neste estudo. Além disso, vários estudos (Archer, Maruish, Imhof & Piotrowski, 1991; Piotrowski & Keller, 1992; Piotrowski, Sherry & Keller, 1985; in Beck et al., 1996) referem que, durante os últimos 35 anos, o BDI se tornou um dos instrumentos mais aceites para aferir a severidade da depressão em pacientes diagnosticados, bem como para detectar a possibilidade de depressão em populações normais.

Desta forma, convém salientar, para melhor compreender, de que forma evoluiu o BDI-II, que, ao contrário do BDI-IA, se apresenta como uma revisão substancial do BDI (Beck et al., 1996).

O BDI baseou-se em observações clínicas e descrições de pacientes psiquiátricos com depressão, que foram sendo sistematicamente consolidadas em 21 itens representativos de

sintomas e atitudes, não tendo sido seleccionados com o intuito de reflectir nenhuma teoria da depressão em particular (Beck et al., 1961, in Beck et al., 1996). Estes 21 itens [humor, pessimismo, sentido de fracasso, insatisfação pessoal, culpa, castigo (ou punição), desprezo por si próprio, auto-acusação, ideação suicida, choro, irritabilidade, isolamento social, indecisão, alteração da imagem corporal, dificuldade no trabalho, insónia, fadigabilidade, perda de apetite, perda de peso, preocupação somática e perda de libido] foram organizados consoante a severidade do conteúdo das afirmações, e cada item foi cotado numa escala de quatro pontos, que ia do zero ao três em termos de severidade (Beck et al., 1996).

Passados dez anos da construção do BDI, em 1971, Beck et al. (1979, in Beck et al., 1996) começaram a utilizar uma versão modificada do instrumento que eliminava palavras alternativas para os mesmos sintomas e evitava as “duplas-negativas”. Além disso, nesta nova versão, que só foi publicada em 1979, o número das alternativas de resposta foi estritamente limitado a quatro para cada item, incluindo a hipótese nula, e foram feitas reformulações mínimas em 15 dos itens, sendo que 6 deles não foram modificados.

A necessidade de se reformular novamente o instrumento começou a surgir no início dos anos 80 com a evolução que o conceito de depressão estava a sofrer, especialmente com a edição do Diagnostic and Statistical Manual of Mental Disorders – Third Edition (DSM-III; 1980, in Beck et al., 1996) e com o estudo efectuado por Moran e Lambert (1983, in Beck et al., 1996), no qual se concluiu que, dos nove critérios de diagnóstico do DSM-III para perturbações afectivas, seis estavam incluídos no BDI, dois estavam parcialmente incluídos e um não estava de todo incluído. Os dois critérios de diagnóstico parcialmente incluídos eram os referentes à perda de apetite e de sono, que se referiam unicamente a diminuições e não admitiam aumentos. Em relação ao critério que não estava de todo incluído, o BDI não contava com nenhum item referente à actividade psicomotora e à agitação, que se constituíam como critérios de diagnóstico para a depressão major.

Também as edições posteriores da American Psychiatric Association vieram reforçar a necessidade duma nova medida psicológica para a depressão, que aferisse sintomas que fossem consonantes com os critérios de diagnóstico do DSM-III-R e do DSM-IV (1987 e 1994, in Beck et al., 1996). Além disso, sintomas como a alteração da imagem corporal, perda de peso e preocupação somática, que estavam incluídos no BDI, foram-se tornando cada vez mais irrelevantes para a aferição da severidade da depressão.

Assim, em 1994 iniciou-se o estudo piloto para o BDI-II, no qual se incluíram itens referentes aos critérios de diagnóstico do DSM-IV, como a agitação, a desvalorização pessoal, a dificuldade de concentração e a perda de energia, que não estavam presentes no BDI-IA; se reformularam as opções de resposta dos itens já existentes, bem como dos itens referentes à insónia e à perda de apetite de forma a reflectirem simultaneamente diminuições e aumentos; e, finalmente, se excluiu o item referente à perda de peso (Beck et al., 1996). Deste estudo resultou a “versão final” do BDI-II constituída por 21 itens – tristeza, pessimismo, fracasso no passado, perda de prazer, sentimentos de culpa, sentimentos de punição, desprezo por si próprio, autocrítica, pensamentos ou desejos suicidas, choro, agitação, perda de interesse, indecisão, desvalorização pessoal, perda de energia, alterações nos hábitos do sono, irritabilidade, alterações no apetite, dificuldades de concentração, cansaço ou fadiga e perda de interesse sexual.

O BDI-II pode ser aplicado individualmente ou em grupo, oralmente ou não, e, duma maneira geral, requer 5 a 10 minutos para ser preenchido. De forma a haver coerência com os critérios de diagnóstico para a depressão major do DSM-IV, a instrução que é dada aos indivíduos é a de que vão responder a um questionário composto por grupos de frases que descrevem comportamentos e sentimentos, e que deverão escolher, em cada grupo, a frase que melhor descreve o que têm sentido nas últimas duas semanas, incluindo o próprio dia. A cotação do BDI-II é feita somando as cotações dos 21 grupos de frases, em que cada um deles é cotado numa escala de quatro pontos que vai do zero ao três (se num grupo de frases for seleccionada mais do que uma frase, cota-se a que tiver cotação mais elevada). Desta forma, o score total máximo do BDI-II é de 63. Através deste score total é possível aferir quatro grupos (pontos de corte) relativos à severidade de depressão em pacientes diagnosticados com depressão major:

Score total entre 0 e 13: Depressão mínima;

Score total entre 14 e 19: Depressão ligeira;

Score total entre 20 e 28: Depressão moderada;

Score total entre 29 e 63: Depressão severa.

Importa salientar que estes grupos se referem ao objectivo de Beck et al. (1996), o de estabelecer pontos de corte para aferir a severidade de depressão em pacientes diagnosticados

com depressão major, e que, como os autores referem, a decisão de se usar diferentes pontos de corte deve basear-se nas características da amostra, bem como no objectivo do uso do BDI-II. Por exemplo, quando o objectivo de utilização do BDI-II é o de detectar o maior número de indivíduos deprimidos, então os pontos de corte devem ser diferentes dos apresentados por Beck et al. (1996). Além disto, é importante referir que o BDI-II foi desenvolvido como um indicador da presença e do grau de sintomas depressivos consistentes com o DSM-IV, e não como um instrumento para especificar um diagnóstico clínico (Beck et al., 1996).

Os dados psicométricos do BDI-II foram obtidos com base numa amostra de 500 pacientes psiquiátricos (população clínica) e numa outra de 120 estudantes universitários (população estudantil). O coeficiente alfa de Cronbach para a população clínica foi de 0,92 e para a população estudantil foi de 0,93. As correlações item-total corrigidas foram significativas em ambas as amostras ($\alpha=0,05$), tendo variado entre 0,39 (perda de interesse sexual) e 0,70 (perda de prazer) na população clínica, e entre 0,27 (perda de interesse sexual) e 0,74 (desprezo por si próprio) na população estudantil. É de referir que estes dados são indicadores da consistência interna do instrumento.

Na análise de quão bem as quatro afirmações ou opções de resposta de cada grupo de sintomas se diferenciam umas das outras, assim como de quão bem o conjunto das quatro afirmações ou opções de resposta de cada grupo de sintomas mede monotonicamente a dimensão depressão auto-descrita, verificou-se que: Todos os itens demonstraram uma relação monotónica crescente com a dimensão depressão; 17 dos 21 itens reflectiram uma cotação ordinal apropriada em relação à discriminação entre indivíduos com depressão mais elevada daqueles com depressão mais baixa; e todas as opções de resposta de 4 itens (sentimentos de punição, pensamentos ou desejos suicidas, agitação e perda de interesse sexual) não apresentaram a cotação ordinal apropriada – por exemplo, a segunda opção de resposta (cotada com um ponto) do item pensamentos e desejos suicidas era mais provável, isto é, tinha maior probabilidade de ser seleccionada pelos pacientes com depressão severa do que a quarta opção de resposta (cotada com três pontos).

Para além dos dados anteriores relativos à fiabilidade do BDI-II, foi também efectuada uma análise da estabilidade teste-reteste da qual resultou uma correlação significativa de 0,93 ($p<0,001$).

No processo de validação do instrumento foram efectuadas várias correlações (todas significativas $p < 0,001$) entre o BDI-II e outros instrumentos psicológicos para estimar a sua validade de construto (validade convergente e discriminante). Para Beck et al. (1996), a correlação entre o BDI-II e o BDI-IA ($r=0,93$) é prova da validade convergente do instrumento, bem como a correlação entre o BDI-II e a Beck Hopelessness Scale ($r=0,68$) e a Scale for Suicide Ideation ($r=0,37$), na medida em que os construtos teóricos subjacentes a estas duas escalas (desespero e ideação suicida) são frequentemente descritos como estando correlacionados com a depressão; a correlação entre o BDI-II e o Beck Anxiety Inventory ($r=0,60$), que poderia evidenciar uma fraca validade discriminante, era de esperar devido ao facto de ansiedade e depressão estarem correlacionadas na avaliação clínica; finalmente, o facto de o BDI-II estar positivamente mais correlacionado com a Revised Hamilton Psychiatric Rating Scale for Depression ($r=0,71$) do que com a Revised Hamilton Anxiety Rating Scale (HARS-R; $r=0,47$), e de a correlação entre estes dois últimos instrumentos ser superior à correlação entre o BDI-II e a HARS-R ($r=0,51$), evidencia, para os autores, uma robusta validade discriminante entre ansiedade e depressão.

São ainda de reter os dados resultantes da análise factorial do instrumento em relação às duas amostras, através da qual foi possível extrair dois factores em cada uma delas. Para a população clínica, o BDI-II reflectiu uma dimensão somático-afectiva da depressão e uma dimensão cognitiva da depressão; para a população estudantil, o BDI-II reflectiu uma dimensão cognitivo-afectiva da depressão e uma dimensão somática da depressão [para mais informações a este respeito ver Beck et al. (1996)].

Em relação aos dados sócio-demográficos, na população clínica não se registaram correlações significativas entre o BDI-II e a etnia e a idade, no entanto, observou-se uma diferença significativa na média no que diz respeito ao género ($x_{\text{♀}}^7=23,61$ e $x_{\text{♂}}^8=20,44$); na população estudantil a idade estava inversamente correlacionada com o BDI-II e observou-se igualmente uma diferença significativa na média no que diz respeito ao género ($x_{\text{♀}}=14,55$ e $x_{\text{♂}}=10,04$).

Finalmente, os dados de Beck et al. (1996) sugerem que o BDI-II tem uma boa capacidade de discriminação de diagnóstico, não só porque a média do score total do BDI-II

⁷ Sempre que aparecer este símbolo leia-se “indivíduos do género feminino”.

⁸ Sempre que aparecer este símbolo leia-se “indivíduos do género masculino”.

da população clínica foi superior à da população estudantil (aproximadamente o dobro), mas também porque se verificou que os pacientes com perturbações de humor obtiveram scores mais elevados do que os pacientes com perturbações de ansiedade, ajustamento e outras, assim como os pacientes com perturbações depressivas mais graves obtiveram scores mais elevados do que os pacientes com perturbações depressivas menos graves.

IACLIDE

O IACLIDE surgiu na década de 90 como um instrumento “alternativo” ao BDI, de origem portuguesa, após vários anos de utilização e de estudos com aquele inventário. A razão que levou à sua criação prende-se fundamentalmente com a insatisfação relativamente aos instrumentos existentes. Como refere Vaz-Serra (1994, p.36-37):

...o processo laborioso da construção de uma escala pode conhecer uma grande variabilidade, consoante o que é mais valorizado pelos respectivos autores, em função dos seus interesses clínicos ou da sua formação teórica. (...). Para além destes pormenores não podemos esquecer que há uma outra variável que igualmente influencia a construção de uma escala: A cultura! Habitualmente um instrumento de medida começa por um número mais elevado de questões do que aquelas com que fica. Os procedimentos estatísticos vão ditando quais os itens que devem ser seleccionados. Este facto significa que uma escala criada num país, com o auxílio dos métodos estatísticos, acaba por ser submetida ao filtro poderoso da *cultura*. Esta molda, de uma forma directa ou indirecta, as atitudes e as condutas “permitidas” perante os acontecimentos e liga-se à génese das emoções. Os respondentes que vivem num determinado contexto cultural assinalam com maior frequência certas questões do que outras.

É por este motivo que o autor salienta que “Uma escala criada num determinado país pode ser excelente e aí muito utilizada para medir Depressão. Mas, pelos motivos aduzidos, se for utilizada noutro país, *apresenta-nos o que tem mas não nos refere se lá fazem falta outras questões*” (Vaz-Serra, 1994, p.37). Desta forma, o IACLIDE foi construído com o objectivo de se criar um instrumento de medida da depressão adequado à população portuguesa, constituído por sintomas que podem não estar presentes noutros instrumentos de medida (contém questões que, de facto, embora se encontrem frequentemente associadas à depressão, não costumam ser referidas noutras escalas, como por exemplo a dependência que o deprimido cria de terceiros, as queixas que debita em excesso, o prejuízo da memória de

curto prazo, a dificuldade de concentração ou a perda da alegria de viver), e que não seja constituído por questões sensíveis a influências discriminadoras.

O IACLIDE é um inventário de auto-avaliação da depressão, criado como uma escala de tipo Likert [neste tipo de escala, “a medida de um conceito é conseguida através da soma de várias questões que, individualmente, podem ser classificadas em diversas classes de resposta, a que corresponde uma pontuação. *São escalas que permitem, ao mesmo tempo, uma avaliação qualitativa e quantitativa*” (Vaz-Serra, 1994, p.19); para além disso, os itens devem ter uma correlação elevada uns com os outros (Ribeiro, 2008)], que permite medir a intensidade dos quadros clínicos depressivos. Foi denominado por “inventário” pois permite obter mais do que uma pontuação global com interesse clínico.

Na sua construção o autor procurou colocar todos os sintomas considerados típicos da depressão, e que foram seleccionados a partir da experiência obtida na observação clínica de pacientes deprimidos, do que foi descrito a respeito da depressão no DSM-III-R, e da revisão de literatura, em especial dos trabalhos de Beck e seus colaboradores e de Rehm (Vaz-Serra, 1994). A este respeito o autor refere que o processo de selecção e construção dos itens para a construção dum instrumento de avaliação da depressão é complexo por dois motivos:

Um deles, porque não existem sintomas que possam ser considerados *patognomónicos* desta entidade, isto é, cuja presença, só por si, abone sobre a existência deste quadro clínico. (...). O outro motivo é que há sintomas bidireccionais, (...) que podem variar tanto num sentido de aumento como de diminuição. (...). Apesar das dificuldades assinaladas, é aceite que a avaliação de uma Depressão pode ser feita tendo em conta uma **constelação de sintomas** que, isoladamente, têm pouco valor representativo mas, em conjunto, revelam a presença deste quadro clínico (Bech, 1992, in Vaz-Serra, 1994, p.41).

O IACLIDE é constituído por 23 questões, correspondentes a quatro tipos distintos de sintomas: Biológicos, cognitivos, interpessoais e de desempenho de tarefa. Estes tipos de sintomas reflectem, respectivamente, a presença de transtornos/perturbações a diferentes níveis da relação que o indivíduo deprimido estabelece: Com o corpo (alterações de funções biológicas fundamentais); consigo próprio como pessoa (alterações do conteúdo do pensamento); com os outros (modificações nas relações interpessoais); e, finalmente, com o trabalho (supressão ou alteração do desempenho de tarefa).

Das 23 questões, 6 referem-se a sintomas biológicos (desassossego interior, prejuízo da memória de fixação, insónia inicial, insónia tardia, acordar fatigado e dificuldade de concentração), 10 referem-se a sintomas cognitivos (desinteresse pelo trabalho, culpabilidade, sentir-se infeliz, auto-depreciação, perda da esperança no futuro, sentimento de fracasso, incapacidade de resolução de problemas, incapacidade de tomar decisões, perda da alegria de viver e ideação suicida), 3 referem-se a sintomas interpessoais (dependência dos outros, débito excessivo de queixas e afastamento social), e finalmente, 4 são referentes a sintomas de desempenho de tarefa (fatigabilidade no desempenho de tarefas, depreciação no desempenho de tarefas, inibição motora e inactividade).

Cada um dos sintomas está dividido em cinco afirmações, que traduzem uma gravidade progressivamente crescente da depressão, das quais existe uma única possibilidade de resposta. À primeira afirmação corresponde, tal como no BDI-II, uma cotação de zero, representativa da inexistência de qualquer transtorno/perturbação, e à última afirmação corresponde uma cotação de quatro, que traduz a gravidade máxima de transtorno/perturbação. O IACLIDE, tal como o BDI-II, baseia-se no pressuposto de que “a gravidade de uma Depressão varia em função do *número de sintomas assinalados* e da *intensidade que atingem*” (Vaz-Serra, 1994, p.31).

O IACLIDE pode ser aplicado individualmente ou em grupo, e, em média, requer 10 a 15 minutos a ser preenchido. A instrução que é dada aos indivíduos é a de que cada um dos sintomas apresentados não se refere à sua maneira de ser habitual, mas sim à forma como se sentem pelo menos desde há duas semanas para cá, na medida em que o IACLIDE se destina a avaliar um estado e não traços de personalidade. O critério da presença de sintomas “desde há pelo menos duas semanas” foi utilizado de forma a reflectir os critérios de diagnóstico do DSM-III-R. A cotação do inventário é feita somando as cotações das 23 questões, sendo que, como as cotações são feitas numa escala de zero a quatro, a cotação máxima é de 92.

Os dados psicométricos do IACLIDE foram obtidos através duma amostra de 340 deprimidos. O coeficiente alfa de Cronbach foi de 0,89, a correlação “par/ímpar” foi de 0,85 e o coeficiente de Spearman-Brown de 0,92, o que traduz uma boa consistência interna do inventário. Foi efectuada uma correlação entre o score total do IACLIDE e o da versão portuguesa do BDI, da qual resultou um coeficiente de validação concorrente de 0,72.

Verificou-se que os itens têm uma boa capacidade discriminativa de grupos extremos, isto é, as várias questões registam valores diferentes quando o score total indica que o caso é “grave” ou “leve”. Também se verificou uma excelente diferenciação entre a nota global e cada um dos sintomas quando a amostra de 340 deprimidos foi comparada com uma de 340 indivíduos normais, isto é, que nunca tiveram qualquer transtorno psicológico.

Através desta comparação foi possível extrair o ponto de corte entre indivíduos normais e deprimidos, que correspondeu ao score de 20 (significa que com um resultado superior a 20 é muito provável estarmos perante indivíduos deprimidos, e abaixo desse valor indivíduos não deprimidos), e concluir que o IACLIDE é um bom diferenciador entre indivíduos normais e deprimidos. Em relação à severidade/gravidade da depressão, tendo em conta os níveis leve, moderada e grave, os valores médios encontrados foram 28, 45 e 58, respectivamente, sendo que é possível considerar que scores entre 20 e 34 correspondem a depressões leves, entre 35 e 52 correspondem a depressões moderadas, e, finalmente, scores superiores a 53 correspondem a depressões graves.

A análise factorial permitiu extrair cinco factores que explicam 54,24% da variância total. No primeiro factor estão representados sintomas ligados à dificuldade do desempenho de tarefa, bem como a queixas biológicas e cognitivas, que correspondem a depressões endógenas; no segundo factor estão maioritariamente representados sintomas cognitivos, o que corresponde basicamente a uma relação perturbada do indivíduo consigo próprio; o terceiro factor é constituído por sintomas cognitivos, correspondentes a uma constelação suicida; o quarto factor constitui-se por sintomas interpessoais e biológicos, correspondendo a uma marcada instabilidade emocional e dependência de terceiros; finalmente, o quinto factor reúne dois sintomas biológicos, os dois referentes aos transtornos do sono.

É ainda de referir que através do IACLIDE, ou mais correctamente, através de alguns itens (sintomas) do inventário, é possível prever as incapacidades para a vida em geral, para o trabalho, para a vida social e para a família dum deprimido [para mais informações a este respeito ver Vaz-Serra (1994)].

STAI

Para o presente estudo utilizou-se a versão traduzida, adaptada e aferida para a população portuguesa do STAI. O trabalho de tradução, adaptação e aferição deste instrumento passou por três grandes estudos encabeçados por vários autores (Santos & Silva, 1997; Silva & Campos, 1998; Silva, Silva, Rodrigues & Luís, 1999/2000; in Silva, 2006).

O STAI é um instrumento de auto-relato, que pode ser usado com adultos e adolescentes (a partir do 10º ano de escolaridade ou com idade equivalente) de ambos os géneros, e o seu campo de utilização é vasto (Silva, 2006). Tal como o BDI, também o STAI foi sofrendo alterações ao longo dos tempos. A este respeito Silva (2006, p.45-46) refere:

O Inventário de Estado-Traço de Ansiedade (STAI) surgiu em 1970 e foi construído por Spielberger, Gorsuch e Lushene, com o intuito de se dispor de uma medida do tipo relato pessoal, breve e dotada de garantia, para avaliar o Estado e o Traço de Ansiedade, na prática da clínica e da investigação (Spielberger, 1994). De início, o plano consistia na construção de um inventário constituído por um só conjunto de itens que pudesse ser aplicado, com instruções diferentes, consoante se pretendesse medir o Estado ou o Traço de Ansiedade. (...). Após uma ampla investigação tendo em vista a validade dos itens, (...), apurou-se um conjunto de 20 itens que constituíram a Forma A, versão preliminar do STAI. A investigação posterior com o Inventário veio a revelar que a alteração das instruções nem sempre lograva ultrapassar o carácter psicolinguístico mais vincadamente ligado a estado ou traço de certas palavras-chave dos itens. Este facto tinha consequências no grau de consistência da medida. Esta e outras dificuldades determinaram a alteração do modo de construção do Inventário, que levou à elaboração de duas escalas separadas de 20 itens cada, uma de Estado e outra de Traço de Ansiedade. Assim surgiu o STAI (Forma X). Em ambas as escalas, há itens que acusam a presença de ansiedade e, outros, a sua ausência. (...). A investigação continuada e o estudo levaram a que, volvidos dez anos, se publicasse uma forma revista, a Forma Y, publicada em 1983.

Esta nova versão do inventário surgiu com o intuito de: Se alcançar uma medida da ansiedade mais precisa, que discriminasse de forma mais consistente sentimentos de ansiedade e sentimentos depressivos, e indivíduos com perturbações de ansiedade e indivíduos com perturbações depressivas (diagnóstico diferencial); substituir itens cujas propriedades psicométricas eram fracas, bem como aperfeiçoar a estrutura factorial da escala de traço de ansiedade, contrabalançando de forma mais equilibrada os itens em que a ansiedade está presente e aqueles em que está ausente (Spielberger et al., 1983).

Segundo Silva (2006), a construção do STAI assenta na concepção de Freud sobre ansiedade e na distinção entre ansiedade-estado e ansiedade-traço proposta por Cattell e Scheier, e foi com esta base que Spielberger elaborou o seu próprio modelo conceptual sobre ansiedade e stress, do qual importa salientar o seguinte:

Spielberger define *estado de ansiedade* como um corte transversal temporal na corrente emocional da vida de uma pessoa e que consiste em sentimentos subjectivos de tensão, apreensão, nervosismo e preocupação, e activação do sistema nervoso autónomo. Por sua vez, *traço de ansiedade* é definido nos termos das diferenças individuais relativamente estáveis quanto à propensão para a ansiedade, isto é, diferenças inter-individuais na tendência para perceber situações stressantes como perigosas ou ameaçadoras e na inclinação para reagir a tais situações com elevações mais frequentes e intensas do estado de ansiedade (Spielberger, 1994, in Silva, 2006, p.48).

O STAI é constituído por duas escalas de tipo Likert de quatro pontos: Uma que mede a ansiedade-estado, isto é, que avalia como se sentem os indivíduos no preciso momento em que respondem ao inventário, em termos de intensidade dos sentimentos (1-Nada; 2-Um Pouco; 3-Moderadamente; 4-Muito); e outra que mede a ansiedade-traço, ou seja, que avalia como se sentem habitualmente os indivíduos, em termos da frequência dos sentimentos (1-Quase Nunca; 2-Algumas Vezes; 3-Frequentemente; 4-Quase Sempre). Ambas as escalas são constituídas por 20 itens, e em qualquer uma delas há itens em que a ansiedade está presente (como por exemplo no terceiro item da escala de ansiedade-estado: “Estou tenso”), e itens em que a ansiedade está ausente (como no item 15 da mesma escala: “Sinto-me descontraído”) – itens de cotação inversa. Na escala de ansiedade-estado existem 10 itens de cotação inversa, enquanto que na escala de ansiedade-traço existem 9.

Sendo de auto-aplicação, o STAI, tal como o BDI-II e o IACLIDE, pode ser aplicado individualmente ou em grupo, e, por norma, cada uma das escalas requer seis minutos a ser preenchida. Segundo Silva (2006), o STAI não deve ser apresentado como um instrumento de avaliação da ansiedade mas antes como “Questionário de Auto-Avaliação”.

No entanto, para o presente estudo optou-se por dizer aos participantes que iriam preencher “alguns instrumentos de medição da depressão e da ansiedade”, sem no entanto estar mencionado, na parte relativa ao STAI, que esta se referia à ansiedade. Este facto deveu-se à convicção de que as suas implicações sobre os resultados seriam mínimas, na medida em que foi dito aos participantes que o objectivo do estudo era o de analisar os

instrumentos de medição e não propriamente os participantes em si, e na medida em que, independentemente de se dizer ou não, havia uma elevada probabilidade de a grande maioria dos indivíduos (alunos de Psicologia) se aperceber dos objectivos de cada inventário. Para além do que, como Silva (2006) refere, os itens do STAI têm uma elevada validade facial (que os torna mais susceptíveis à falsificação), o que significa que é muito fácil os indivíduos se aperceberem qual o construto teórico que está por detrás daqueles itens; ainda mais fácil se torna quando os indivíduos estão dentro da área. Além disso, a acontecerem implicações, ocorreriam em todos os instrumentos, pois tanto foi referido o termo “ansiedade” como o termo “depressão”, o que não põe em causa os resultados obtidos.

As instruções do STAI são diferentes nas duas escalas: Enquanto que na escala de ansiedade-estado se pede aos indivíduos que respondam como se sentem no preciso momento em que respondem (“agora”); na de ansiedade-traço pede-se que indiquem como se sentem geralmente (“habitualmente”). É de referir que quando se faz a aplicação total do inventário, a escala de ansiedade-estado deve ser aplicada antes da escala de ansiedade-traço, de forma a “...evitar a influência do sentimento despertado pela aplicação da segunda escala na primeira” (Silva, 2006, p.57).

A cotação do STAI é feita, à semelhança dos outros instrumentos, somando as cotações de cada um dos itens, para cada uma das escalas (resultados independentes), sendo que cada um deles é cotado numa escala que vai do um ao quatro – em que o valor quatro assume uma severidade máxima em termos de intensidade e/ou frequência de sentimentos de ansiedade. No entanto, deve ter-se em atenção os itens de cotação inversa, na medida em que, se nos itens em que a ansiedade está presente a cotação é feita com o próprio número da resposta (com a ordem crescente 1, 2, 3, 4), para os itens em que a ansiedade está ausente deve-se inverter os valores da resposta – à resposta um corresponde o valor quatro, à dois o valor três, à três o valor dois, e à quatro o valor um (com uma ordem decrescente 4, 3, 2, 1). Desta forma, o score total máximo é de 80, sendo o mínimo de 20. Se por ventura um ou dois itens não forem respondidos “...o procedimento a observar será o seguinte: Calcular a nota ponderada média dos itens da escala respondidos, multiplicar esse valor por 20 e arredondar o produto obtido para o número inteiro mais alto seguinte. Havendo três ou mais itens omissos, a validade da escala é posta em questão” (Silva, 2006, p.58).

Para o presente estudo tirou-se partido duma das grandes vantagens do STAI – a possibilidade de se poder usar unicamente uma das escalas, tendo em conta os objectivos de cada utilizador. Desta forma, utilizou-se unicamente a escala de ansiedade-estado, na medida em que interessava avaliar a forma como os indivíduos se sentiam no momento presente (estado) e não como se sentem geralmente (traço).

Os dados psicométricos da versão portuguesa do STAI, obtiveram-se através de vários estudos, sendo que “O primeiro (Santos & Silva, 1997) descreve os aspectos conceptuais do Inventário, a construção e validação de uma forma experimental portuguesa” (Silva, 2006, p.50); “O segundo trabalho (Silva & Campos, 1998) compreende diversas análises com a forma final portuguesa estabelecida no estudo anterior: (1) análises de consistência interna; (2) cálculo das médias e desvios-padrão de cada uma das escalas nos grupos estudados; (3) estudos sobre garantia e validade; (4) estudos com grupos de pacientes” (Silva, 2006, p.51-52); finalmente, o terceiro trabalho foi em tudo semelhante ao segundo, mas efectuado com uma população de militares portugueses (Silva et al., 1999/2000). Apresentam-se na Tabela 2 os principais resultados (dados psicométricos) dos dois últimos estudos, pois são estes que se debruçam sobre a versão final portuguesa do STAI.

Tabela 2 – Resumo dos dados referentes à consistência interna do STAI nos vários grupos⁹

		ESCALA DE ANSIEDADE-ESTADO		ESCALA DE ANSIEDADE-TRAÇO	
		Coefficiente Alfa de Cronbach	Correlações Item-Total Corrigidas	Coefficiente Alfa de Cronbach	Correlações Item-Total Corrigidas
Grupo de Estudantes do Ensino Secundário (n♂: 361; n♀:376)	♂	0,90	0,41-0,64	0,90	0,36-0,63
	♀	0,89	0,37-0,62	0,88	0,36-0,67
Grupo de Estudantes Universitários (n♂: 263; n♀:438)	♂	0,89	0,25-0,67	0,90	0,37-0,72
	♀	0,91	0,41-0,93	0,90	0,44-0,73
Grupo de Adultos (n♂: 233; n♀:284)	♂	0,91	0,34-0,67	0,89	0,25-0,67
	♀	0,93	0,46-0,75	0,90	0,35-0,69
Grupo de Militares (n♂: 807; n♀:193)	♂	0,91	0,35-0,70	0,89	0,34-0,66
	♀	0,93	0,37-0,71	0,89	0,31-0,72

⁹ Dados retirados de Silva et al. (1999/2000) e Silva (2006).

Estes dados, relativos à consistência interna do STAI, tanto os que se referem à amostra total (alfa de Cronbach) como os que se referem aos itens (correlações item-total corrigidas), são para os autores bastante satisfatórios, pelo que consideram que a versão portuguesa do STAI detém boas qualidades psicométricas. Para além disto, foi também efectuada uma análise da estabilidade teste-reteste da qual resultou uma correlação de 0,59 para a escala de ansiedade-estado e de 0,80 para a de ansiedade-traço, valores também bastantes satisfatórios. Também a validade de construto do STAI foi provada.

Procedimento

Tal como referido anteriormente, o objectivo do presente estudo é o de estudar as características psicométricas do BDI-II na população académica do ISPA. Para tal foi utilizada uma versão traduzida do instrumento, que, apesar de ser largamente utilizada em contexto de investigação, nomeadamente no ISPA, carece de ser validada e/ou adaptada para a população portuguesa.

No processo de preparação do instrumento foi efectuada uma tradução do BDI-II, através dum método de tradução muito utilizado e referido por Hill e Hill (2008): O método “traduz-retraduz”. Este método consiste basicamente em três passos: Primeiro, faz-se a tradução do instrumento para a língua portuguesa; seguidamente, testa-se a tradução, traduzindo-se a versão em português para a língua inglesa; e finalmente, compara-se a versão original do instrumento com a versão que foi traduzida para a língua inglesa no segundo passo. Se as duas versões forem iguais ou muito semelhantes considera-se que a versão portuguesa é adequada. Após este processo, e tendo-se verificado que a versão portuguesa era adequada, construiu-se o instrumento de recolha de dados, constituído pelos três instrumentos de avaliação referidos anteriormente.

A recolha dos dados foi efectuada no ISPA e dividida em dois grandes momentos: Um primeiro, no qual se recolheram os dados do primeiro ano de Psicologia; e um segundo, no qual se recolheram os dados do segundo ao quarto anos de Psicologia e do primeiro ao

quarto anos de RIS e de DCSM. Tal como referido anteriormente, esta divisão ocorreu devido ao facto dos alunos do primeiro ano de Psicologia terem sido alvo dum estudo de investigação, anterior a este, no qual tiveram que preencher um questionário muito idêntico. Este estudo tinha como objectivo avaliar o impacto dos resultados escolares no bem-estar emocional dos estudantes universitários, em função do seu autoconceito académico. O bem-estar emocional era avaliado através de três dimensões: Depressão, ansiedade e auto-estima, que foram medidas em dois momentos (um primeiro momento anterior à época de frequências – nas duas semanas antes das frequências; e um segundo momento posterior à época de frequências – na semana depois de todos os resultados escolares terem sido lançados).

Apesar dos estudos terem objectivos diferentes, foram utilizados os mesmos instrumentos (à excepção do IACLIDE que não foi utilizado no primeiro momento), e as recolhas foram feitas nas mesmas condições experimentais: Em qualquer dos momentos foi o mesmo investigador que efectuou a recolha de dados; os dados foram recolhidos no mesmo ambiente (ISPA), em contexto de sala de aula, nos primeiros ou nos últimos 30 minutos das aulas (em função da disponibilidade dos professores); e nas mesmas condições, pois ambas as recolhas foram efectuadas nas duas semanas anteriores à época de frequências (no entanto, o primeiro momento de recolha ocorreu nas duas semanas anteriores à primeira época de frequências, enquanto que o segundo foi efectuado nas duas semanas anteriores à segunda época de frequências).

Em ambos os momentos, o investigador foi convidado a entrar na sala, no início ou no final da aula, sendo de imediato apresentado pelo professor da respectiva aula como sendo uma aluna do quinto ano de Psicologia Clínica que estava a realizar o seu trabalho de final de curso. Após esta pequena apresentação, o investigador fez a sua própria apresentação e explicação do objectivo do seu estudo, dizendo o seguinte:

...estou a fazer uma análise de alguns instrumentos de avaliação da depressão e da ansiedade, que são utilizados tanto em contexto clínico como de investigação. Como tal, estou aqui para pedir a vossa colaboração para este estudo.

Gostaria de salientar que a vossa participação é voluntária, os vossos dados e respostas são confidenciais e só servem para fins científicos.

O questionário que vão preencher é constituído por três partes distintas. Peço-vos que leiam atentamente as suas instruções, e, no caso de não perceberem alguma coisa, chamem-me para eu vos esclarecer. Tentem responder de forma espontânea, e sendo o mais verdadeiros possível. Certifiquem-se

que responderam a todas as questões, e tenham em atenção que o questionário tem frente e verso e que a última página é para ser preenchida por mim.

É de referir que esta foi a explicação dada no segundo momento de recolha de dados. Para os alunos do primeiro ano de Psicologia, a explicação foi em tudo semelhante à dada aos restantes alunos, exceptuando a parte do objectivo do estudo, na qual foi referido que se estava a fazer um estudo longitudinal de caracterização da população académica do ISPA, mais especificamente dos alunos do primeiro ano de Psicologia, no que dizia respeito aos seus pensamentos e sentimentos em relação a si próprios, como pessoas e como estudantes. No entanto, acredita-se que este facto não põe em causa os dados obtidos, ainda para mais porque foram os únicos aos quais não foi dito que iam preencher inventários que mediam a depressão e a ansiedade – algo que, como referido anteriormente, poderia pôr em causa os resultados.

Só depois desta explicação é que os questionários foram distribuídos aos indivíduos. Após o preenchimento, os questionários foram recolhidos, e, no caso das recolhas feitas no final das aulas, os alunos foram saindo da sala de aula conforme terminavam, no caso das recolhas feitas no início, o investigador saiu da sala após todos os alunos terminarem.

RESULTADOS

No estudo das características psicométricas de qualquer teste psicológico é necessário e imprescindível verificar uma série de critérios e/ou qualidades que garantem o seu valor métrico. Entre outros critérios, aqueles que se revelam como mais pertinentes para esse fim são: A fiabilidade, a sensibilidade e a validade do instrumento em questão (Maroco, 2005); critérios que são verificados através de inúmeras metodologias estatísticas, as quais se passa de imediato a apresentar.

Fiabilidade

A fiabilidade refere-se à capacidade do teste medir de forma repetida e consistente. Existem várias medidas de fiabilidade: A consistência interna; a fiabilidade de formas paralelas; a fiabilidade split-half (ou par-ímpar); a fiabilidade teste-reteste; e a fiabilidade inter-avaliador.

Para o presente estudo não foi estimada a fiabilidade teste-reteste, na medida em que não fazia grande sentido fazê-lo, pois esta é uma medida da estabilidade temporal do instrumento. Como o BDI-II é um instrumento que mede um estado (que se modifica ao longo dos tempos), e não traços de personalidade (que tendem a manter-se ao longo da vida),

seria de prever que qualquer reteste apresentasse correlações fracas com o teste inicial. Além disso, segundo Bech (1992, in Vaz-Serra, 1994, p.55-56) “O método teste-reteste não é relevante (como medida de fiabilidade nas escalas para medir depressão)... pois as pontuações têm tendência a variar, no período entre dois testes, devido à flutuação de sintomas”.

Para além da fiabilidade teste-reteste, também não foi estimada a fiabilidade inter-avaliador e a fiabilidade de formas paralelas, na medida em que a primeira não era exequível e a segunda não fazia sentido. Além disso, sendo a consistência interna considerada uma boa estimativa da fiabilidade dum instrumento, também não se recorreu à fiabilidade split-half (ou par-ímpar), até porque esta técnica

...constitui uma forma de provar se uma das metades dos itens da escala é tão consistente a medir o construto como a outra metade. Um problema que se levanta é que a correlação “par/ímpar” representa a correlação que se estabelece entre duas versões menores da própria escala, o que subestima a sua verdadeira fiabilidade, uma vez que esta é directamente proporcional ao número de itens que a contêm. Este valor costuma ser corrigido pelo Coeficiente de Spearman-Brown... (Vaz-Serra, 1994, p.53-54).

Desta forma, obteve-se no presente estudo um coeficiente alfa de Cronbach de 0,89 (e um valor idêntico para os itens standardizados), pelo que se pode considerar que o instrumento tem uma boa consistência interna (ver Anexo F). Além disso, é de referir que todos os itens da escala são significativos para a consistência interna do conjunto dos itens, na medida em que, quando retirados um a um, o valor do alfa de Cronbach baixa sempre. É ainda de referir que as correlações item-total corrigidas variaram entre 0,34 (agitação) e 0,64 (perda de interesse).

Sensibilidade

Apesar de se ter referido uma série de parâmetros em relação à análise dos itens, neste estudo, assim como acontece na grande maioria dos estudos de análise das características

psicométricas ou de validação e/ou adaptação de instrumentos psicométricos, optou-se por se analisar o parâmetro que se revela mais importante e pertinente: A sensibilidade (ou capacidade de discriminação). Esta refere-se à capacidade que os itens têm de diferenciar os indivíduos no que toca ao traço latente que o teste está a medir, ou seja, os itens têm capacidade de discriminação quando diferenciam os indivíduos que obtêm scores elevados dos que obtêm scores baixos.

Para a análise da sensibilidade dos itens usou-se a metodologia proposta por Maroco (2005), que refere que um item tem elevada sensibilidade "...quando consegue discriminar os indivíduos, estruturalmente diferentes, que mede". Segundo o autor, a sensibilidade dos itens poderá ser analisada através da forma da distribuição das respostas (ver Anexo C), considerando-se que eles têm capacidade de discriminação, ou sensibilidade elevada, quando apresentam um enviesamento e um achatamento a tender para zero e que valores acima de um revelam problemas de sensibilidade. Para além desta metodologia, utilizou-se também uma outra, a que foi usada por Beck et al. (1996) no estudo de validação do BDI-II, para reforçar os resultados da análise anterior. Esta metodologia baseia-se na análise da probabilidade de cada uma das opções de resposta de cada item em função do score total, considerando-se que os itens são sensíveis quando reflectem uma cotação ordinal apropriada, isto é, quando as várias opções de resposta se diferenciam umas das outras quanto ao score total.

Forma da Distribuição dos Itens

Existem duas medidas que permitem caracterizar a forma da distribuição dos elementos da população: As medidas de assimetria e as de achatamento. Em relação à forma da distribuição dos itens do BDI-II, dada pelas medidas de assimetria e achatamento, obtiveram-se os dados que se encontram em anexo (ver Anexo G), em relação a cada um deles, através dos quais se pode constatar que todos os itens têm uma distribuição assimétrica à direita, o que significa que a distribuição das respostas se concentra do lado esquerdo, ou seja, se concentra nas opções de resposta zero e um, e que, à excepção de dois itens (perda de energia e dificuldades de concentração), todos os outros têm uma distribuição leptocúrtica,

isto é, uma distribuição pontiaguda, o que significa que a distribuição das respostas foi muito “intensa” na vizinhança dos valores centrais.

Pela análise dos dados é possível constatar que dos 21 itens 5 deles se apresentam com uma elevada sensibilidade (sentimentos de culpa, perda de energia, alterações nos hábitos do sono, dificuldades de concentração e cansaço ou fadiga), pois os seus coeficientes de assimetria e de achatamento têm valores menores que um e a tender para zero. Dos restantes itens, poder-se-á considerar que seis deles se apresentam não com uma elevada sensibilidade, mas com uma sensibilidade “boa ou moderada” (pessimismo, autocrítica, agitação, indecisão, irritabilidade e alterações no apetite), na medida em que um dos dois coeficientes tem valores inferiores ou iguais a um e o outro coeficiente tem valores entre 1,06 e 1,85.

Em relação aos 10 itens que se apresentam com uma fraca sensibilidade (tristeza, fracasso no passado, perda de prazer, sentimentos de punição, desprezo por si próprio, pensamentos ou desejos suicidas, choro, perda de interesse, desvalorização pessoal e perda de interesse sexual), é possível considerar que 4 deles revelam graves problemas de sensibilidade (tristeza, sentimentos de punição, pensamentos ou desejos suicidas e perda de interesse sexual), pois os seus coeficientes de assimetria têm valores entre 2,47 e 5,11 e os de achatamento têm valores entre 6,55 e 33,32.

Probabilidade das Opções de Resposta em Função do Score Total

Uma outra forma possível de se avaliar a sensibilidade dos itens do BDI-II é analisando a probabilidade de cada uma das opções de resposta, de cada um dos itens, em função do score total, que, no fundo, reflecte quão bem as quatro opções de resposta de cada um dos itens se diferenciam umas das outras. Esta análise permite então inferir se os itens reflectem ou não uma cotação ordinal apropriada em relação à discriminação entre indivíduos com depressão severa (com scores mais altos no inventário) daqueles com depressão mínima ou ligeira (com scores mais baixos no inventário).

Esta análise realiza-se facilmente através das representações gráficas das probabilidades de cada uma das opções de resposta, para cada um dos itens, em função do score total do BDI-II, que se encontram em anexo (ver Anexo H), através das quais ressalta a constatação de que dos 21 itens 2 deles não reflectem uma cotação ordinal apropriada (sentimentos de punição e pensamentos ou desejos suicidas).

Constata-se que no item sentimentos de punição, a probabilidade da terceira opção de resposta (com uma cotação de dois pontos) é sempre zero seja qual for o score total associado, isto é, é sempre inferior à probabilidade dos indivíduos seleccionarem as restantes opções de resposta, nomeadamente a primeira e a segunda (cotadas, respectivamente, com zero e um pontos), o que significa que mesmo os indivíduos com depressão severa (com scores entre 29 e 63) têm maior probabilidade de seleccionar as duas primeiras opções de resposta do que a terceira. Em relação ao item pensamentos ou desejos suicidas pode-se observar que a segunda opção de resposta (cotada com um ponto) tem maior probabilidade de ser seleccionada por indivíduos com scores entre 30 e 50 (que correspondem a depressões severas) do que a terceira e a quarta opções de resposta (cotadas com dois e três pontos); além disso, é mais provável de ser seleccionada do que a terceira opção de resposta (cotada com dois pontos) pelos indivíduos com scores entre os 50 e os 63 pontos.

Validade

A validade dum instrumento prende-se fundamentalmente com o facto dele medir exactamente o que pretende medir, isto é, “Um instrumento ou teste diz-se «válido» se mede o que é suposto medir e não existirem erros na interpretação dos resultados e/ou lógica na elaboração das conclusões” (Maroco, 2005). Considera-se que existem três tipos de validade: Validade de construto, validade de critério e validade de conteúdo. A escolha dos procedimentos de validação depende do uso que vai ser dado teste, podendo desta forma o processo de validação recair unicamente num dos tipos de validade referidos (Anastasi & Urbina, 2000).

O presente estudo não se vai deter na validade de conteúdo, que diz respeito ao facto dos itens medirem exactamente o que dizem medir, isto é, um teste tem validade de conteúdo se for considerado uma amostra representativa do domínio de comportamento que ele mede. A justificação para não se fazer a validação do conteúdo, prende-se com o facto desta validação ser iniciada aquando da própria construção do instrumento e terminar com a análise teórica e empírica (estatística) dos itens. Assim sendo, estando a analisar as qualidades psicométricas dum instrumento, deveria-se passar automaticamente para a fase da análise teórica e empírica dos itens. Ora a análise teórica dos itens não era exequível, pois seriam necessários especialistas da área para os avaliarem. No que diz respeito à análise empírica dos itens esta foi feita, não de forma exaustiva, mas recaindo no parâmetro que se considera ser o mais importante para este fim – a referida análise da sensibilidade (ou da discriminação) dos itens. Para além disso, sendo que a validade de conteúdo “É a componente da validade menos relevante” (Maroco, 2005), e sabendo que na maioria dos estudos das características psicométricas de instrumentos não se analisa a validade de conteúdo dos mesmos, a análise da validade do BDI-II vai centrar-se nas validades de construto e de critério.

Validade de Critério

A validade de critério tem a ver com o facto do teste ser capaz de predizer, ou de estar em concordância com, um determinado comportamento (desempenho ou estado), comportamento que se constitui como o critério contra o qual o próprio teste está a ser avaliado. Existem dois tipos de validade de critério: A validade preditiva e a validade concorrente, sendo que a sua diferenciação se baseia nas relações temporais que se estabelecem entre a recolha dos dados do teste e da medida de critério. Para além disso, a sua diferenciação também se pode basear no próprio objectivo da testagem, considerando-se que a validação concorrente faz sentido para os testes cujo objectivo é a avaliação do status existente, e não a predição de resultados futuros (Anastasi & Urbina, 2000). Cronbach (1996) refere ainda que a validação concorrente é perfeitamente lógica quando se propõe um teste como substituto dum outro já existente, esperando-se que o teste proposto indique aquilo que o outro já existente diria a respeito da pessoa nesse momento.

Desta forma, considera-se ter fundamentada a opção de analisar a validade de critério através da validação concorrente – neste estudo pretende-se analisar as características psicométricas dum instrumento que mede o grau de depressão nos indivíduos (estado) e quer-se basicamente confirmar que ele de facto o faz, isto é, que ele de facto mede a depressão, contrapondo-o com um outro instrumento, existente no mercado, que é considerado válido para esse fim.

Validade Concorrente

Na validação concorrente, a recolha dos dados do teste e da medida de critério ocorre, normalmente, em simultâneo. Este facto não é de todo relevante para a análise da validade de critério. O que é de facto importante e imprescindível é que se utilize um critério válido para a análise. Para este estudo, o critério usado para determinar a validade concorrente do BDI-II foi um outro instrumento, que mede o mesmo construto teórico (depressão) – o IACLIDE. Como Pasquali (2003, p.187-188) refere,

O que vale aqui é o princípio de que se houver um teste comprovadamente válido para a medida de algum traço latente, ele certamente pode servir de critério para a validação de um novo teste. Espera-se neste caso que a correlação do novo teste seja elevada, de pelo menos 0,75. (...). Pode-se concluir que a validade concorrente só faz sentido se existirem testes comprovadamente válidos que possam servir de critério contra o qual se quer validar um novo teste e que este novo teste tenha algumas vantagens sobre o antigo (como, por exemplo, economia de tempo, etc.).

Assim, na análise da validade concorrente, a correlação de Pearson que se obteve entre os scores do BDI-II e os do IACLIDE, correlação que se constitui como a medida do coeficiente de validade concorrente, foi de 0,91, correlação positiva e altamente significativa (ver Anexo I). Fica desta forma demonstrada a validade de critério do BDI-II.

Validade de Construto

Passa-se agora para a análise da validade de construto, para a qual a perspectiva de Maroco (2005) vai servir de base. O autor refere que esta validade

Avalia a lógica de construção de escalas de conceitos psicométricos a partir de itens de mensuração fácil. Se uma escala tem validade de construto, então mede o que se propõe medir e a operacionalização dos itens no construto, permitem inferir com confiança para os construtos teóricos que estão na base da construção da escala.

O estabelecimento da validade do construto é um processo longo e complexo. Em primeiro lugar exige uma sólida base teórica, que depois deve ser operacionalizada claramente nos itens ou indicadores que medem esse construto. Um construto com reduzida validade pode resultar de uma fraca operacionalização da teoria, ou da utilização de itens que medem coisas diferentes para sujeitos diferentes.

Para Maroco (2005), assim como para Cronbach (1996), é através de dois tipos de validade que se determina a validade de construto: A validade convergente e a validade discriminante. A primeira verifica-se quando os itens que constituem o construto estão correlacionados entre si e quando o próprio construto se correlaciona positivamente com outros construtos que medem o mesmo conceito teórico. A validade discriminante existe quando os itens que medem um determinado construto não se encontram correlacionados com itens que medem um construto diferente e quando o próprio construto não se correlaciona com outros construtos que medem conceitos teóricos diferentes. A avaliação da validade de construto é feita “...com a Análise Factorial Confirmatória, especialmente quando a construção dos construtos e dos itens que os definem é assente numa base teórica sólida; e com a análise de Correlação Bivariada” (Maroco, 2005).

Para a análise da validade de construto, vai-se acrescentar à validade convergente e discriminante a validade factorial, na medida em que ambas se fazem através da análise factorial confirmatória e na medida em que nos estudos das características psicométricas ou de validação e/ou adaptação de instrumentos tem que se fazer, no caso dos instrumentos reflectirem factores ou dimensões subjacentes, o “teste de adequação” do modelo factorial à população em estudo. Para além disso, Hill e Hill (2008) referem mesmo que a validade de construto é de três tipos: Convergente, divergente (ou discriminante) e factorial.

Validade Factorial

Quando se faz um estudo das características psicométricas ou a validação e/ou adaptação dum instrumento para uma determinada população e esse instrumento tem um modelo factorial, intrínseco à sua população de “origem”, tem que se testar se esse modelo factorial é adequado ou não para a população em estudo. Para tal, é necessário recorrer à análise factorial confirmatória do modelo (ver Anexo A). Caso o modelo não se adequa à população em

estudo, deverá proceder-se a uma análise factorial exploratória com essa população e, de seguida, proceder-se a uma nova análise factorial confirmatória para se comprovar que esse modelo factorial é, de facto, verdadeiramente adequado à população em estudo. Como Tukey (1977, in Murteira, 1993, p.12) refere “...se a análise exploratória de dados é trabalho de detective, que procura pistas e evidência, a análise confirmatória de dados é trabalho judicial ou quase-judicial, que analisa e avalia a força das provas e da evidência”.

Tal como referido anteriormente, na análise factorial do BDI-II, Beck et al. (1996) extraíram dois factores em cada uma das suas amostras – na população clínica e na população estudantil. Para o presente estudo teve-se em consideração o modelo factorial que foi extraído para a população estudantil, pois é a única população passível de comparações. Desta forma, na população estudantil o BDI-II reflectiu uma dimensão cognitivo-afectiva da depressão, constituída pelos itens: Tristeza, fracasso no passado, perda de prazer, sentimentos de culpa, sentimentos de punição, desprezo por si próprio, auto-crítica, pensamentos ou desejos suicidas, choro, agitação, perda de interesse, indecisão, desvalorização pessoal e irritabilidade; e uma dimensão somática da depressão, constituída pelos itens: perda de energia, alterações nos hábitos do sono, alterações no apetite, dificuldades de concentração e cansaço ou fadiga. Foi com base neste modelo que foi efectuada a análise factorial confirmatória através do programa de análise estatística Amos 4.0 (Arbuckle & Wothke, 1995-1999) e que se encontra em anexo (ver Anexo J).

Através da análise factorial confirmatória é possível extrair inúmeros dados, dos quais, para a verificação da adequação do modelo à população em estudo, são importantes de reter os seguintes indicadores (indicadores da qualidade da análise factorial e da adequação do modelo à população em causa): O CFI (comparative fit index), o GFI (goodness of fit index), e o RMSEA (root mean square error of approximation).

Segundo Arbuckle e Wothke (1995-1999) existem valores para estes indicadores de qualidade que reflectem uma boa adequação do modelo factorial à população em estudo. Desta forma, consideram que o RMSEA deverá ser igual ou inferior a 0,05 para que se admita uma boa adequação do modelo em relação aos graus de liberdade, caso contrário, isto é, valores superiores a 0,05 põem em causa a adequação do modelo; o CFI atinge valores entre zero e um, pelo que valores próximos de um indicam uma óptima adequação do modelo;

finalmente, o GFI também se encontra entre zero e um, sendo que um significa uma adequação perfeita.

É possível verificar que o modelo factorial de Beck et al. (1996) para a população estudantil não se adequa à população académica do ISPA (ver Anexo J), na medida em que os indicadores, referidos anteriormente, reflectiram os seguintes valores: O RMSEA foi de 0,066 (valor superior a 0,05, valor máximo para a consideração duma boa adequação do modelo); o CFI foi de 0,88 e o GFI de 0,91, valores razoáveis, mas não “perfeitos” (o ideal seria obter valores superiores a 0,90).

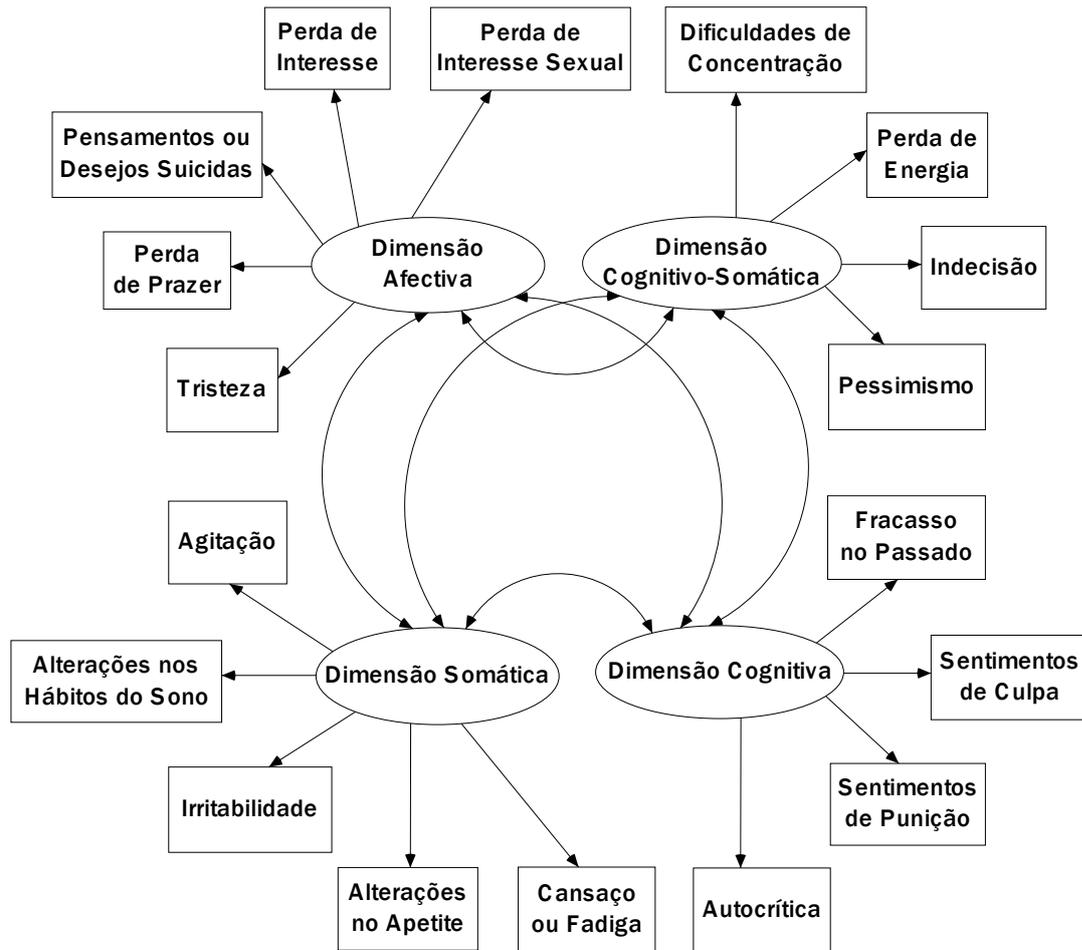
Desta forma, não havendo uma adequação do modelo de Beck et al. (1996) à população académica do ISPA, procedeu-se a uma análise factorial exploratória para a população em estudo e, de seguida, a uma análise factorial confirmatória para comprovar se o modelo factorial resultante da análise factorial exploratória é ou não adequado. Através da análise factorial exploratória (ver Anexo K) foi possível extrair os seguintes dados:

- A partir da intercorrelação dos 21 itens do BDI-II obteve-se um KMO [Kaiser-Meyer-Olkin; “...uma medida da homogeneidade das variáveis...” (Maroco, 2003, p.267)] de 0,93, o que significa que a qualidade dos dados é excelente, isto porque o KMO “avalia a factorabilidade do instrumento, isto é, se as correlações entre os itens permitem definir construtos ou factores” (Maroco, 2005) e porque se pode assumir que valores entre 0,9 e 1,0 são excelentes (Maroco, 2003).
- Como o teste de Esfericidade de Bartlett apresenta um p-value inferior a 0,001, podemos concluir que as variáveis estão significativamente correlacionadas.
- Através da análise de componentes principais, e cumprindo a regra de retenção de factores com valores próprios superiores a um, foi possível reter quatro factores com os seguintes valores próprios: 6,93; 1,56; 1,10 e 1,04, que explicam cerca de 51% (50,65%) da variabilidade total, sendo que o primeiro factor explica 33,02% da variabilidade total; o segundo factor 7,43%; o terceiro factor explica 5,24% e, finalmente, o quarto factor explica 4,96% do total da variância.

- Na medida em que na análise factorial de Beck et al. (1996) se constatou que o BDI-II reflectia factores fortemente correlacionados, e na medida em que é difícil conceber dimensões da depressão que não estejam correlacionadas entre si, foi usada uma rotação Promax para identificar as dimensões da depressão.
- Considerando os itens com pesos factoriais nos factores iguais ou superiores a 0,47, os itens salientes no primeiro factor são: Tristeza, perda de prazer, pensamentos ou desejos suicidas, perda de interesse e perda de interesse sexual, pelo que se pode considerar que esta é uma dimensão afectiva da depressão, na medida em, apesar de englobar sintomas de todos os tipos (tristeza, perda de prazer e perda de interesse – sintomas afectivos; pensamentos ou desejos suicidas – sintoma cognitivo e perda de interesse sexual – sintoma somático), são os sintomas afectivos que estão em maior número e com maior peso factorial.
- Os itens salientes no segundo factor são: Agitação, alterações nos hábitos do sono, irritabilidade, alterações no apetite e cansaço ou fadiga. Com a excepção da irritabilidade, que é um sintoma afectivo, todos os outros são sintomas somáticos e na medida em que são estes que têm pesos factoriais mais elevados, considera-se que esta é uma dimensão somática da depressão.
- No terceiro factor incluem-se os seguintes itens: Pessimismo, indecisão, perda de energia e dificuldades de concentração. Sendo os dois primeiros sintomas cognitivos e os dois últimos sintomas somáticos, e tendo estes sintomas pesos factoriais idênticos, pode-se considerar que esta é uma dimensão cognitivo-somática da depressão.
- Finalmente, no quarto e último factor foram os seguintes itens que se salientaram: Fracasso no passado, sentimentos de culpa, sentimentos de punição e autocrítica, que, sendo todos sintomas cognitivos, fazem com que esta seja uma dimensão cognitiva da depressão.
- É ainda de referir que as correlações entre os factores variaram entre 0,21 (dimensão somática e dimensão cognitiva da depressão) e 0,49 (dimensão afectiva e dimensão cognitiva da depressão). Além disso, foi a dimensão afectiva da depressão que se correlacionou mais fortemente com todas as outras dimensões.

Submetendo estes dados a uma análise factorial confirmatória (ver Anexo L), constata-se que existe uma boa adequação do modelo factorial à população em estudo, na medida em que o RMSEA foi de 0,049 (valor inferior a 0,05), o CFI foi de 0,94 e o GFI de 0,95, indicadores que reflectem uma óptima adequação do modelo, que se esquematiza na Figura 3.

Figura 3 – Modelo factorial da população académica do Ispa



Validade Convergente

A validade convergente, tal como referido, verifica-se quando os itens que constituem o construto se correlacionam entre si e quando o próprio construto (instrumento) se correlaciona positivamente com outros construtos (instrumentos) que medem o mesmo conceito teórico. Segundo Maroco (2005), existem vários critérios que têm de ser comprovados para que se verifique esta validade. Um deles é o KMO, de que já se falou anteriormente. Como o KMO observado para os dados é de 0,93, o que significa que a qualidade dos dados é excelente e

que as correlações entre os itens permitem definir construtos ou factores, a validade convergente está verificada.

Para além do KMO, também os pesos factoriais dos itens nos factores, bem como a correlação inter-itens e/ou consistência interna dos itens que constituem os factores, servem para a verificação desta validade. Segundo Maroco (2005), os pesos factoriais dos itens nos factores deverão ser iguais ou superiores a 0,50. Para a presente análise, consideraram-se os itens com pesos factoriais iguais ou superiores a 0,47¹⁰, na medida em que se revelaram úteis para a constituição e “caracterização” das dimensões e porque não põem em causa a qualidade dos dados – através da análise factorial confirmatória do modelo da população académica do ISPA com itens com pesos factoriais nos factores iguais ou superiores a 0,50, obtiveram-se os seguintes dados: RMSEA de 0,046; CFI de 0,96 e GFI de 0,97, que evidenciam igualmente uma óptima adequação do modelo, melhor do que a considerada é certo, mas que não justifica o “sacrifício” de quatro itens (ver Anexo M). Desta forma, não existe motivo para não considerar que a validade convergente está verificada.

Em relação à correlação inter-itens que constituem cada factor (ver Anexo N), verifica-se que todas as correlações são significativas (para $\alpha=0,01$). No primeiro factor as correlações variaram entre 0,16 (pensamentos ou desejos suicidas e perda de interesse sexual) e 0,60 (perda de prazer e perda de interesse); no segundo factor variaram entre 0,22 (agitação e alterações no apetite) e 0,43 (alterações nos hábitos do sono e cansaço ou fadiga); no terceiro factor variaram entre 0,27 (pessimismo e perda de energia) e 0,44 (perda de energia e dificuldades de concentração); e, finalmente, no quarto factor as correlações variaram entre 0,18 (sentimentos de punição e autocrítica) e 0,32 (fracasso no passado e sentimentos de culpa). Na medida em que se verifica que os itens estão relacionados entre si de forma significativa, apesar de algumas das correlações serem baixas, pode-se dizer que estes dados reforçam a verificação da validade convergente.

Também a consistência interna dos itens que constituem os factores, sendo de 0,76 para a primeira dimensão, 0,72 para a segunda dimensão, 0,68 para a terceira dimensão e de 0,61 para a quarta dimensão (ver Anexo O), não põe em causa a validade convergente, apesar dos valores se encontrarem entre 0,70 e 0,80 para as duas primeiras dimensões e entre 0,60 e

¹⁰ É de referir que só existem quatro itens com pesos factoriais inferiores a 0,50 – um na primeira dimensão, um na terceira e dois na quarta dimensão.

0,70 para a terceira e quarta dimensões, valores que estão abaixo daquilo que é habitualmente aceitável. No entanto, sabe-se à partida que este coeficiente é fortemente afectado quer pela variabilidade da amostra (que não está aqui em causa), quer pelo comprimento do instrumento. Ora, se se está a analisar a consistência interna de factores constituídos por quatro e cinco itens, não se pode esperar que estes tenham uma elevada consistência interna, pois sabe-se que esta é tanto maior quanto maior for o número de itens. Para além disso, como Ribeiro (2008) refere, aceitam-se valores de alfa entre 0,60 e 0,80 quando os instrumentos são constituídos por poucos itens, como é o caso dos factores em questão.

Finalmente, outra das formas de se verificar a validade convergente, e a mais usada nos estudos das características psicométricas ou de validação e/ou adaptação de instrumentos, é através da correlação do instrumento com outros instrumentos com construtos teóricos equivalentes. Admite-se que um instrumento apresenta validade convergente quando produz resultados positivamente correlacionados com outros instrumentos que medem o mesmo construto teórico e essas correlações deverão ser superiores a 0,80. Assim, foi efectuada uma correlação de Pearson (ver Anexo P) entre os scores do BDI-II e os do IACLIDE, tendo resultado uma correlação fortemente positiva entre os dois instrumentos ($r=0,91$), pelo que fica mais uma vez comprovada a validade convergente.

Validade Discriminante

No sentido oposto à validade convergente, a validade discriminante verifica-se quando os itens que definem um construto não se relacionam com itens que medem um construto diferente e quando o próprio construto (instrumento) não se correlaciona com outros que medem conceitos teóricos diferentes. Ou seja, considera-se que um instrumento apresenta validade discriminante quando não mede aquilo que não é suposto medir. Segundo Maroco (2005) há três formas ou indicadores para verificar a validade discriminante, que se passa a citar: Existirem itens com pesos factoriais baixos (inferiores a 0,50) em factores que medem “conceitos” diferentes, bem como correlações inter-itens de outros factores baixas, e, finalmente, haverem correlações nulas com factores ou construtos teoricamente diferentes.

É possível verificar que, à excepção do item sentimentos de punição, que “pontuou” tanto na quarta dimensão como na primeira (com pesos factoriais de 0,49 e de 0,40, respectivamente), e do item pensamentos ou desejos suicidas, que pontuou igualmente na

primeira e na quarta dimensões (com pesos factoriais de 0,47 e de 0,41, respectivamente), todos os itens têm pesos factoriais baixos nas outras dimensões que não as próprias (ver Anexo K). Este facto não põe em causa a validade discriminante, não só por serem unicamente dois itens, mas porque os pesos factoriais que eles assumiram nas outras dimensões que não as suas são inferiores a 0,47 (valor mínimo estabelecido para os itens fazerem parte dos factores).

Em relação à correlação inter-itens de factores diferentes (ver Anexo N), pode-se constatar que todos os itens, seja de que dimensão for, se correlacionam de forma significativa entre si (para $\alpha=0,01$ e $\alpha=0,05$). Facto que não é de estranhar, pois assumiu-se à partida que o modelo tinha factores correlacionados, o que implica que também os itens estejam correlacionados entre si. Pode-se contudo constatar, que duma maneira geral as correlações que se estabelecem entre os itens dum factor com os itens dos restantes factores tendem a ser mais baixas do que as correlações que se estabelecem entre os itens dentro do mesmo factor. Esta é mais uma prova da validade discriminante do BDI-II, apesar de existirem itens que se correlacionam mais fortemente com itens de outros factores do que com os do próprio factor.

Finalmente, quanto à verificação de haverem correlações nulas com factores ou construtos teoricamente diferentes, efectuou-se uma correlação de Pearson (ver Anexo P) entre os scores do BDI-II e os do STAI, na medida em que um avalia o grau de depressão e outro o grau de ansiedade. Constata-se que a correlação entre os dois instrumentos foi significativa (para $\alpha=0,01$) e moderada ($r=0,73$). Este facto não é de todo inesperado, na medida em que, apesar de serem construtos teóricos diferentes, são construtos que se encontram habitualmente correlacionados na avaliação clínica da depressão. No entanto, o facto do BDI-II se correlacionar fortemente com o IACLIDE ($r=0,91$) e moderadamente com o STAI ($r=0,73$), juntamente com o facto destes últimos terem uma correlação superior ($r=0,78$) à encontrada entre o BDI-II e o STAI, pode evidenciar uma validade discriminante, entre ansiedade e depressão, fazendo o mesmo tipo de raciocínio que Beck et al. (1996) fizeram na sua análise da validade discriminante do instrumento.

Dados Sócio-Demográficos do BDI-II na População Académica do ISPA

Resta referir que não foram encontradas diferenças significativas entre os géneros, os cursos, os anos e os grupos etários, no que diz respeito aos scores do BDI-II (ver Anexo Q). Ou seja, não existem diferenças significativas entre os scores do BDI-II dos estudantes do género feminino e do género masculino. Da mesma forma que não foram encontradas diferenças significativas entre os alunos de Psicologia, de RIS e de DCSM. Relativamente aos vários anos que foram amostrados, também não existem diferenças significativas entre os estudantes do primeiro, do segundo, do terceiro, e do quarto anos. Finalmente, no que diz respeito à idade, constata-se que também não existem diferenças significativas entre os vários grupos etários.

DISCUSSÃO

O estudo das características/qualidades psicométricas de qualquer instrumento psicológico é, tal como referido anteriormente, muito semelhante aos estudos de validação e/ou adaptação desses instrumentos. A este respeito Bradley (1994, in Ribeiro, 2008) refere que para que a adaptação dum instrumento psicológico seja adequada é necessário proceder-se de forma muito semelhante ao processo de construção inicial do mesmo, ou seja, a forma traduzida do instrumento tem que ser submetida aos mesmos procedimentos psicométricos pelos quais a sua versão original passou, de maneira a verificar os critérios e/ou qualidades que garantem o seu valor métrico.

Como referido por Hill e Hill (2008), quando se constrói um instrumento para definir uma variável latente (neste caso, a depressão) que seja medida adequadamente, é necessário efectuar dois passos, a saber: A selecção de itens adequados para definir a variável latente e a determinação da adequação do instrumento para medir essa mesma variável. Ou seja, no processo de construção dum qualquer instrumento é imprescindível a verificação da adequacidade dos vários itens e do instrumento em si para medir a variável latente, pelo que tem que se testar, para além de algumas características dos itens, se a fiabilidade e a validade da medida são suficientemente adequadas para que ela possa ser usada com confiança (Hill & Hill, 2008).

Desta forma, e tal como referido a propósito dos resultados, foram verificados no presente estudo os seguintes critérios e/ou qualidades psicométricas do BDI-II: A fiabilidade (consistência interna), a validade de critério (concorrente), a validade de construto

(convergente, discriminante e factorial) e a sensibilidade dos itens. A análise efectuada não só vai ao encontro do que a grande maioria dos estudos das características psicométricas ou de validação e/ou adaptação de instrumentos referem como análise pertinente, mas também vai ao encontro do que foi efectuada por Beck et al. (1996) no processo de validação do instrumento em causa.

No entanto, apesar de no processo de validação do BDI-II, Beck et al. (1996) para além de terem verificado a fiabilidade do instrumento em termos da sua consistência interna (através das técnicas alfa), também o fizeram em termos da sua estabilidade temporal (através da metodologia teste-reteste), esta metodologia não foi usada no presente estudo. Isto porque a metodologia de teste-reteste não é relevante para a estimativa da fiabilidade em escalas que medem depressão, na medida em que os seus resultados tendem a variar ao longo dos tempos, pois os próprios sintomas flutuam (Bech, 1992, in Vaz-Serra, 1994).

Para além disso, no manual de utilização do BDI-II, Beck et al. (1996) fazem referência à validade de conteúdo do instrumento, considerando-a adequada, pelo facto dos itens do BDI-II reflectirem os critérios de diagnóstico do DSM-IV para as perturbações depressivas, o que vai ao encontro dos objectivos do próprio instrumento – foi especialmente desenvolvido para aferir sintomas depressivos correspondentes aos referidos critérios.

Não tendo sido efectuada a validação do conteúdo do BDI-II, este facto pode ser apontado como uma das limitações do presente estudo. Segundo Pasquali (2003) a validação de conteúdo passa, entre outras coisas, por uma análise teórica dos itens, na qual se pretende avaliar a compreensão e a pertinência dos mesmos. Para tal, seria necessário recorrer a juízes que, no caso da avaliação da compreensão dos itens, seriam os próprios indivíduos testados, e, no caso da avaliação da pertinência dos itens, seriam peritos na área em questão (neste caso, na área da depressão). Esta análise é habitualmente efectuada após o processo de “tradução-retroversão” do instrumento, mas tal não era exequível no âmbito do presente estudo.

No entanto, apesar de ser uma limitação do estudo, acredita-se que esta não põe em causa os dados, não só pelo facto deste tipo de validade ser considerada por alguns autores, como Maroco (2005), como a componente da validade menos importante e/ou relevante, mas também porque se se considerar que a validade de conteúdo dum instrumento está garantida se “os atributos inclusos no teste são uma amostra representativa dos que se pretende medir”

(Ribeiro, 2008, p.94), e tendo em conta que o BDI-II pretende reflectir os critérios de diagnóstico do DSM-IV para as perturbações depressivas e que, de facto, o faz, pode-se então considerar que a validade de conteúdo do BDI-II está assegurada. Para além disso, é de referir que não é habitual nos estudos das características psicométricas de instrumentos psicológicos realizar-se uma análise à validade de conteúdo dos mesmos.

Beck et al. (1996) analisaram também a capacidade de discriminação de diagnóstico, que é uma forma de análise da validade concorrente do instrumento. Este tipo de validação também não foi efectuada no presente estudo, o que remete para outra das limitações do mesmo: O facto de se ter trabalhado unicamente com uma amostra de estudantes universitários (população académica, dita normal), torna impossível qualquer generalização dos resultados a outras populações, nomeadamente à população clínica, e impossibilita também a análise da validade concorrente baseada no critério “diagnóstico psiquiátrico”, tal como referido por Pasquali (2003).

Como tal sugere-se um novo estudo, que incida numa população clínica, não só para se obterem dados relativos à validade concorrente do BDI-II baseada no diagnóstico psiquiátrico, que permita inferir acerca da capacidade de discriminação de diagnóstico do instrumento, mas também para se obterem dados semelhantes aos do presente estudo (quanto à fiabilidade, validade de critério e construto, e sensibilidade dos itens) que, conjuntamente, podem fornecer dados relevantes para uma futura validação e/ou adaptação do BDI-II para a população portuguesa.

Apesar destas limitações, é de referir novamente que a análise dos vários critérios e/ou qualidades psicométricas aqui efectuada, é em tudo semelhante à grande maioria dos estudos de análise das características psicométricas ou de validação e/ou adaptação de instrumentos, que habitualmente se centram na análise da fiabilidade (consistência interna), validade de critério (concorrente) e da validade de construto (principalmente na factorial).

Após a exaustiva análise estatística que foi efectuada pode-se concluir que o BDI-II demonstra uma fiabilidade adequada, isto é, demonstra capacidade de medir de forma repetida e consistente, pois obteve-se um valor de alfa de Cronbach elevado, correspondente a uma boa consistência interna do instrumento. Além disso, da análise de quais os itens contribuem

bem para a consistência interna do BDI-II, e de quais não o fazem, pode-se concluir que todos eles são significativos para a consistência interna do conjunto dos itens.

Relativamente à análise da sensibilidade dos itens do BDI-II, os dados não abonam muito a favor do instrumento, na medida em que se constata que quase metade dos itens revela problemas de sensibilidade. De facto, verifica-se que quase metade dos itens tem muitas dificuldades em discriminar os indivíduos estruturalmente diferentes – indivíduos que obtêm scores elevados (deprimidos) e indivíduos que obtêm scores baixos (não deprimidos), ou seja, mesmo os indivíduos que podem ser considerados deprimidos tendem a seleccionar as opções de resposta zero e um, que são habitualmente seleccionadas (ou assim se espera) por indivíduos não deprimidos. Esta situação é particularmente evidente em quatro dos dez itens que revelam problemas de sensibilidade, itens que se referem à tristeza, aos sentimentos de punição, aos pensamentos e desejos suicidas e à perda de interesse sexual.

A análise gráfica da probabilidade de cada uma das opções de resposta, de cada um dos itens, em função do score total, efectuada pois esta foi a análise que Beck et al. (1996) fizeram, veio reforçar os resultados da análise da sensibilidade dos itens. De facto, os dados mais salientes resultantes da análise gráfica da probabilidade das opções de resposta vão ao encontro dos dados resultantes da análise da forma da distribuição dos itens, pelo que se verifica que há claramente dois itens, os que se referem aos sentimentos de punição e aos pensamentos e desejos suicidas, que não reflectem uma cotação ordinal apropriada em relação à discriminação entre indivíduos com depressão severa (com scores mais elevados) e indivíduos com depressão mínima ou ligeira (com scores mais baixos). É de referir que estes dados vão ao encontro dos reportados por Beck et al. (1996), que encontraram quatro itens que não apresentaram uma cotação ordinal apropriada e dois deles são os aqui referidos.

Pode-se tentar encontrar uma série de justificações para os dados resultantes da análise da sensibilidade, mas a que parece ser mais plausível é o facto dos indivíduos accionarem mecanismos de defesa quando se deparam com certas e determinadas questões, que têm um impacto emocional elevado (assumir ideação suicida não é tarefa fácil, pelo que negá-la é ou pode ser a solução); já para não falar do facto de haver inúmeros indivíduos que têm dificuldade em perceber o seu próprio estado de ânimo de forma correcta (falta de insight), pelo que tendem a dar as respostas que lhes parecem socialmente mais correctas ou aceitáveis, enviesando desta forma os dados. É portanto necessário considerar a existência de

eventuais respostas enviesadas. Para além disso, deve-se sempre ter em conta que “As descrições dos sintomas podem ser comprometidas por dificuldades de concentração, deficiência da memória, ou tendência para negar, dissimular ou desenvolver explicações para os sintomas” (American Psychiatric Association, 1996, p.331), características dos indivíduos deprimidos que dificultam o diagnóstico clínico da depressão, e poderão também enviesar os resultados do BDI-II, pois, apesar da população alvo ser não clínica, é nossa convicção que nela se encontram indivíduos deprimidos (com diferentes graus de severidade de depressão).

Ainda em relação à análise da sensibilidade dos itens do BDI-II, é de referir que, apesar desta ter demonstrado que quase metade dos seus itens não é sensível (ou apresenta graves problemas de sensibilidade), este facto pode ser relativizado. Assim, estes dados podem não ser de todo negativos e comprometedores da qualidade do BDI-II, na medida em que se verifica que os itens do IACLIDE (instrumento criado e validado para a população portuguesa, e considerado como dotado de qualidades psicométricas) não se revelam melhores que os do BDI-II, pelo contrário (ver Anexo R).

Apesar desta constatação, na análise do IACLIDE Vaz-Serra (1994) concluiu que todos os seus itens tinham uma boa capacidade de discriminação. Facto que não é de surpreender, pois o autor fez a sua análise com um método que tende a evidenciar a diferenciação entre grupos – o método dos grupos extremos (constituídos por 25% dos indivíduos com scores mais baixos e por 25% dos indivíduos com scores mais elevados de nível de depressão), que obviamente são diferentes, não fossem eles extremos. Para além disto, como refere Ribeiro (2008), este método de avaliação da capacidade de discriminação dos itens é um método utilizado para testes de capacidade/de realização máxima, e não para testes de personalidade como o IACLIDE e o BDI-II. De qualquer forma, é nossa convicção que caso se utilizasse este método, também se chegaria à conclusão que o BDI-II é constituído por itens com uma boa capacidade de discriminação ou com elevada sensibilidade.

Relativamente à validade, qualidade fundamental e indispensável de qualquer instrumento psicológico, pois diz se o instrumento mede aquilo que pretende medir, e não qualquer outra coisa parecida ou diferente, e diz quão bem ele o faz (Ribeiro, 2008), pode-se concluir que o BDI-II tem validade garantida e adequada, sendo portanto uma medida da variável latente que pretende medir – da depressão. Esta conclusão resulta da convergência de vários dados, nomeadamente os provenientes da análise da validade de critério

(concorrente) e da validade de construto (convergente, discriminante e factorial). Além disso, independentemente de não se ter efectuado uma análise da validade de conteúdo, pode-se concluir, tal como referido, que o BDI-II é também dotado deste tipo de validade, pois os seus itens são uma amostra representativa do domínio de conteúdo que pretende medir.

Antes de se passar à reflexão acerca dos vários dados que convergiam no sentido de demonstrar que o BDI-II tem validade garantida e adequada, é de referir que não foi efectuada nenhuma análise da validade preditiva do instrumento, o que pode ser visto como outra limitação do presente estudo. No entanto, acredita-se ter justificado conveniente e logicamente esta decisão. Sendo que o que difere a validade preditiva da validade concorrente é a relação temporal que se estabelece entre a recolha dos dados do teste e da medida de critério (neste caso o IACLIDE), de facto, a validade que está patente neste estudo é a concorrente. Quanto à razão que levou a que não fosse feita uma análise preditiva, esta prende-se fundamentalmente com duas questões relativas ao próprio objectivo da testagem: A validação concorrente é relevante para os testes que avaliam o estado actual da situação (como a depressão no BDI-II), e que não pretendem prever futuros estados (Anastasi & Urbina, 2000); para além disso, Cronbach (1996) refere mesmo que a validação concorrente é perfeitamente lógica quando se está a propor um teste como substituto de outro existente (que, no fundo, poderá ser uma consequência do presente estudo, caso se verifique que o BDI-II apresenta qualidades psicométricas superiores às do IACLIDE, entre outras).

Desta forma, através da análise da validade concorrente pode-se considerar que o BDI-II é dotado de validade de critério, e concluir que o instrumento mede de facto a depressão, pois existe concordância entre os seus resultados e os resultados da medida de critério (IACLIDE). Esta conclusão é apoiada por diferentes autores, nomeadamente Pasquali (2003) que refere que a correlação entre os dois instrumentos deverá ser elevada, de pelos menos 0,75. Já Bech (1992, in Vaz-Serra, 1994) diz que dois instrumentos têm validade concorrente, ou seja, medem a mesma dimensão/construto teórico quando apresentam uma correlação de pelo menos 0,70.

Apesar de Anastasi e Urbina (2000) referirem que as correlações entre os dois instrumentos devem ser moderadamente altas, mas não altas demais, porque se assim for o novo instrumento (BDI-II) poderá representar uma duplicação desnecessária do instrumento já existente (IACLIDE), é nossa convicção que tal não é o caso, ainda para mais porque se

considera que o BDI-II tem vantagens extras (como brevidade, simplicidade, facilidade de aplicação, e é menos dispendioso) em relação ao IACLIDE, facto que, segundo as autoras, já justifica correlações mais elevadas, como as encontradas no presente estudo ($r=0,91$).

No que se refere à validade de construto, e reportando aos dados resultantes das análises factoriais confirmatórias e exploratória, bem como da análise de correlação bivariada, comprova-se que o BDI-II é também dotado de validade de construto. Esta conclusão advém da convergência de dados de inúmeras análises e/ou indicadores, nomeadamente os referentes à validade factorial, à validade convergente e à validade discriminante.

Relativamente à validade factorial, e contrariando alguns estudos de validação e/ou adaptação, ou simplesmente de análise das qualidades psicométricas do BDI-II (e.g. Chang, 2005; Dozois, Dobson & Ahnberg, 1998; Ghassemzadeh, Mojtabai, Karamghadiri & Ebrahimkhani, 2005; Steer & Clark, 1997; Storch, Roberti & Roth, 2004), o modelo factorial de Beck et al. (1996) para a população estudantil não se adequa à população em estudo. A explicação para este facto é algo difícil; no entanto, é nossa convicção que são os factores culturais que melhor podem explicar esta situação. De facto, a forma como os indivíduos experienciam os vários sintomas de depressão varia de cultura para cultura, e portanto, é de prever que as dimensões de depressão encontradas num determinado país através do BDI-II sejam diferentes das encontradas noutra país, da mesma forma que as dimensões encontradas por Beck et al. (1996) para a população clínica são diferentes das encontradas para a população estudantil.

Estando verificada a qualidade dos dados do presente estudo, que é excelente para a definição de factores, constata-se que o BDI-II apresenta um modelo factorial constituído por quatro factores (que explicam cerca de 51% da variabilidade total do modelo), denominados de dimensão afectiva da depressão, dimensão somática da depressão, dimensão cognitivo-somática da depressão e dimensão cognitiva da depressão, em função dos seus itens constituintes e respectivos pesos factoriais. Na dimensão afectiva da depressão encontram-se itens como a tristeza, a perda de prazer, os pensamentos ou desejos suicidas, a perda de interesse e a perda de interesse sexual; na dimensão somática da depressão estão os itens agitação, alterações nos hábitos do sono, irritabilidade, alterações no apetite e cansaço ou fadiga; a dimensão cognitivo-somática é constituída pelo pessimismo, indecisão, perda de energia e dificuldades de concentração; e, finalmente, a dimensão cognitiva da depressão é

representada pelos itens referentes ao fracasso no passado, sentimentos de culpa e de punição, e autocrítica. Este modelo factorial adequa-se, como seria de esperar, à população em estudo, não fosse ele resultante dos dados provenientes desta população.

Para além da qualidade e da boa adequação do modelo factorial encontrado para a população académica do ISPA, todos os dados apontam para a comprovação de que o BDI-II tem uma boa validade convergente e discriminante. De facto, todos os critérios referidos por Maroco (2005) para comprovar as validades convergente e discriminante se verificam, apesar de alguns dados serem mais consistentes e abonatórios do que outros. Este facto leva a que alguns dados possam ser alvo de críticas, pelo que se passa de imediato à sua defesa.

No que se refere aos pesos factoriais dos itens nos factores, Maroco (2005) refere que estes devem ser iguais ou superiores a 0,50. No entanto, no presente estudo foram considerados os itens com pesos factoriais iguais ou superiores a 0,47, na medida em que, não só os quatro itens com pesos factoriais entre 0,47 e 0,50 se revelaram úteis para a constituição e “caracterização” dos factores, mas também porque a sua consideração nos factores não põe em causa a qualidade do modelo factorial, como provam as análises factoriais confirmatórias realizadas com e sem esses quatro itens (ver Anexos L e M).

Além disso, segundo Boyle e Harrison (1981, in Ribeiro, 2008) o modelo factorial só tem utilidade teórica se em cada factor se encontrarem pelo menos três itens que tenham um peso factorial no factor que não viole a relação entre o peso factorial mais elevado e o mais baixo, tendo em conta a fórmula referida pelos autores e que se encontra em anexo a propósito da análise factorial (ver Anexo A). Ora se estes quatro itens fossem excluídos do modelo factorial, o quarto factor (dimensão cognitiva da depressão), por exemplo, passaria a ter, não quatro, mas unicamente dois itens, o que para os referidos autores é inaceitável, pois põe em causa a utilidade teórica do modelo factorial.

A reforçar ainda mais a opção de incluir estes quatro itens no modelo factorial está o facto de que, se considerarmos a fórmula que Boyle e Harrison (1981, in Ribeiro, 2008) referem, o valor mínimo do peso factorial considerado aceitável para que um item defina um factor deveria ser de 0,45 para o primeiro factor, de 0,42 para o segundo, de 0,40 para o terceiro e de 0,45 para o quarto, valores ainda mais baixos do que o considerado no presente estudo. Ainda mais baixo do que estes valores foi o valor utilizado por Beck et al. (1996) no

seu modelo factorial, que foi de 0,35. Assim, é nossa convicção que este facto não põe em causa a validade convergente do BDI-II.

Também a consistência interna dos factores pode ser alvo de crítica pelo facto dos seus valores se encontrarem entre 0,70 e 0,80 nos dois primeiros factores, e entre 0,60 e 0,70 no terceiro e quarto factores, pois valores entre 0,70 e 0,80 são para Maroco (2005) considerados baixos, e valores inferiores a 0,70 são considerados inaceitáveis. No entanto, para Hill e Hill (2008), valores entre 0,70 e 0,80 são razoáveis e entre 0,60 e 0,70 são considerados fracos. Para além disso, apesar de Ribeiro (2008) referir que para que se considere a existência duma boa consistência interna o valor de alfa deverá ser superior a 0,80, o autor refere também que valores de alfa entre 0,60 e 0,80 são aceitáveis quando os instrumentos são constituídos por um número reduzido de itens, como é o caso dos factores em questão. De facto, sabendo-se que a consistência interna é fortemente afectada quer pela variabilidade da amostra, quer pelo comprimento do instrumento, não se pode esperar que factores constituídos por quatro e cinco itens tenham uma elevada consistência interna, como preconiza Maroco (2005), podendo então concluir-se que este facto também não põe em causa a validade convergente do BDI-II.

Relativamente à correlação que se estabelece entre os itens de diferentes factores, Maroco (2005) refere que estas devem ser baixas. Pela observação dos dados, é possível constatar que, apesar das correlações que se estabelecem entre os itens dum factor com os itens dos restantes factores tenderem a ser mais baixas do que as correlações que se estabelecem entre os itens dentro do mesmo factor, todos os itens se correlacionam entre si, independentemente dos factores, e existem mesmo itens que se correlacionam mais fortemente com itens de outros factores do que com os do próprio factor. Apesar destes dados não irem ao encontro do referido por Maroco (2005), não parece ser uma situação de todo inesperada, se se assumiu à priori que o modelo contém factores correlacionados, que também os itens se correlacionem entre si, independentemente do factor do qual fazem parte. Ainda para mais convém lembrar que, apesar do BDI-II reflectir um modelo factorial constituído por quatro factores, ele não deixa de ser um instrumento unidimensional, que avalia a dimensão depressão através da soma de todos os itens, que estão correlacionados entre si e que medem a mesma coisa, a depressão – isto porque, não existindo sintomas patognomónicos da depressão, ou seja, que só por si garantam a existência de depressão, se assume que esta pode ser avaliada a partir duma constelação de sintomas que em conjunto

revelam então a presença de depressão, mas que isoladamente nada representam (Bech, 1992, in Vaz-Serra, 1994).

O último dado que pode ser criticado, isto se se considerar, como Maroco (2005), que as correlações entre instrumentos que medem construtos teóricos diferentes devem ser nulas, é o facto de o BDI-II se correlacionar significativa e moderadamente com o STAI (que mede ansiedade). No entanto, esta situação não é de surpreender na medida em que, apesar dos instrumentos medirem construtos teóricos diferentes, ansiedade e depressão, estes construtos encontram-se habitualmente correlacionados na avaliação clínica da depressão. Como refere Montgomery (2000), é difícil encontrar um indivíduo deprimido que não apresente sintomas de ansiedade. Além disso, há evidências empíricas da sobreposição entre os dois estados psicopatológicos (Angst & Dobler-Mikola, 1985; Breier, Charney & Heninger, 1984; in Montgomery, 2000). Assim, é nossa convicção que não está em causa se o BDI-II avalia tanto depressão como ansiedade, mas sim se o BDI-II reflecte o facto, clinicamente comprovado, de a depressão ser frequentemente acompanhada de ansiedade, como referido por Rehm (1989, in Vaz-Serra, 1994) a propósito da primeira versão deste instrumento.

Além disso, se se tiver em consideração a posição de Herdman, Fox-Rushby e Badia (1998, in Ribeiro, 2008) em relação à validade convergente e discriminante dos instrumentos psicológicos, a validade discriminante do BDI-II não pode ser posta em causa, mesmo perante a correlação significativa e moderada que estabelece com o STAI. Segundo os referidos autores, a validade convergente refere-se à extensão em que a correlação do instrumento com instrumentos que medem o mesmo construto é maior do que a correlação com os que medem construtos diferentes” (Herdman, Fox-Rushby & Badia, 1998, in Ribeiro, 2008, p.96), o que acontece no presente estudo; já a validade discriminante refere-se “...à extensão em que a correlação do instrumento com instrumentos que medem diferentes construtos é menor do que a correlação com os que medem o mesmo construto” (Herdman, Fox-Rushby & Badia, 1998, in Ribeiro, 2008, p.96), o que também se verifica. Fazendo o mesmo tipo de raciocínio que Beck et al. (1996) fizeram a propósito da validade discriminante, o facto do BDI-II se correlacionar fortemente com o IACLIDE ($r=0,91$) e moderadamente com o STAI ($r=0,73$), juntamente com o facto destes últimos terem uma correlação superior ($r=0,78$) à encontrada entre o BDI-II e o STAI, pode evidenciar uma validade discriminante entre os construtos.

Para além da convergência de todos os dados no sentido de comprovar a validade do BDI-II, também os dados sócio-demográficos reportados são abonatórios para o instrumento, na medida em que mostram que ele não é sensível a diferenças de género, idade, curso e ano frequentado, ou seja, mede de forma idêntica todos os indivíduos amostrados.

Desta forma, pode-se concluir que o BDI-II é um instrumento dotado de qualidades psicométricas, com fiabilidade e validade adequadas, pelo que se considera que pode ser usado (desde que usado adequadamente) com confiança para avaliar a depressão. Provou-se neste estudo que o BDI-II mede de facto a depressão (validade) e fá-lo de forma repetida e consistente (fiabilidade). No entanto, é importante ter sempre em mente que “...os méritos individuais de cada prova (identificados através das suas propriedades psicométricas) não asseguram só por si uma avaliação com qualidade” (Gonçalves et al., 2006, p.16).

Algumas reticências são colocadas no que diz respeito à sensibilidade (capacidade de discriminação) de alguns itens, pelo que se adverte os usuários do BDI-II que tenham em atenção que, em algumas situações, mesmo indivíduos deprimidos poderão dar respostas típicas de indivíduos não deprimidos (isto é, com uma cotação mais baixa em termos de severidade dos sintomas), o que pode mascarar os dados, e logo as conclusões que deles advêm. Esta situação é particularmente importante em relação a alguns sintomas, nomeadamente à ideação suicida, na medida em que é do conhecimento geral que o risco e a taxa de suicídio são particularmente elevados em indivíduos deprimidos – entre um terço e metade dos suicídios e tentativas de suicídio podem ser atribuídas à depressão e o seu risco de morte por suicídio é 30 vezes superior ao da população geral (Braconnier, 2007) –, e logo torna-se necessária uma avaliação cuidadosa a este nível.

Além disso, tem de ser feita uma outra advertência aos usuários deste instrumento, que resulta da análise da estrutura factorial do BDI-II. O presente estudo permitiu evidenciar que a versão traduzida para a língua portuguesa do BDI-II não apresenta o mesmo modelo factorial que o instrumento original reflectiu para uma população idêntica à aqui usada. Esta é mais uma prova de que não se devem usar critérios oriundos de outros países, sem se ter a certeza se eles se adequam ou não à população que se está a estudar. Resta saber se o modelo factorial aqui encontrado se adequa ou não a outras populações, académicas ou clínicas, o que remete para futuros estudos. Ainda a respeito da estrutura factorial do BDI-II, considera-se, assim como Gonçalves e Fagulha (2006) a respeito da CES-D, que, apesar de poder ser

interessante analisar os resultados por factores para fins de investigação, o BDI-II avalia uma única dimensão, a depressão, como referido anteriormente, o que significa que quanto mais elevado for o score obtido, mais intensa é a sintomatologia depressiva experienciada.

Finalmente, é de referir que, para além das já mencionadas limitações do presente estudo, existe uma outra (à parte de outras tantas que certamente existem, mas que não foi possível identificar): Não se procedeu a estudos de normatização dos resultados, que são “considerados como a etapa final dum processo de aferição” (Simões, 1994, in Canavarro, 2007, p.326). Esta é uma limitação subsequente da segunda limitação referida, ou seja, parece advir do facto de não se ter trabalhado com uma população clínica, na medida em que é nossa convicção que é impossível chegar a esta “etapa final da aferição”, na qual se estabelecem normas para a população, sem se realizarem estudos com uma população clínica. No entanto, chegar a esta etapa final sai fora do âmbito do presente trabalho, que se propunha analisar as qualidades/características psicométricas do BDI-II. Apesar de um processo de validação e/ou adaptação do BDI-II (que abranja uma população mais vasta, nomeadamente a clínica) ultrapassar os objectivos do presente estudo, com a experiência adquirida no decorrer do mesmo, este poderá constituir-se como um futuro projecto a realizar. Ainda assim, acredita-se que o presente estudo poderá ser uma referência útil para os investigadores e usuários do BDI-II que trabalhem com populações semelhantes à que foi aqui examinada.

Apesar de todas as limitações, todas as dificuldades, todos os obstáculos, alguns mesmo intransponíveis, crê-se que foi dado o primeiro passo para a validação e adaptação deste instrumento para a população portuguesa, e logo mais um passo para que esteja disponível no nosso país mais um instrumento, dentro do pequeno leque de instrumentos nacionais e estrangeiros, validados para a população portuguesa para avaliar a depressão, que nos permita andar a par e passo com os estudos internacionais, usando medidas universais (que permitam a comparação de dados). Instrumento este que é, comparado com os disponíveis, o mais recente e que, comparado com o IACLIDE, por exemplo, se apresenta com características a seu favor, como brevidade, simplicidade e facilidade de aplicação, económico e pouco invasivo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, 96, 358-372.
- Abramson, L.Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87, 49-74.
- American Psychiatric Association (1996). [*Manual de diagnóstico e estatística das perturbações mentais*] (4ª ed.). (A. Baptista, A. P. Vieira, L. C. Pestana, P. Casquinha, P. Levy, & P. Varandas, Trans.). Lisboa: Climepsi Editores.
- American Psychiatric Association (2002). [*Manual de diagnóstico e estatística das perturbações mentais*] (4ª ed. rev.). (J. N. de Almeida, Trans.). Lisboa: Climepsi Editores.
- Anastasi, A., & Urbina, S. (2000). [*Testagem psicológica*] (7ª ed.). (M. A. V. Veronese, Trans.). Porto Alegre: Artmed Editora.
- Arbuckle, J. L., & Wothke, W. (1995-1999). *Amos 4.0 user's guide*. Chicago: SmallWaters Corporation.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: The Guilford Press.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II: Manual*. San Antonio: The Psychological Corporation.
- Braconnier, A. (2007). [*Manual de psicopatologia*]. (F. Fonseca, Trans.). Lisboa: Climepsi Editores.
- Canavarro, M. C. (2007). Inventário de sintomas psicopatológicos (BSI): Uma revisão crítica dos estudos realizados em Portugal. In M. R. Simões, C. Machado, M. M. Gonçalves, & L. S. Almeida (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa* (Vol. 3, pp. 305-331). Coimbra: Quarteto.

- Champion, L. A., & Power, M. J. (1995). Social and cognitive approaches to depression: Towards a new synthesis. *British Journal of Clinical Psychology*, 34, 485-503.
- Chang, Hsiaowen (2005). Dimensions of the chinese Beck depression inventory – II in a university sample. *Individual Differences Research*, 3 (3), 193-199.
- Cláudio, V. (2004). *Da trama das minhas memórias o fio que tece a depressão: Esquecimento dirigido e memória autobiográfica na depressão major*. Lisboa: ISPA.
- Cronbach, L. J. (1996). [*Fundamentos da testagem psicológica*] (5ª ed.). (C. A. S., Neto, & M. A. V., Veronese, Trans.). Porto Alegre: Editora Artes Médicas Sul, Lda.
- Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck depression inventory – II. *Psychological Assessment*, 10 (2), 83-89.
- Ghassemzadeh, H., Mojtabai, R., Karamghadiri, N., & Ebrahimkhani, N. (2005). Psychometric properties of a persian-language version of the Beck depression inventory – second edition: BDI-II-Persian. *Depression and Anxiety*, 21, 185-192.
- Gonçalves, B., & Fagulha, T. (2006). Escala de depressão do centro de estudos epidemiológicos (CES-D). In M. M. Gonçalves, M. R. Simões, L. S. Almeida, & C. Machado (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa* (2ª ed. rev., Vol. 1, pp. 37-44). Coimbra: Quarteto.
- Gonçalves, M., Simões, M., Almeida, L., & Machado, C. (2006). Introdução: Contributos para a divulgação de instrumentos de avaliação psicológica. In M. M. Gonçalves, M. R. Simões, L. S. Almeida, & C. Machado (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa* (2ª ed. rev., Vol. 1, pp. 13-19). Coimbra: Quarteto.
- Gotlib, I. H., & Hammen, C. L. (2009). Introduction. In I. H. Gotlib, & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 1-2). New York: The Guilford Press.
- Graziani, P. (2005). [*Ansiedade e perturbações da ansiedade*]. (E. Pestana, Trans.). Lisboa: Climepsi Editores.
- Hammen, C. (2006). Stress generation in depression: Reflections on origins, research, and future directions. *Journal of Clinical Psychology*, 62 (9), 1065-1082.
- Hammen, C., & Mazure, C. M. (2003). Symposium: Understanding the interaction of stress and gender in the prediction of major depression – Introduction. *Archives of Women's Mental Health*, 6, 3-4.
- Hill, M. M., & Hill, A. (2008). *Investigação por questionário* (2ª ed. rev.). Lisboa: Edições Sílabo.
- Joormann, J. (2009). Cognitive aspects of depression. In I. H. Gotlib, & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 298-321). New York: The Guilford Press.

- Kessler, R. C., & Wang, P. S. (2009). Epidemiology of depression. In I. H. Gotlib, & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 5-22). New York: The Guilford Press.
- Maroco, J. (2005). *Validade, fiabilidade e sensibilidade*. Manuscrito não publicado em formato PowerPoint.
- Maroco, J. (2003). *Análise estatística com utilização do SPSS* (2^a ed.). Lisboa: Edições Sílabo.
- Matos, A. C. (2001a). Nota introdutória. In *Depressão: Episódios de um percurso em busca do seu sentido* (pp. XV-XVIII). Lisboa: Climepsi Editores.
- Matos, A. C. (2001b). Narcisismo e depressão. In *Depressão: Episódios de um percurso em busca do seu sentido* (pp. 494-501). Lisboa: Climepsi Editores.
- Montgomery, S. A. (2000). [*Ansiedade e depressão*] (2^a ed.). (A. P. Carneiro, Trans.). Lisboa: Climepsi Editores.
- Murteira, B. J. F. (1993). *Análise exploratória de dados: Estatística descritiva*. Alfragide: McGraw-Hill de Portugal.
- Nezu, A. M., Nezu, C. M., Friedman, J., & Lee, M. (2009). Assessment of depression. In I. H. Gotlib, & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 44-68). New York: The Guilford Press.
- Pais-Ribeiro, J., Silva, I., Ferreira, T., Martins, A., Meneses, R., & Baltar, M. (2007). Validation study of a portuguese version of the hospital anxiety and depression scale. *Psychology, Health & Medicine*, 12 (2), 225-237.
- Pasquali, L. (2003). *Psicometria: Teoria dos testes na psicologia e na educação*. Petrópolis, RJ: Editora Vozes.
- Power, M. J., & Champion, L. A. (1988). Perspectivas cognitivas da depressão: Crítica teórica. *Análise Psicológica*, 2 (6), 183-196.
- Power, M., & Dalgleish, T. (2008). *Cognition and emotion: From order to disorder* (2nd ed.). New York: Psychology Press.
- Reis, E. (1996). *Estatística descritiva* (3^a ed.). Lisboa: Edições Sílabo.
- Ribeiro, J. L. P. (1999). *Investigação e avaliação em psicologia e saúde*. Lisboa: Climepsi Editores.
- Ribeiro, J. L. P. (2008). *Metodologia de investigação em psicologia e saúde* (2^a ed.). Porto: Legis Editora.
- Roudinesco, E., & Plon, M. (2000). [*Dicionário de psicanálise*]. (V. Ribeiro, & L. Magalhães, Trans.). Mem Martins: Editorial Inquérito.

- Silva, D. R., Silva, J. A., Rodrigues, A., & Luís, R. (1999/2000). Estudo de adaptação e estabelecimento de normas do inventário de estado-traço de ansiedade (STAI) – forma Y de Spielberger, para a população militar portuguesa. *Revista de Psicologia Militar*, 12, 8-26.
- Silva, D. R. (2006). O inventário de estado-traço de ansiedade (STAI). In M. M. Gonçalves, M. R. Simões, L. S. Almeida, & C. Machado (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa* (2ª ed. rev., Vol. 1, pp. 45-60). Coimbra: Quarteto.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory (form Y) (“self-evaluation questionnaire”)*. Palo Alto: Consulting Psychologists Press, Inc.
- Steer, R. A., & Clark, D. A. (1997, October). Psychometric characteristics of the Beck depression inventory – II with college students. *Measurement and Evaluation in Counseling and Development*, 30 (3), 128-136. Retrieved February 23, 2006 from EBSCOHost Academic Search Elite database.
- Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck depression inventory – second edition in a sample of college students. *Depression and Anxiety*, 19, 187-189.
- Vaz-Serra, A. (1994). *Inventário de avaliação clínica da depressão*. Coimbra: Edição Psiquiatria Clínica.

ANEXOS

Anexo A: Metodologias Estatísticas que Permitem Avaliar a Validade de Construto

Tal como referido, existe uma série de técnicas que permitem analisar a validade de construto. Pasquali (2003) refere as que apresentamos de seguida:

- O erro de estimação da TCT;
- A análise da representação comportamental do construto, que poderá ser efectuada através da análise da consistência interna e da análise factorial;
- A análise por hipótese, na qual se poderão utilizar várias metodologias, como a validação convergente-discriminante, a idade, outros testes com o mesmo construto teórico e a experimentação;
- E a curva de informação da TRI.

➤ O **erro de estimação** da TCT diz-nos o quanto o score que o indivíduo obteve no teste se afasta do score verdadeiro – a validade é vista como a precisão com que o teste prediz o score verdadeiro. Limitação: Existe um engano ao pressupor que o score verdadeiro é a medida daquilo que o teste pretende medir, pois, de facto, o score verdadeiro engloba em parte aquilo que o teste pretende medir, mas engloba também todas as características dos itens que fazem parte do teste e que não se relacionam com o que o teste pretende medir.

➤ A análise da **consistência interna** dá-nos essencialmente uma medida da homogeneidade – “uma vez que ele ajuda a caracterizar o domínio de comportamento ou traço amostrado pelos testes, o grau de homogeneidade de um teste tem certa relevância para a sua validade de construto” (Anastasi & Urbina, 2000, p.119). Limitação: Apesar de facultar uma medida de homogeneidade do teste, a consistência interna pouco nos diz acerca do que o teste mede. Além disso, como Pasquali (2003, p.172) refere:

A análise da consistência interna do teste implica no cálculo das correlações de cada item individualmente com o restante teste. Esta análise apresenta um problema lógico que se situa no score total. Na verdade, o score total é o critério contra o qual cada item é avaliado; mas acontece que os itens são os que vão constituir o score total, antes mesmo de se saber se eles são itens válidos e somáveis (unidimensionais, isto é, que estão medindo um e o mesmo traço latente). O score total constitui, assim, uma dificuldade, dado que ele somente faz sentido se o teste já é à priori homogéneo. (...). Além disso, a consistência interna implica que os itens estejam intercorrelacionados, isto é, que as correlações entre eles mesmos sejam elevadas. Entretanto, as intercorrelações entre os itens não são uma demonstração de que estes estejam medindo um e mesmo construto.

➤ A **análise factorial** pretende verificar quantos construtos comuns ou factores são necessários para explicar as intercorrelações dos itens, sendo portanto o seu objectivo simplificar a descrição do comportamento, como se confirma no seu postulado que diz que “...um número menor de traços latentes (variáveis-fonte) é suficiente para explicar um número maior de variáveis observadas (itens)” (Pasquali, 2003, p.173). As variáveis (X) explicam-se através de um factor comum a todas elas (F), bem como através de um factor específico a cada uma delas (U); assim, as variáveis podem ser descritas da seguinte forma: $X_n = a_n F + d_n U_n$, sendo que a_n é o peso factorial da variável no factor comum (F), que representa a percentagem de relação que a variável tem com o factor, ou seja, a percentagem da sua representação com o traço latente. Desta forma, a validade de construto de um teste determina-se através da “...grandeza das cargas factoriais (que são correlações que vão de -1 a +1) das variáveis no factor, sendo aquelas a representação comportamental deste factor, que, por sua vez, é o traço latente para o qual elas foram inicialmente elaboradas como representação empírica” (Pasquali, 2003, 174). Limitação: Para além de se fundamentar exclusivamente em equações lineares entre variáveis e factores, o que, já de si, é algo difícil de encontrar em ciências sociais, existe também na análise factorial uma grande limitação que se prende com a rotação dos eixos, na medida em que não há nenhum critério objectivo na interpretação dos factores – o que na prática pode levar à descoberta de qualquer factor que se deseje, pois esta depende duma interpretação subjectiva (semântica) dos factores, podendo-se então considerar as soluções um pouco arbitrárias.

Segundo Ribeiro (2008), a análise factorial é um procedimento estatístico muito utilizado em psicologia, quando se quer avaliar a estrutura dos testes psicológicos, pois é o “procedimento de referência para o estudo da validade de construto” (p.131). Admite-se que a análise factorial é de dois tipos: Exploratória e confirmatória. A análise factorial exploratória

...constitui um método indutivo que visa descobrir um conjunto óptimo de variáveis latentes. O seu papel enquanto procedimento exploratório pode ser usado sem consideração por hipóteses prévias. Por outro lado, a análise factorial confirmatória é um método hipotético dedutivo, que visa testar hipóteses acerca das relações entre variáveis que constituem factores comuns, cujo número e interpretação é pré-existente à realização da análise factorial confirmatória, e que a guia (Mulaik, 1988; Nunnally, 1978; in Ribeiro, 2008, p.131).

Relativamente ao número de indivíduos que uma amostra deverá ter para que se possa realizar a análise factorial, Bartlett, Kotrlik e Higgins (2001, in Ribeiro, 2008) referem que deverão existir pelo menos 100 indivíduos ou 5 indivíduos para cada variável – por exemplo, no caso do presente estudo, como o BDI-II é constituído por 21 itens, a amostra deveria ser constituída por pelo menos 105 indivíduos.

Em relação ao número de factores que podem e/ou devem ser retirados existem vários critérios que podem ser adoptados, uns práticos outros teóricos (Ribeiro, 2008). Há dois critérios práticos muito utilizados: o critério Kaiser (no qual são retirados os factores com valores próprios de um ou superiores a um) e o scree (no qual, na análise gráfica, se considera que o limite para o número de factores se verifica no cotovelo do gráfico, ou seja, onde se encontra uma inclinação acentuada). No critério teórico assume-se que o número de factores a reter não é mais do que o número de factores que o modelo teórico propõe à partida. Existem, contudo, outros critérios. Tabachnick e Fidell (1996, in Ribeiro, 2008) propõem que o número de factores a reter deverá expressar o número de processos teóricos que subjazem o domínio de comportamento em estudo, e que uma maneira rápida de decidir o número de factores a reter é dividindo o número de variáveis em questão por três ou por cinco, sendo que este critério é aceitável para amostras grandes e quando o número de variáveis é inferior a 40 – por exemplo, no caso do BDI-II, que é constituído por 21 itens, a sua solução factorial deverá ter quatro a sete factores. Para Boyle e Harrison (1981, in Ribeiro, 2008, p.132)

...a solução factorial só é útil do ponto de vista teórico se inclui, para cada factor, pelo menos 3 itens que possuam uma carga que respeite a relação entre o valor da carga factorial mais elevada com a menor, segundo a seguinte fórmula: $C = \sqrt{H/2}$, em que H é o valor do item com a carga factorial mais elevada e C o valor mínimo de carga factorial considerado aceitável para um item que define o factor em jogo.

➤ A **análise por hipótese** diz-nos basicamente qual a capacidade de um teste de discriminação ou de predição de um critério externo a ele próprio. Limitação: Dificuldade “...na definição inequívoca do critério a ser utilizado como representante da manifestação do traço” (Pasquali, 2003, p.181).

❖ Na validação **convergente-divergente** a validade de construto baseia-se na demonstração simultânea de que o teste se correlaciona significativamente com outras variáveis com as quais ele deve correlacionar-se teoricamente (validade convergente)

e de que o teste não se correlaciona significativamente com variáveis com as quais ele deverá teoricamente diferir (validade discriminante).

❖ No processo de validação do construto através da **idade** pretende-se provar que o teste, que mede um traço que altera com a idade, tem capacidade de discriminar grupos etários diferenciados. Limitação: Este método não é aplicável a testes que medem traços não dependentes de mudanças desenvolvimentais (a nível cognitivo e afectivo), e, além disso, uma vez que o desenvolvimento de diferentes características de comportamento pode ser influenciado pela cultura, não se pode pressupor que este seja um critério universal.

❖ A **correlação com outros testes** que medem o mesmo construto teórico (traço latente) também pode ser utilizada como demonstração da validade de construto, na medida em que se um teste mede um determinado construto teórico e um novo teste se correlaciona significativamente com o anterior, então pode-se considerar que o novo teste mede o mesmo construto teórico. Segundo Anastasi e Urbina (2000, p.118) “...essas correlações devem ser moderadamente altas, mas não altas demais. Se o novo teste apresenta uma correlação muito alta com um teste já existente, sem vantagens extras como brevidade ou facilidade de aplicação, o novo teste representa uma duplicação desnecessária”. Limitação: “...normalmente um teste de um traço qualquer não se apresenta com tal pureza a se poder afirmar que ele mede exclusivamente o tal traço” (Pasquali, 2003, p.179).

❖ Através da **intervenção experimental** pretende-se provar que o teste tem capacidade de discriminar grupos-critério produzidos experimentalmente no que diz respeito ao traço medido pelo teste – se se puder garantir que as manipulações efectuadas nos grupos-critério se reflectem exclusivamente no traço medido, então a validade de construto fica demonstrada. Limitação: Pode acontecer que a manipulação de uma variável tenha repercussões noutras variáveis, principalmente quando elas interagem, fazendo com que seja difícil perceber se o teste discrimina os grupos-critério exclusivamente no traço medido ou não.

➤ A **curva de informação da TRI** mostra-nos para que faixas do traço latente o teste é particularmente válido e para que faixas ele não é válido.

A respeito da validade de construto, Cronbach (1996) refere algumas técnicas de pesquisa pertinentes, tais como: Inspeccionar itens; verificar correlações internas; verificar a estabilidade de resultados; administrar o teste a indivíduos que “pensam em voz alta”; tentar melhorar os resultados; estabelecer a correlação com critérios práticos e com outros testes; e fazer análises de diferenças de grupos. Cronbach refere também que na busca pela validade de construto, procura-se basicamente confirmar a convergência de indicadores que se correlacionam, assim como a divergência de indicadores que não se correlacionam.

Maroco (2005) diz que a validade de construto é determinada pela validade convergente e pela validade discriminante, e avalia-se através da análise factorial e da análise de correlação bivariada. Para o autor, existe uma série de critérios que terão de ser verificados para que exista validade convergente e discriminante¹¹, assim:

- Para a validade convergente: KMO (as correlações entre os itens permitem definir construtos ou factores); pesos factoriais dos itens nos factores de pelo menos 0,50; correlações inter-itens e/ou consistência interna dos itens que constituem o factor superior a 0,80; e correlação com outros construtos teoricamente equivalentes superior a 0,80.
- Para a validade discriminante: Pesos factoriais inferiores a 0,50 em factores que medem “conceitos” diferentes; correlações inter-itens de outros factores baixas; e correlações nulas com factores teoricamente diferentes.

Anastasi e Urbina (2000) por seu turno, para além de referirem a maioria das técnicas citadas por Pasquali, a saber, a técnica de validação através da diferenciação por idades (mudanças desenvolvimentais), a técnica da correlação com outros testes, a técnica da análise factorial, a técnica da consistência interna, a técnica da validação convergente e discriminante e a técnica de validação através de intervenções experimentais, mencionam também a técnica da modelagem de equação estrutural e referem as contribuições que a psicologia cognitiva tem dado a este respeito¹².

Note-se que, apesar dos diferentes autores usarem alguns termos distintos, existe uma grande equivalência entre as várias técnicas referidas.

¹¹ Estas medidas são usadas quando se realiza análise factorial exploratória – como no presente estudo.

¹² Para mais informações sobre estas duas técnicas aconselhamos a leitura de Anastasi e Urbina (2000, Cap.5).

Anexo B: Metodologias Estatísticas que Permitem Estimar a Fiabilidade

É possível através da **técnica da correlação**, e tendo como base cada um dos delineamentos experimentais, obter um coeficiente de fiabilidade, que para cada delineamento apresenta um nome diferente. Mas antes de se passar para cada um dos delineamentos convém referir que um coeficiente de correlação indica "...o grau de correspondência, ou relacionamento, entre dois conjuntos de scores" (Anastasi & Urbina, 2000, p.85) e que

...quanto mais próximo de 1 positivo seja este coeficiente, mais próxima de 100% será a correspondência directa entre as duas classificações; quanto mais próxima de -1 (negativo) a correlação, tanto mais certeza temos que os primeiros classificados na primeira aplicação do teste, serão os últimos na segunda e vice-versa. Um coeficiente de 0 indicaria que não haveria nenhuma relação entre o que aconteceu na primeira aplicação e na segunda; isto é, a partir da classificação dos sujeitos obtida na primeira aplicação, não dá para prever nada sobre como será esta classificação na segunda aplicação (Pasquali, 2003, p.197).

Ou seja, a análise correlacional não permite fazer inferências acerca da causalidade do tipo "a" provoca "b"; unicamente descreve a variância partilhada, isto é, só permite fazer inferências acerca da forma como as variáveis variam em conjunto, indicando-nos o grau de associação entre elas – como uma variável muda quando a outra muda ou qual a proporção de variância de uma variável que é prevista por outra variável à qual se associa (Ribeiro, 2008).

Para facilitar a compreensão deste tema, apresenta-se na página seguinte uma tabela que resume o que de mais importante há a referir acerca da estimação do coeficiente de fiabilidade para cada um dos delineamentos, tabela esta que foi adaptada de Pasquali (2003).

É de referir que as desvantagens de cada um dos delineamentos não os tornam inválidos quanto à estimação do coeficiente de fiabilidade, apenas alertam que se deve garantir que não ocorreram, ou que, tendo ocorrido, o seu impacto não teve um efeito significativo nos dados.

No que se refere à fiabilidade de teste-reteste, o intervalo temporal entre aplicações não é fixo, mas deverá ser de pelo menos uma semana ou um mês (Hill & Hill, 2008). Em relação à fiabilidade do tipo split-half, os autores referem que este método de estimativa da fiabilidade apresenta dois problemas (a redução do nº de itens que aumenta o erro amostral, o que tende a ser corrigido pela fórmula de Spearman-Brown; e a forma arbitrária de partição dos itens que reduz a confiança neste método), problemas estes que são facilmente ultrapassados através do método de estimativa da fiabilidade interna alfa de Cronbach.

TÉCNICA	PROCEDIMENTO	COEFICIENTE	VANTAGENS	DESVANTAGENS
DUAS METADES OU SPLIT-HALF	<ul style="list-style-type: none"> ▪ Aplica-se 1 mesmo teste a 1 amostra representativa; ▪ Divide-se o teste em 2 partes/metades equivalentes; ▪ Calcula-se a correlação entre as 2 metades, aplicando a correcção de Spearman-Brown. 	Consistência interna do teste	<ul style="list-style-type: none"> ▪ Exige uma única aplicação, o que permite evitar possíveis variáveis temporais. 	<ul style="list-style-type: none"> ▪ É difícil garantir a equivalência das 2 metades (igual nº de itens, nível de dificuldade e discriminação, assim como igual índice de consistência interna); ▪ É difícil garantir que as respostas à 1ª metade não afectam as respostas à 2ª metade; ▪ É difícil controlar o efeito da fadiga quando se responde à 2ª metade, que normalmente contém os itens mais difíceis.
FORMAS PARALELAS	<ul style="list-style-type: none"> ▪ Aplica-se 2 formas paralelas/equivalentes de 1 teste a 1 amostra que seja representativa; ▪ Calcula-se a correlação entre as 2 formas paralelas. 	Equivalência	<ul style="list-style-type: none"> ▪ Vai de encontro ao conceito de fiabilidade; ▪ Exige uma única aplicação. 	<ul style="list-style-type: none"> ▪ É difícil garantir 2 formas perfeitamente paralelas/equivalentes, isto é, 2 testes que meçam o mesmo traço latente/construto teórico com itens diferentes; ▪ É difícil garantir que a aplicação da 1ª forma não influencia a aplicação da 2ª forma.
TESTE-RETESTE	<ul style="list-style-type: none"> ▪ Aplica-se 1 mesmo teste a 1 amostra representativa em 2 situações diferentes; ▪ Calcula-se a correlação entre os resultados obtidos nas 2 situações. 	Estabilidade ou constância	<ul style="list-style-type: none"> ▪ Garante-se a equivalência (paralelismo) pois trata-se do mesmo teste 	<ul style="list-style-type: none"> ▪ É difícil definir qual o intervalo de tempo ideal entre as aplicações; ▪ É difícil controlar eventos que ocorreram entre as aplicações e seus efeitos aquando da 2ª aplicação; ▪ É difícil controlar os efeitos que a 1ª aplicação terá nas respostas à 2ª aplicação – memória e má atitude.

Tabela 3 – Coeficientes de fiabilidade para cada um dos delineamentos

Além disto, gostaríamos de acrescentar que no que toca à fiabilidade de formas paralelas, um teste é considerado preciso quando o coeficiente de correlação se aproxima de um, não sendo suficiente que o coeficiente seja estatisticamente significativo, isto é, não é suficiente que os testes estejam relacionados, os seus resultados devem ser idênticos – sendo que,

...coeficientes em torno de 0,90 ou maiores são normalmente os esperados para expressar a fidedignidade de um teste. Coeficientes em torno de 0,80 usualmente são considerados razoáveis, enquanto coeficientes de precisão abaixo de 0,70 não são normalmente suficientes como demonstração de uma fidedignidade aceitável para um teste (Pasquali, 2003 p.200).

Através das **técnicas alfa** é possível estimar um coeficiente de fiabilidade de um teste que tenha sido aplicado a uma amostra representativa de indivíduos numa mesma e única ocasião. Esta estimativa baseia-se na análise da consistência interna dos itens, tal como referido anteriormente, isto é, baseia-se na análise da congruência que cada um dos itens do teste tem com os restantes itens do teste. A consistência interna é um indicador da precisão/fiabilidade de um teste, na medida em que sabendo a variabilidade dos itens, sabemos a sua precisão, tal como referido por Pasquali (2003, p.205) “...quanto menos variabilidade um mesmo item produz numa amostra de sujeitos, menos erro ele provoca. Assim, quanto menor a variância do item, mais preciso é o item”. O coeficiente mais usado para este fim é o coeficiente alfa de Cronbach (α), e é nele que nos vamos deter. O que a fórmula de Cronbach diz é que

...se todos os itens variarem do mesmo jeito, isto é, se não houver variância entre os itens individualmente, o alfa será igual a 1; quer dizer que os itens serão totalmente homogêneos, de facto idênticos, produzindo exactamente a mesma variância. Como tal evento não é de se esperar, o alfa dará o tanto de congruência ou covariação que os itens têm dentro do teste. O coeficiente alfa vai de 0 a 1, o 0 indicando ausência total de consistência interna dos itens e o 1, presença de consistência de 100% (Pasquali, 2003, p.204).

Segundo Maroco (2005) “o α pode interpretar-se como a % da variância que a escala (teste, instrumento) observada explicaria na escala real (hipotética) composta de todos os itens possíveis desse domínio”. O autor refere ainda que o alfa de Cronbach não corresponde a uma “estimativa de quão bem um conjunto de itens medem um construto unidimensional”, e que o alfa de Cronbach é considerado a medida universal de consistência interna e, de uma maneira geral, se assume que a consistência interna é elevada quando o alfa de Cronbach é

igual ou superior a 0,90; boa quando se encontra entre 0,80 e 0,90; baixa quando está entre 0,70 e 0,80; e inaceitável quando é inferior a 0,70.

Para Ribeiro (2008), para que se considere a existência duma boa consistência interna o alfa de Cronbach deverá ser superior a 0,80, no entanto, aceitam-se valores de alfa superiores a 0,60, sendo que valores de alfa baixos (como entre 0,60 e 0,80) são justificados quando os instrumentos são constituídos por um número de itens muito baixo.

Ainda a respeito do alfa de Cronbach, Rotter (1990, p.491, in Ribeiro, 2008, p.114) refere que “podemos ter alfas muito elevados se fizermos a mesma pergunta inúmeras vezes, mas as limitações preditivas de tal teste são óbvias (...) e um teste com um alfa de 0,90 é suspeito na medida em que pode estar a medir o estilo de resposta em vez de fornecer uma amostra representativa do comportamento e de situações que se desejam prever, ou pode simplesmente consistir em itens redundantes”. Sabemos também que o valor do alfa de Cronbach aumenta com o número de itens do instrumento e com correlações mais elevadas entre os itens (Hill & Hill, 2008).

Regra geral, os vários coeficientes de fiabilidade têm valores diferentes, sendo que o de maior valor é o alfa de Cronbach, seguido do de formas paralelas/equivalentes, logo a seguir o de split-half e, finalmente, o coeficiente de menor valor é o de estabilidade temporal, coeficiente que está mais sujeito a fontes de erro de medição (Hill & Hill, 2008). Segundo estes autores coeficientes de fiabilidade superiores a 0,90 são excelentes, entre 0,80 e 0,90 são bons, entre 0,70 e 0,80 razoáveis, entre 0,60 e 0,70 fracos e inferiores a 0,60 são inaceitáveis.

Anexo C: Metodologias Estatísticas que Permitem Analisar a Qualidade dos Itens

Unidimensionalidade

A unidimensionalidade pode ser analisada através de vários índices, índices esses que foram categorizados por Hattie (1985, in Pasquali, 2003) da seguinte forma: 1) baseados em padrões de resposta; 2) baseados na fiabilidade; 3) baseados na análise de componentes principais; 4) baseados na análise factorial e 5) baseados na TRI. Pasquali (2003) refere ainda outro método, baseado tanto na análise factorial como na TRI, que considera ser o mais útil para a análise da unidimensionalidade – a análise factorial full information. Sendo que segundo o autor (Pasquali, 2003) os três primeiros métodos têm muitas críticas, e na medida em que para o presente trabalho não era exequível trabalhar com a análise da TRI e com a análise factorial full information, vamo-nos deter no índice baseado na análise factorial.

Através da análise factorial¹³ simplifica-se a descrição do comportamento, de tal forma que os itens possam ser reduzidos a uma única dimensão ou factor, com a qual estão relacionados – esta é a situação ideal para que se considere existir unidimensionalidade. A relação que se estabelece entre eles (itens e factores) é determinada pelo peso factorial que os itens têm nos factores. Assim, considera-se que os itens são unidimensionais quando têm um peso factorial alto no factor (de pelo menos 0,30) e que os itens são “estranhos” quando têm um peso factorial perto de 0, pois não medem o mesmo que os outros itens (Pasquali, 2003). No entanto, segundo Maroco (2005) os itens deverão ter um peso factorial no factor de pelo menos 0,50, na medida em que, se assim for, o factor só explica 25% da variabilidade do item, sendo os outros 75% erros, o que significa que se se tiver pesos factoriais inferiores, menor será a variabilidade do item que é explicada pelo factor, sendo tudo o resto erros. Caso se verifique que é necessário mais do que uma dimensão ou factor para explicar as relações entre os itens, a sua análise deverá ser dividida em tantas partes quantas as dimensões existentes, como se se tratassem de vários testes. Mas antes de se chegar a esta solução deve-se sempre analisar os itens, e os seus pesos factoriais, com base numa única dimensão – só quando se verifica que a maioria dos itens não pontua, isto é, não tem um peso factorial alto no factor, é que se deverá extrair mais factores.

Dificuldade

O grau de dificuldade dos itens depende obviamente do objectivo da testagem – se se pretende uma discriminação máxima entre os indivíduos testados, o grau de dificuldade dos itens

¹³ Para mais informações acerca da análise factorial ver Anexo A.

deverá situar-se em torno de 50% ($p=0,50$), isto é, os itens deverão ter uma dificuldade média; mas, caso se pretenda diferenciar os “melhores” indivíduos testados para um processo de selecção, aí, o grau de dificuldade dos itens deverá ser aumentado, por exemplo para 20% ($p=0,20$) no caso de se querer seleccionar apenas 20% da amostra, o que significa que se terão itens para os quais somente 20% da amostra tem probabilidade de acertar (Pasquali, 2003; Anastasi & Urbina 2000).

Discriminação (ou sensibilidade)

Segundo Pasquali (2003), é possível calcular o índice de discriminação de duas formas: Através de grupos-critério ou através da correlação item-total. Quando a análise de discriminação é realizada com base em grupos-critério, a escolha destes grupos, que depende obviamente do objectivo da testagem, pode recair em critérios internos ou externos (Pasquali, 2003). Se se utilizam *critérios externos*, a discriminação dos itens é feita com referência a grupos definidos que se diferenciam entre si em algum aspecto do comportamento que é relevante para o objectivo da testagem (Pasquali, 2003). Quando se utilizam *critérios internos*, a discriminação dos itens baseia-se na diferenciação entre os indivíduos quanto ao próprio score do teste, utilizando-se então os chamados grupos-extremos de indivíduos – grupo superior e grupo inferior – que, no caso de amostras grandes, podem ser compostos pelos 27% superiores e os 27% inferiores (Kelley, 1939, in Anastasi & Urbina, 2000; Kelley, 1939, in Pasquali, 2003). Definidos os grupos-critério, o índice de discriminação pode ser calculado através de duas metodologias estatísticas: A estatística D e o teste t-student¹⁴, através das quais se analisam as diferenças entre os grupos-critério – existindo diferenças, os itens têm capacidade de discriminação.

Convém mencionar que para Ribeiro (2008) a análise da discriminação por grupos extremos se destina a testes de realização máxima ou de capacidade, e que é para Maroco (2005) uma demonstração de validade de critério e não da sensibilidade (capacidade de discriminação) dos itens. No entanto, no processo de adaptação do IACLIDE (Vaz-Serra, 1994) a análise da capacidade de discriminação dos itens foi efectuada através dos grupos-extremos, o que vai ao encontro da perspectiva de Pasquali (2003), e o estudo de validade foi efectuada com base na validação por grupos-extremos (deprimidos vs normais), o que está de acordo com Maroco.

¹⁴ Para mais informações a este respeito ver Anastasi e Urbina (2000) e Pasquali (2003).

Também é possível com a correlação item-total calcular o índice de discriminação, através de inúmeras metodologias estatísticas (Anastasi & Urbina, 2000; Pasquali, 2003), a saber: A correlação ponto-bisserial, a correlação bisserial, a correlação phi e a correlação tetracórica, que dependem do tipo de itens que está a ser analisado¹⁵. Quanto mais elevada for a correlação item-total, melhor o item discrimina os indivíduos com melhor resultado dos indivíduos com pior resultado, e, se for elevada, significa também que o item e o score total medem o mesmo construto (Ribeiro, 2008).

Pasquali (2003) sugere ainda a análise da discriminação através da TRI – na qual a dificuldade do próprio item não tem interferência, e a diferenciação dos indivíduos se baseia nas “magnitudes diferentes de traço do qual o item constitui a representação comportamental. Quanto mais próximas forem as magnitudes do traço que o item puder diferenciar, mais discriminativo ele é” (Pasquali, 2003, p.139) –, pois considera as formas anteriores limitadas, referindo que

O cálculo do índice de discriminação com base no score total do sujeito, seja para permitir estabelecer os grupos-critério com o próprio teste ou para estabelecer a correlação entre o item e o escore no teste, apresenta um problema teórico. Na verdade, procura-se analisar a adequação do item (em termos de discriminação) baseada nas informações obtidas de todo o elenco de itens (score total). Tal procedimento parece incongruente, dado que a adequação dos demais itens também está por ser demonstrada, inclusive a esta altura da análise do teste ainda não se sabe se os itens do teste são homogêneos, isto é, se o teste é unidimensional (medindo um único construto), suposição necessária para se poder obter um score total. Tenta-se resolver este problema procedendo-se a uma análise factorial dos itens antes da própria análise individual dos mesmos, para se certificar, pelo menos, que todos os itens estejam se referindo ao mesmo factor. Além disso, como a base dos cálculos da discriminação são os grupos-critério, acontece que para o caso de itens muito fáceis, os quais todos os sujeitos acertam, ou de itens muito difíceis, os quais todo o mundo erra, um dos grupos-critério desaparece, tornando o cálculo da discriminação inviável (Pasquali, 2003, p.139).

Para Maroco (2005), a análise da sensibilidade (capacidade de discriminação) dos itens pode e deve ser realizada através da forma da distribuição dos itens, com as medidas de assimetria e de achatamento – os itens discriminam ou são sensíveis quando apresentam um enviesamento e um achatamento a tender para zero. Segundo Maroco (2003, p.30) “A assimetria de uma distribuição pode ser caracterizada pelo enviesamento (*skewness*) que essa

¹⁵ Para mais informações a este respeito ver Anastasi e Urbina (2000) e Pasquali (2003).

distribuição apresenta relativamente à média”, sendo que se considera que “A assimetria resulta, em geral, da presença de valores extremamente elevados ou pequenos” (Reis, 1996, p.127). Assim,

Se a distribuição for simétrica (em torno da média) o valor de g_1 ¹⁶ é nulo; se a distribuição se concentrar no lado direito com uma longa cauda para o lado esquerdo o valor de g_1 é menor que zero e a distribuição diz-se com enviesamento negativo ou assimétrica à esquerda. Finalmente, se a distribuição se concentrar no lado esquerdo com uma longa cauda para a direita, g_1 toma valores maiores do que zero e a distribuição diz-se com enviesamento positivo ou assimétrica à direita (Maroco, 2003, p.30-31).

Além disto, “A distribuição da variável pode ainda ser caracterizada pela sua forma mais ou menos achatada (...)” (Maroco, 2003, p.31), que se “avalia” através do coeficiente de achatamento ou de curtose. Sendo a curtose o grau de achatamento de uma distribuição, “As medidas de curtose dão-nos, assim, uma indicação da intensidade das frequências na vizinhança dos valores centrais” (Reis, 1996, p.130). Como refere Maroco (2003, p.31)

Se g_2 ¹⁷ for próximo de zero a distribuição diz-se mesocúrtica. Se a distribuição for achatada, g_2 toma valores menores que zero, e a distribuição diz-se platocúrtica. Pelo contrário se a distribuição for pontiaguda, g_2 toma valores maiores do que zero, e a distribuição diz-se leptocúrtica.

Uma outra forma de se analisar a sensibilidade dos itens, e que foi usada por Beck et al. (1996) no seu estudo de validação do BDI-II, é através da representação gráfica da probabilidade de cada uma das opções de resposta, de cada item, em função do score total – os itens discriminam ou são sensíveis quando reflectem uma cotação ordinal apropriada, isto é, quando as várias opções de resposta se diferenciam umas das outras quanto ao score total.

Respostas enviesadas

Tal como referido anteriormente, as respostas enviesadas podem derivar de factores culturais, de respostas aleatórias e de respostas estereotipadas. Os erros de resposta provocados por questões culturais relacionam-se, sobretudo, com o facto de se utilizarem testes psicológicos em amostras para as quais estes não foram criados e validados. Existem várias metodologias estatísticas para lidar com este problema (“viés do item” e “viés do teste”), que se baseiam na TRI e no DIF (Differential Item Functioning), que têm como objectivo “...identificar itens

¹⁶ g_1 =coeficiente de assimetria.

¹⁷ g_2 =coeficiente de achatamento ou curtose.

para os quais pessoas igualmente capazes de diferentes grupos culturais têm diferentes probabilidades de sucesso. Para este objectivo, habilidade igual significa igual com relação ao construto que o teste deve avaliar, ou o comportamento de critério que foi planejado para predizer” (Anastasi & Urbina, 2000, p.172).

As respostas aleatórias, ou respostas ao acaso, são respostas enviesadas por factores aleatórios, não sistemáticos, como por exemplo a incompreensão das instruções, o facto do indivíduo responder de má fé (contrariado) ao teste ou responder gozando, etc., e, portanto, são respostas muito difíceis de se controlar. Tanto a TRI como a TCT tentam dar, de alguma forma, uma solução a este problema identificando a ocorrência deste tipo de resposta¹⁸. Por último, as respostas estereotipadas, que podem ser vistas como erros ou tendências, dependem das características dos indivíduos testados e são mais evidentes nos testes de personalidade (Pasquali, 2003). Estas respostas podem ser de dois tipos (Pasquali, 2003): As respostas socialmente desejáveis e as respostas sistemáticas.

A *desejabilidade social* é, segundo Edwards (1957, in Pasquali, 2003, p.154-155) “a tendência dos sujeitos em atribuir a si mesmos, em caso de autodescrição, afirmações de personalidade com valores socialmente desejáveis e em rejeitar aquelas com valores socialmente indesejáveis”. Esta atitude não representa, necessariamente, uma vontade deliberada de falsificar os dados (Anastasi & Urbina, 2000; Pasquali, 2003), aliás, ela “...pode indicar falta de insight em relação às próprias características, auto-engano ou relutância em enfrentar as próprias limitações” (Edwards, 1957, citado por Anastasi & Urbina, 2000, p.311). Outras investigações (Crown & Marlowe, 1964, & Frederiksen, 1965, in Anastasi & Urbina, 2000) dão conta que a tendência para dar respostas socialmente desejáveis se prende com questões como a auto-protecção, o evitar críticas, e a conformidade e aprovação social; enquanto a tendência para dar respostas socialmente indesejáveis se liga a uma necessidade de atenção, de simpatia ou de ajuda. Apesar de ser impossível controlar este tipo de resposta, pode-se contudo confirmar se existe esta tendência de resposta, introduzindo-se uma escala de desejabilidade social à situação de testagem.

Enquanto a desejabilidade social representa um traço de personalidade, a *resposta sistemática* representa uma falha de julgamento (Pasquali, 2003), que pode ser de vários tipos:

¹⁸ A este respeito ver Pasquali (2003).

O efeito de halo, o erro de leniência, a tendência central, o erro de contraste, a aquiescência, entre outros. Assim como a desejabilidade social, também é extremamente difícil controlar este tipo de resposta, que constitui um grande problema na avaliação da personalidade. Independentemente das tentativas diversificadas de controlo destas tendências, uma conclusão é aceite pela maioria dos investigadores e criadores de testes psicológicos: “...os scores dos inventários de personalidade tendem a reflectir uma combinação de auto-engano, manejo de impressão¹⁹ e auto-retrato realista, e que o peso de cada um destes componentes varia de acordo com o indivíduo e a ocasião” (Anastasi & Urbina, 2000, p.313).

No processo de validação dos itens, podem ser utilizados outros índices, para além dos já referidos, que se baseiam na relação entre a resposta ao item e o score total do teste (Anastasi, 1990, in Ribeiro, 2008), e são eles:

- *Dificuldade do item* – este índice destina-se aos itens dos testes de realização máxima ou de capacidade, na medida em que indica a percentagem de pessoas que responde acertadamente;
- *Discriminação do item* – este índice também se destina a testes de realização máxima e indica a capacidade de discriminação entre os indivíduos com melhor resultado e os indivíduos com pior resultado (pode ser feita através da comparação de grupos extremos, como referido anteriormente);
- *Distribuição das respostas* – este índice destina-se a testes de realização típica ou habitual e diz que para que um item seja considerado bom, deverão existir respostas em todas as alternativas que o constituem;
- *Correlação item-total* – este índice diz que quanto mais elevada for a correlação, melhor o item discrimina os indivíduos com melhor resultado dos indivíduos com pior resultado, e, se for elevada, diz que o item e o score total medem o mesmo construto;
- *Consistência interna do item* – uma boa consistência interna do item indica que o item tem uma boa correlação com o score total do teste de que faz parte, resultando numa maior homogeneidade;
- *Validade convergente-discriminante do item* – indica que o item mede o mesmo construto do teste a que pertence e não outro;
- *Itens de testes referentes a critério ou a conteúdos* – um bom item deve discriminar eficazmente um indivíduo antes e após a aprendizagem.

¹⁹ Sendo manejo de impressão definido por Anastasi e Urbina (2000, p.311) como a “simulação consciente destinada a criar um efeito específico desejado pelo respondente”.

Anexo D: Estatística Descrita Relativa à Amostra

Frequencies

Statistics

Género		
N	Valid	670
	Missing	0

Género

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Feminino	572	85,4	85,4	85,4
	Masculino	98	14,6	14,6	100,0
	Total	670	100,0	100,0	

Statistics

Idade

N	Valid	670
	Missing	0
Mean		22,3836
Std. Deviation		4,39636
Minimum		18,00
Maximum		48,00

Idade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18,00	31	4,6	4,6	4,6
	19,00	63	9,4	9,4	14,0
	20,00	111	16,6	16,6	30,6
	21,00	162	24,2	24,2	54,8
	22,00	108	16,1	16,1	70,9
	23,00	71	10,6	10,6	81,5
	24,00	24	3,6	3,6	85,1
	25,00	21	3,1	3,1	88,2
	26,00	15	2,2	2,2	90,4
	27,00	10	1,5	1,5	91,9
	28,00	13	1,9	1,9	93,9
	29,00	5	,7	,7	94,6
	30,00	2	,3	,3	94,9
	31,00	7	1,0	1,0	96,0
	32,00	4	,6	,6	96,6
	33,00	4	,6	,6	97,2
	34,00	2	,3	,3	97,5
	36,00	1	,1	,1	97,6
	38,00	1	,1	,1	97,8
	39,00	1	,1	,1	97,9
	40,00	2	,3	,3	98,2
	41,00	1	,1	,1	98,4
	42,00	1	,1	,1	98,5
	43,00	1	,1	,1	98,7
	44,00	1	,1	,1	98,8
	45,00	1	,1	,1	99,0
	46,00	3	,4	,4	99,4
	47,00	2	,3	,3	99,7
	48,00	2	,3	,3	100,0
	Total	670	100,0	100,0	

Frequencies

Statistics

Curso		
N	Valid	670
	Missing	0

Curso

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	RIS	41	6,1	6,1	6,1
	DCSM	23	3,4	3,4	9,6
	PSIC	606	90,4	90,4	100,0
	Total	670	100,0	100,0	

Statistics

Ano		
N	Valid	670
	Missing	0

Ano

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1º	125	18,7	18,7	18,7
	2º	168	25,1	25,1	43,7
	3º	200	29,9	29,9	73,6
	4º	177	26,4	26,4	100,0
	Total	670	100,0	100,0	

Statistics

Área		
N	Valid	158
	Missing	512

Área

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clínica	100	14,9	63,3	63,3
	Educacional	38	5,7	24,1	87,3
	Social	20	3,0	12,7	100,0
	Total	158	23,6	100,0	
Missing	System	512	76,4		
Total		670	100,0		

Statistics

Turno		
N	Valid	606
	Missing	64

Turno

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Manhã	338	50,4	55,8	55,8
	Tarde	144	21,5	23,8	79,5
	Noite	124	18,5	20,5	100,0
	Total	606	90,4	100,0	
Missing	System	64	9,6		
Total		670	100,0		

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Curso * G�nero	670	100,0%	0	,0%	670	100,0%
Curso * Idade	670	100,0%	0	,0%	670	100,0%
Curso * Ano	670	100,0%	0	,0%	670	100,0%
Curso * �rea	158	23,6%	512	76,4%	670	100,0%
Curso * Turno	606	90,4%	64	9,6%	670	100,0%

Curso * G nero Crosstabulation

Count

		G�nero		Total
		Feminino	Masculino	
Curso	RIS	39	2	41
	DCSM	19	4	23
	PSIC	514	92	606
Total		572	98	670

Curso * Ano Crosstabulation

Count

		Ano				Total
		1�	2�	3�	4�	
Curso	RIS	9	10	6	16	41
	DCSM	6	7	7	3	23
	PSIC	110	151	187	158	606
Total		125	168	200	177	670

Curso *  rea Crosstabulation

Count

		�rea			Total
		Cl�nica	Educacional	Social	
Curso	PSIC	100	38	20	158
Total		100	38	20	158

Curso * Turno Crosstabulation

Count

		Turno			Total
		Manh�	Tarde	Noite	
Curso	PSIC	338	144	124	606
Total		338	144	124	606

Curso *Idade Crosstabulation

Count

		Curso			Total
		RIS	DCSM	PSIC	
Idade	18,00	0	1	30	31
	19,00	1	3	59	63
	20,00	8	2	101	111
	21,00	12	8	142	162
	22,00	7	4	97	108
	23,00	4	3	64	71
	24,00	5	0	19	24
	25,00	1	0	20	21
	26,00	2	1	12	15
	27,00	1	0	9	10
	28,00	0	1	12	13
	29,00	0	0	5	5
	30,00	0	0	2	2
	31,00	0	0	7	7
	32,00	0	0	4	4
	33,00	0	0	4	4
	34,00	0	0	2	2
	36,00	0	0	1	1
	38,00	0	0	1	1
	39,00	0	0	1	1
	40,00	0	0	2	2
	41,00	0	0	1	1
	42,00	0	0	1	1
	43,00	0	0	1	1
	44,00	0	0	1	1
	45,00	0	0	1	1
	46,00	0	0	3	3
	47,00	0	0	2	2
	48,00	0	0	2	2
Total		41	23	606	670

Summarize

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Idade * Curso	670	100,0%	0	,0%	670	100,0%

Case Summaries

Idade

Curso	N	Mean	Std. Deviation
RI5	41	21,9756	1,89061
DCSM	23	21,4783	2,21297
PSIC	606	22,4455	4,57295
Total	670	22,3836	4,39636

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
BDITotal * Curso	670	100,0%	0	,0%	670	100,0%
STAITotal * Curso	670	100,0%	0	,0%	670	100,0%
IACLIDETotal * Curso	559	83,4%	111	16,6%	670	100,0%

Case Summaries

Curso		BDITotal	STAITotal	IACLIDETotal
RI5	N	41	41	41
	Mean	9,2439	37,4146	10,8537
	Std. Deviation	7,12313	9,09938	9,14757
DCSM	N	23	23	23
	Mean	11,5652	38,8261	12,7826
	Std. Deviation	5,79832	8,11089	8,36636
PSIC	N	606	606	495
	Mean	9,4059	38,3119	12,0384
	Std. Deviation	7,68944	11,07018	11,74857
Total	N	670	670	559
	Mean	9,4701	38,2746	11,9821
	Std. Deviation	7,60054	10,86261	11,44895

Anexo E: Instrumento de Recolha de Dados

Obrigado por participar nesta investigação.

O presente estudo tem como objectivo analisar alguns instrumentos de medição da depressão e da ansiedade, utilizados na prática clínica e de investigação em Psicologia. Desta forma, o questionário que se segue, é composto por 3 partes.

É importante que saiba que a sua participação é **voluntária**, os seus dados e as suas respostas são **confidenciais** e servirão apenas para **fins científicos** e não avaliativos.

Por favor, indique os seguintes dados pessoais:

Sexo: Masculino Feminino

Idade: _____

Leia atentamente as instruções de cada parte do questionário.

Se não perceber alguma coisa pergunte ao investigador.

Certifique-se que respondeu a todas as perguntas.

PRIMEIRA PARTE

Instruções: Este questionário é composto por grupos de frases que descrevem sentimentos e comportamentos. Escolha, em cada grupo, aquela frase que melhor descreve o que tem sentido **nas últimas duas semanas incluindo hoje**. Faça um círculo à volta do número que se encontra no início da afirmação que escolheu.

Exemplo**Felicidade**

- | | |
|---|-----------------------------|
| 0 | Nunca me sinto feliz. |
| ① | Raramente me sinto feliz. |
| 2 | Normalmente sinto-me feliz. |
| 3 | Sinto-me sempre feliz. |

1. Tristeza

- 0 Não me sinto triste.
- 1 Sinto-me triste a maior parte do tempo.
- 2 Estou sempre triste.
- 3 Estou tão triste ou infeliz que não consigo suportar isto.

2. Pessimismo

- 0 Não me sinto desencorajado em relação ao meu futuro.
- 1 Sinto-me mais desencorajado em relação ao meu futuro do que costumava sentir-me.
- 2 Não espero que as coisas me corram bem.
- 3 Não tenho qualquer esperança no futuro e sinto que a minha situação só vai piorar.

3. Fracasso no passado

- 0 Não me sinto um falhado.
- 1 Falhei mais do que deveria ter falhado.
- 2 Quando olho para o passado, vejo muitos fracassos.
- 3 Sinto que sou uma pessoa completamente falhada.

4. Perda de prazer

- 0 Continuo a ter o mesmo prazer que tinha nas coisas que gosto.
- 1 Já não tenho tanto prazer nas coisas como anteriormente.
- 2 Tenho muito pouco prazer nas coisas que anteriormente me satisfaziam.
- 3 Não tenho prazer nenhum nas coisas que anteriormente me satisfaziam.

5. Sentimentos de culpa

- 0 Não me sinto culpado de nada em particular.
- 1 Sinto-me culpado por muitas coisas que fiz ou que deveria ter feito.
- 2 Sinto-me bastante culpado a maior parte do tempo.
- 3 Sinto-me sempre culpado.

6. Sentimentos de punição

- 0 Não sinto que estou a ser castigado.
- 1 Sinto que posso vir a ser castigado.
- 2 Espero vir a ser castigado.
- 3 Sinto que estou a ser castigado.

7. Desprezo por si próprio

- 0 Sinto-me tão bem comigo próprio agora como anteriormente.
- 1 Perdi a confiança em mim.
- 2 Estou desiludido comigo.
- 3 Não gosto de mim próprio.

8. Autocrítica

- 0 Não me critico ou culpo mais do que é habitual.
- 1 Sou mais crítico em relação a mim do que costumava ser.
- 2 Culpo-me por todos os meus defeitos.
- 3 Culpo-me por tudo o que acontece de mal.

9. Pensamentos ou desejos suicidas

- 0 Não penso em suicidar-me.
- 1 Penso em suicidar-me, mas não seria capaz de fazê-lo.
- 2 Gostaria de me suicidar.
- 3 Gostaria de me suicidar se tivesse oportunidade.

10. Choro

- 0 Não choro mais do que costumava chorar.
- 1 Choro mais agora do que costumava chorar.
- 2 Choro por qualquer pequena coisa.
- 3 Apetece-me chorar, mas não sou capaz.

11. Agitação

- 0 Não estou mais agitado ou inquieto do que é habitual.
- 1 Sinto-me mais agitado ou inquieto do que é habitual.
- 2 Estou tão agitado que é difícil ficar quieto.
- 3 Estou tão agitado que tenho de estar em constante movimento ou a fazer algo.

12. Perda de interesse

- 0 Não perdi o interesse por outras pessoas ou actividades.
- 1 Estou menos interessado nas outras pessoas ou actividades do que anteriormente.
- 2 Perdi grande parte do interesse pelas outras pessoas ou actividades.
- 3 É difícil interessar-me por qualquer coisa.

13. Indecisão

- 0 Tomo decisões tão bem como anteriormente.
- 1 É mais difícil tomar decisões do que anteriormente.
- 2 Tenho muito mais dificuldade em tomar decisões do que costumava ter.
- 3 Não consigo tomar nenhuma decisão.

14. Desvalorização pessoal

- 0 Não sinto que não tenho valor.
- 1 Não me considero tão merecedor ou útil como me costumava considerar.
- 2 Sinto que tenho menos valor quando me comparo com outras pessoas.
- 3 Sinto-me completamente sem valor.

15. Perda de energia

- 0 Continuo a ter a mesma energia que anteriormente.
- 1 Tenho menos energia do que costumava ter.
- 2 Não tenho energia suficiente para fazer muita coisa.
- 3 Não tenho energia suficiente para fazer alguma coisa.

16. Alterações nos hábitos do sono

- 0 Não senti nenhuma alteração no meu padrão de sono.

1a Durmo um pouco mais do que é habitual.

1b Durmo um pouco menos do que é habitual.

2a Durmo muito mais do que é habitual.

2b Durmo muito menos do que é habitual.

3a Durmo a maior parte do dia.

3b Acordo 1-2 horas mais cedo e não consigo voltar a adormecer.

17. Irritabilidade

- 0 Não me irrita mais do que é habitual.
- 1 Irrito-me mais do que é habitual.
- 2 Irrito-me muito mais do que é habitual.
- 3 Irrito-me com tudo.

18. Alterações no apetite

- 0 Não senti qualquer alteração no meu apetite.

1a O meu apetite é pouco menor do que é habitual.

1b O meu apetite é pouco maior do que é habitual.

2a O meu apetite é muito menor do que é habitual.

2b O meu apetite é muito maior do que é habitual.

3a Não tenho apetite absolutamente nenhum.

3b Tenho sempre vontade de comer.

19. Dificuldades de concentração

- 0 Consigo concentrar-me tão bem como antes.
- 1 Não consigo concentrar-me tão bem como era habitual.
- 2 É difícil concentrar-me em algo durante muito tempo.
- 3 Acho que não consigo concentrar-me em nada.

20. Cansaço ou Fadiga

- 0 Não estou mais cansado ou fatigado do que é habitual.
- 1 Canso-me ou fatigo-me mais facilmente do que é habitual.
- 2 Estou demasiado cansado ou fatigado para fazer uma quantidade de coisas que costumava fazer.
- 3 Estou demasiado cansado ou fatigado para fazer a maior parte das coisas que costumava fazer.

21. Perda de interesse sexual

- 0 Não notei qualquer alteração recente no meu interesse sexual.
- 1 Estou menos interessado em sexo do que costumava estar.
- 2 Agora estou muito menos interessado em sexo.
- 3 Perdi completamente o interesse sexual.

SEGUNDA PARTE

Instruções: Em baixo tem uma série de frases que são habitualmente utilizadas para descrever pessoas. Leia cada uma delas e assinale com uma cruz (X) o algarismo da direita que melhor indica **como se sente neste momento**. Não há respostas certas ou erradas. Não demore muito tempo com cada frase; responda de modo a descrever o melhor possível a maneira **como se sente agora**.

	NADA	UM POUCO	MODERA DAMENTE	MUITO
1. Sinto-me calmo	1	2	3	4
2. Sinto-me seguro	1	2	3	4
3. Estou tenso	1	2	3	4
4. Sinto-me cansado	1	2	3	4
5. Sinto-me à vontade	1	2	3	4
6. Sinto-me perturbado	1	2	3	4
7. Presentemente, preocupo-me com possíveis desgraças	1	2	3	4
8. Sinto-me satisfeito	1	2	3	4
9. Sinto-me amedrontado	1	2	3	4
10. Sinto-me confortável	1	2	3	4
11. Sinto-me auto-confiante	1	2	3	4
12. Sinto-me nervoso	1	2	3	4
13. Sinto-me trémulo	1	2	3	4
14. Sinto-me indeciso	1	2	3	4
15. Sinto-me descontraído	1	2	3	4
16. Sinto-me contente	1	2	3	4
17. Estou preocupado	1	2	3	4
18. Sinto-me confuso	1	2	3	4
19. Sinto-me firme	1	2	3	4
20. Sinto-me bem	1	2	3	4

TERCEIRA PARTE

Instruções: Cada um dos sintomas que a seguir é apresentado não se refere à sua maneira de ser habitual mas sim ao **estado em que se sente pelo menos desde há duas semanas para cá**. Não há respostas certas ou erradas. Há apenas a sua resposta. Responda de forma rápida, honesta e espontânea. Se lhe parecer que duas respostas diferentes já foram experimentadas por si, assinale com uma cruz (X), no quadrado respectivo , aquela que se aproxime mais do seu estado actual.

1. - O interesse pelo meu trabalho é o mesmo de sempre.
- Ultimamente ando a sentir um interesse menor pelo meu trabalho.
- Presentemente sinto, com bastante frequência, desinteresse pelo que tenho de fazer.
- Actualmente ando a sentir, muitíssimas vezes, um grande desinteresse pelo meu trabalho.
- Perdi por completo todo o interesse que tinha pelo meu trabalho.

2. - No meu dia-a-dia não preciso de recorrer à ajuda das outras pessoas mais do que o costume.
- Presentemente recorro mais ao auxílio dos outros do que costumava fazer.
- No momento actual recorro bastante mais ao auxílio dos outros do que costumava fazer.
- Agora recorro muitíssimo mais ao auxílio dos outros do que costumava fazer.
- Actualmente sinto-me completamente incapaz de fazer seja o que for sem a ajuda dos outros.

3. - Ando a sentir-me tão sossegado como de costume.
- Ando a sentir-me interiormente mais desassossegado do que o habitual.
- Sinto-me interiormente bastante mais desassossegado do que o costume.
- Sinto-me interiormente muitíssimo mais desassossegado do que o habitual.
- Sinto-me interiormente num estado de permanente agitação.

4. - Presentemente, no contacto com os outros (familiares, amigos ou conhecidos), sou capaz de falar sobre os temas habituais de conversas comuns.
- No momento actual tenho tendência, no contacto com os outros, a falar mais do que o habitual daquilo que sinto ou me preocupa, embora consiga conversar também sobre outros assuntos.
- Reconheço que agora passo bastante tempo a falar aos outros daquilo que me anda a preocupar e estou a sentir, embora possa também abordar outros assuntos.
- No momento presente, no contacto com os outros, falo muitíssimo mais do que o costume daquilo que me preocupa e ando a sentir, evitando abordar assuntos diferentes.
- Actualmente, no contacto com os outros, só sei falar daquilo que me preocupa e ando a sentir, mostrando-me incapaz de conversar de quaisquer outros assuntos.

5. - A minha consciência não me acusa de nada em particular que tenha feito.
- Ando a sentir-me um fardo para a minha família, amigos e colegas.
- Sinto actualmente que há falhas no meu passado que não devia ter cometido.
- Penso, muitas vezes, que cometi certas falhas graves no meu passado.
- Passo o tempo a pensar que o estado em que me sinto é consequência das falhas graves que cometi na minha vida.
6. - A minha capacidade de fixação é a mesma que sempre tive.
- Às vezes custa-me fixar aquilo que preciso.
- Sinto, bastantes vezes, que ando a fixar pior aquilo que preciso.
- Sinto, muitíssimas vezes, que me esqueço facilmente das coisas.
- Sinto-me incapaz de fixar seja o que for.
7. - Não me sinto mais infeliz do que o costume.
- Tenho períodos em que me sinto mais infeliz do que habitualmente.
- Ando a sentir-me bastante mais infeliz do que o costume, embora tenha ocasiões em que me sinto tão feliz como dantes.
- Só em raros momentos é que consigo sentir-me feliz.
- Actualmente não tenho um único momento em que me sinta feliz.
8. - No momento presente não tenho qualquer dificuldade em começar a dormir.
- Quando me deito passo cerca de meia hora sem conseguir adormecer.
- Actualmente, quando me deito, demoro cerca de 1 – 2 horas antes de conseguir adormecer.
- Em algumas noites, quando me deito, não consigo dormir nada.
- Se não tomar nada para adormecer as minhas noites são de insónia completa.
9. - O valor que dou às coisas que faço e a mim próprio não é maior nem é menor do que o costume.
- Actualmente penso por vezes que, como indivíduo, tenho pouco mérito naquilo que faço.
- Ultimamente apanho-me a pensar, bastantes vezes, que não valho nada.
- Penso, com muita frequência, que não tenho qualquer valor como pessoa.
- No momento actual sinto constantemente que não valho nada como pessoa.
10. - A minha confiança em relação ao futuro é a mesma de sempre.
- Ao contrário de antigamente, sinto-me por vezes inseguro em relação ao futuro.
- No momento actual sinto, bastantes vezes, falta de confiança no futuro.
- Sinto agora, muitíssimas vezes, falta de confiança no futuro.
- Presentemente perdi por completo a esperança que tinha no futuro.
11. - Depois de conseguir adormecer não tenho tendência a acordar durante a noite.
- Actualmente acordo 1 hora antes do habitual e depois não consigo voltar a dormir.
- Presentemente ando a acordar ao fim de 4 – 5 horas de sono e depois não consigo voltar a dormir.
- Após ter adormecido, acordo ao fim de 1 ou 2 horas e não sou capaz de voltar a dormir.
- Se não tomar nada para dormir, a minha noite é de completa insónia.

12. - Considerando a minha vida passada, não sinto que tenha sido um fracassado.
- No momento presente, observando o meu passado, acho que tive alguns fracassos.
- Penso que no meu passado existiu um número bastante grande de fracassos.
- Actualmente considero que tive, no meu passado, um número muitíssimo grande de fracassos.
- Quando observo a minha vida passada, considero-a como um acumular constante de fracassos.
13. - Quando acordo de manhã sinto-me renovado para enfrentar o meu dia.
- Apesar de ir fazendo o que necessito de fazer, quando acordo de manhã tenho a sensação de estar sem energia para o que preciso de efectuar.
- Logo que acordo de manhã, e posteriormente ao longo do dia, sinto-me bastante fatigado e é com dificuldade que vou realizando o que preciso de fazer.
- Desde que acordo de manhã, e depois durante o dia, sinto-me tão fatigado que já não consigo fazer tudo o que tenho a fazer.
- Quando acordo de manhã o meu cansaço é tão grande que me sinto incapaz de fazer seja o que for.
14. - O meu convívio com as outras pessoas é o mesmo de sempre.
- O meu estado de espírito actual leva-me, por vezes, e mais do que o costume, a evitar conviver com as outras pessoas.
- No momento presente fujo bastante do convívio com as outras pessoas.
- Devido à maneira como me ando a sentir, raramente me sinto capaz de conviver com as outras pessoas.
- Presentemente não me sinto capaz de conviver seja com quem for.
15. - No momento presente não noto quaisquer dificuldades na minha capacidade de concentração.
- Sinto por vezes dificuldade em me concentrar no que estou a fazer.
- Sinto com frequência dificuldade em me concentrar naquilo que executo.
- Sinto muitíssimas vezes dificuldade em me concentrar naquilo que faço.
- Sinto-me actualmente incapaz de me concentrar seja no que for.
16. - Não penso que os meus problemas sejam mais graves do que o habitual.
- Ando com medo de não ser capaz de resolver alguns dos problemas que tenho.
- Penso com frequência que não é fácil encontrar solução para alguns dos meus problemas.
- Passo o tempo preocupado com os meus problemas pois sinto que são de solução muitíssimo difícil.
- Considero que os meus problemas não têm qualquer solução possível.
17. - Quando é necessário sou capaz de tomar decisões tão bem como dantes.
- Agora sinto-me mais hesitante do que o costume para tomar qualquer decisão.
- No momento actual demoro bastante mais tempo do que dantes para tomar qualquer decisão.
- Presentemente custa-me tanto tomar decisões que passo o tempo a adiá-las.
- Actualmente sinto-me incapaz de tomar qualquer decisão sem a ajuda das outras pessoas.

18. - A satisfação que sinto com a vida é a mesma de sempre.
- Actualmente sinto-me mais aborrecido do que o costume com a vida que tenho.
- Sinto, com frequência, que a minha vida não me dá o prazer de antigamente.
- Sinto, com muitíssima frequência, que a minha vida não me traz qualquer satisfação.
- Perdi por completo a alegria de viver.
19. - No momento presente acho que a vida vale a pena ser vivida e não sinto qualquer desejo de fazer mal a mim mesmo.
- Actualmente sinto que a vida não vale a pena ser vivida.
- No momento presente não me importaria de morrer.
- Tenho com muita frequência desejos de que a morte venha.
- Ultimamente tenho feito planos para pôr termo à vida.
20. - Ao executar as minhas tarefas diárias usuais não sinto maior fadiga do que o costume.
- No cumprimento das minhas tarefas, ando a sentir-me mais fatigado do que o habitual sem razão para isso.
- Naquilo que tenho de fazer ao longo do dia, embora sem motivo para isso, sinto-me bastante mais fatigado do que o costume.
- Sinto-me, sem motivo justificativo, muitíssimo fatigado naquilo que tenho de fazer, deixando de executar algumas das tarefas usuais.
- Sinto-me exausto, ao menor esforço, deixando de cumprir a maior parte das tarefas que antigamente fazia.
21. - Quando me comparo com os outros, sei que sou capaz de me sair tão bem como eles a fazer o que é preciso.
- Actualmente, ao comparar-me com os outros, acho por vezes que não me desempenho tão bem como eles nas tarefas que precisam de ser feitas.
- Presentemente considero que as outras pessoas fazem as coisas bastante melhor do que eu.
- Acho que me saio muitíssimo pior do que os outros a fazer o que é necessário ser feito.
- Considero que os outros são sempre capazes de fazer melhor do que eu as coisas que precisam de ser feitas.
22. - Quando executo qualquer tarefa faço-a com a rapidez do costume.
- No momento presente sinto-me mais lento a fazer o que tenho para fazer.
- Actualmente demoro bastante mais tempo do que o habitual a fazer o que tenho para fazer.
- Sinto-me muitíssimo mais lento a fazer tudo quanto tenho para fazer.
- Presentemente demoro uma enormidade de tempo a fazer seja o que for.
23. - A capacidade de me manter activo é a mesma do costume.
- Não me sinto capaz de me manter tão activo como o habitual.
- Actualmente dá-me para passar largos momentos sem fazer absolutamente nada.
- No momento actual raramente faço o que preciso de fazer.
- Só tenho vontade de estar todo o tempo sentado ou deitado na cama, sem fazer nada.

A preencher pelo investigador

BDI Total = _____	STAI Total = _____
IACLIDE Total = _____ Biol = _____ Cogn = _____ IP = _____ DT = _____ F ₁ = _____ F ₂ = _____ F ₃ = _____ F ₄ = _____ F ₅ = _____	

Anexo F: Análise da Consistência Interna

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	664	99,1
	Excluded(a)	6	,9
	Total	670	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,893	,896	21

Item Statistics

	Mean	Std. Deviation	N
1. Tristeza	,2033	,45217	664
2. Pessimismo	,3675	,55244	664
3. Fracasso no Passado	,3389	,60759	664
4. Perda de Prazer	,3148	,54263	664
5. Sentimentos de Culpa	,4383	,54850	664
6. Sentimentos de Punição	,2319	,65251	664
7. Desprezo Por Si Próprio	,3343	,73869	664
8. Autocrítica	,6160	,65802	664
9. Pensamentos ou Desejos Suicidas	,0813	,32402	664
10. Choro	,4217	,74898	664
11. Agitação	,5858	,63710	664
12. Perda de Interesse	,3991	,63725	664
13. Indecisão	,4819	,64026	664
14. Desvalorização Pessoal	,2937	,67685	664
15. Perda de Energia	,6175	,67807	664
16. Alterações nos Hábitos do Sono	,8509	,72682	664
17. Irritabilidade	,5934	,71423	664
18. Alterações no Apetite	,6130	,74454	664
19. Dificuldades de Concentração	,7063	,75680	664
20. Cansaço ou Fadiga	,7666	,73975	664
21. Perda de Interesse Sexual	,2485	,54923	664

Inter-Item Correlation Matrix

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
1. Tristeza	1,000																				
2. Pessimismo	,310	1,000																			
3. Fracasso no Passado	,336	,231	1,000																		
4. Perda de Prazer	,440	,313	,335	1,000																	
5. Sentimentos de Culpa	,212	,224	,364	,200	1,000																
6. Sentimentos de Punição	,326	,228	,258	,330	,322	1,000															
7. Desprezo Por Si Próprio	,424	,375	,379	,444	,315	,368	1,000														
8. Autocrítica	,278	,268	,277	,255	,266	,239	,379	1,000													
9. Pensamentos ou Desejos Suicidas	,330	,288	,266	,360	,283	,360	,340	,295	1,000												
10. Choro	,361	,241	,226	,352	,240	,315	,355	,305	,318	1,000											
11. Agitação	,183	,159	,153	,173	,166	,166	,160	,185	,127	,186	1,000										
12. Perda de Interesse	,461	,358	,312	,644	,219	,332	,447	,291	,369	,367	,163	1,000									
13. Indecisão	,286	,334	,297	,357	,257	,201	,338	,329	,196	,293	,213	,363	1,000								
14. Desvalorização Pessoal	,391	,352	,363	,384	,270	,269	,551	,413	,352	,380	,125	,434	,383	1,000							
15. Perda de Energia	,352	,291	,235	,393	,232	,204	,286	,262	,197	,324	,195	,427	,380	,370	1,000						
16. Alterações nos Hábitos do Sono	,294	,189	,152	,238	,210	,111	,189	,240	,154	,271	,277	,256	,242	,181	,365	1,000					
17. Irritabilidade	,350	,245	,189	,346	,148	,216	,212	,277	,208	,324	,325	,347	,284	,282	,385	,348	1,000				
18. Alterações no Apetite	,319	,141	,147	,272	,161	,222	,238	,244	,206	,263	,224	,304	,208	,214	,319	,359	,347	1,000			
19. Dificuldades de Concentração	,307	,305	,279	,339	,267	,199	,313	,246	,196	,307	,223	,415	,339	,339	,442	,310	,345	,234	1,000		
20. Cansaço ou Fadiga	,350	,280	,264	,367	,234	,222	,314	,290	,243	,347	,319	,377	,305	,321	,573	,454	,448	,307	,457	1,000	
21. Perda de Interesse Sexual	,300	,226	,213	,405	,174	,235	,330	,219	,251	,233	,187	,410	,242	,323	,288	,206	,281	,206	,256	,291	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1. Tristeza	9,3012	54,057	,581	,373	,887
2. Pessimismo	9,1370	54,052	,464	,257	,889
3. Fracasso no Passado	9,1657	53,725	,452	,281	,889
4. Perda de Prazer	9,1898	53,026	,608	,500	,886
5. Sentimentos de Culpa	9,0663	54,490	,412	,257	,890
6. Sentimentos de Punição	9,2726	53,550	,434	,276	,890
7. Desprezo Por Si Próprio	9,1702	51,378	,585	,460	,886
8. Autocrítica	8,8886	53,043	,485	,281	,889
9. Pensamentos ou Desejos Suicidas	9,4232	55,795	,459	,288	,891
10. Choro	9,0828	51,892	,526	,299	,888
11. Agitação	8,9187	54,479	,344	,184	,892
12. Perda de Interesse	9,1054	51,859	,638	,530	,884
13. Indecisão	9,0226	52,921	,514	,304	,888
14. Desvalorização Pessoal	9,2108	51,955	,585	,448	,886
15. Perda de Energia	8,8870	51,953	,584	,449	,886
16. Alterações nos Hábitos do Sono	8,6536	52,857	,448	,311	,890
17. Irritabilidade	8,9111	52,202	,524	,348	,888
18. Alterações no Apetite	8,8916	52,878	,433	,250	,891
19. Dificuldades de Concentração	8,7982	51,645	,543	,344	,887
20. Cansaço ou Fadiga	8,7380	51,162	,606	,481	,885
21. Perda de Interesse Sexual	9,2560	54,127	,457	,247	,889

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
9,5045	58,124	7,62389	21

Anexo G: Análise da Sensibilidade dos Itens – Forma da Distribuição dos Itens

Summarize

Case Processing Summary

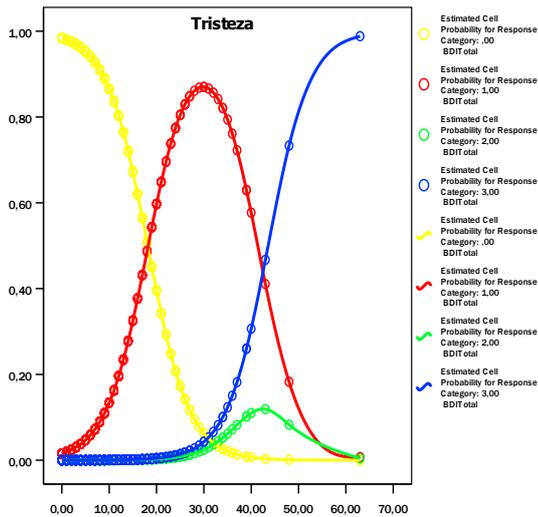
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
1. Tristeza	670	100,0%	0	,0%	670	100,0%
2. Pessimismo	669	99,9%	1	,1%	670	100,0%
3. Fracasso no Passado	670	100,0%	0	,0%	670	100,0%
4. Perda de Prazer	669	99,9%	1	,1%	670	100,0%
5. Sentimentos de Culpa	669	99,9%	1	,1%	670	100,0%
6. Sentimentos de Punição	668	99,7%	2	,3%	670	100,0%
7. Desprezo Por Si Próprio	668	99,7%	2	,3%	670	100,0%
8. Autocrítica	668	99,7%	2	,3%	670	100,0%
9. Pensamentos ou Desejos Suicidas	667	99,6%	3	,4%	670	100,0%
10. Choro	667	99,6%	3	,4%	670	100,0%
11. Agitação	668	99,7%	2	,3%	670	100,0%
12. Perda de Interesse	668	99,7%	2	,3%	670	100,0%
13. Indecisão	668	99,7%	2	,3%	670	100,0%
14. Desvalorização Pessoal	669	99,9%	1	,1%	670	100,0%
15. Perda de Energia	669	99,9%	1	,1%	670	100,0%
16. Alterações nos Hábitos do Sono	669	99,9%	1	,1%	670	100,0%
17. Irritabilidade	669	99,9%	1	,1%	670	100,0%
18. Alterações no Apetite	669	99,9%	1	,1%	670	100,0%
19. Dificuldades de Concentração	669	99,9%	1	,1%	670	100,0%
20. Cansaço ou Fadiga	669	99,9%	1	,1%	670	100,0%
21. Perda de Interesse Sexual	668	99,7%	2	,3%	670	100,0%
BDIScore Total	670	100,0%	0	,0%	670	100,0%

Case Summaries

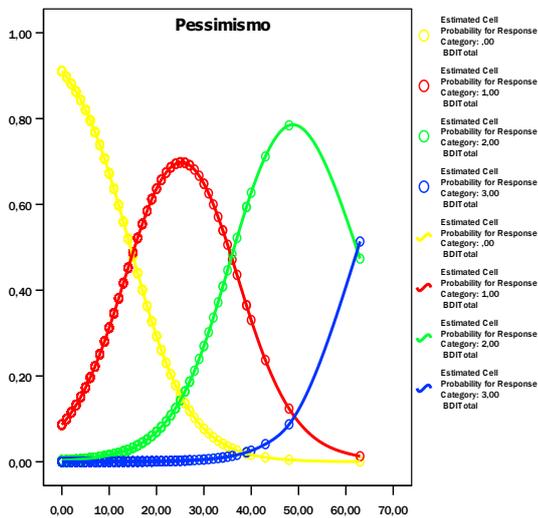
	N	Skewness	Kurtosis
1. Tristeza	670	2,627	9,104
2. Pessimismo	669	1,250	,891
3. Fracasso no Passado	670	1,815	2,997
4. Perda de Prazer	669	1,737	3,251
5. Sentimentos de Culpa	669	,849	,305
6. Sentimentos de Punição	668	3,298	10,688
7. Desprezo Por Si Próprio	668	2,203	3,862
8. Autocrítica	668	,979	1,411
9. Pensamentos ou Desejos Suicidas	667	5,112	33,315
10. Choro	667	1,858	2,888
11. Agitação	668	1,042	1,846
12. Perda de Interesse	668	1,707	3,086
13. Indecisão	668	1,091	,541
14. Desvalorização Pessoal	669	2,192	3,600
15. Perda de Energia	669	,793	,084
16. Alterações nos Hábitos do Sono	669	,612	,263
17. Irritabilidade	669	1,057	,733
18. Alterações no Apetite	669	1,139	,990
19. Dificuldades de Concentração	669	,755	-,177
20. Cansaço ou Fadiga	669	,831	,619
21. Perda de Interesse Sexual	668	2,473	6,552
BDIScore Total	670	1,827	6,104

Anexo H: Análise da Sensibilidade dos Itens – Probabilidade das Opções de Resposta em
Função do Score Total²⁰

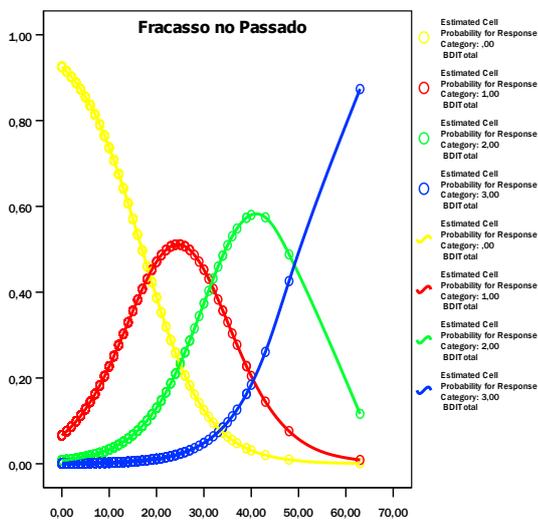
²⁰ Tendo em conta que scores entre 0 e 13 se referem a depressões mínimas; entre 14 e 19 a depressões ligeiras; entre 20 e 28 a depressões moderadas; e entre 29 e 63 se referem a depressões severas.

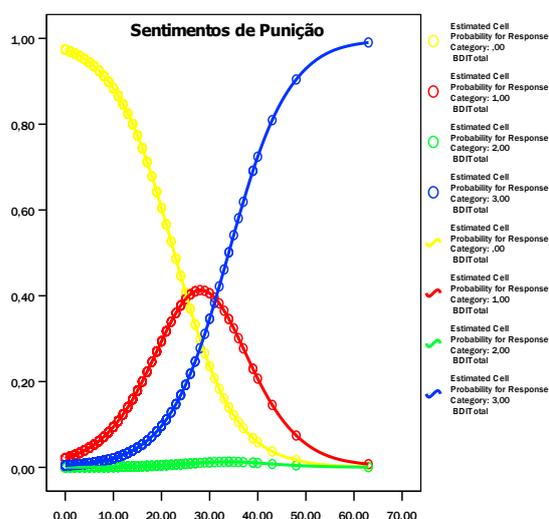
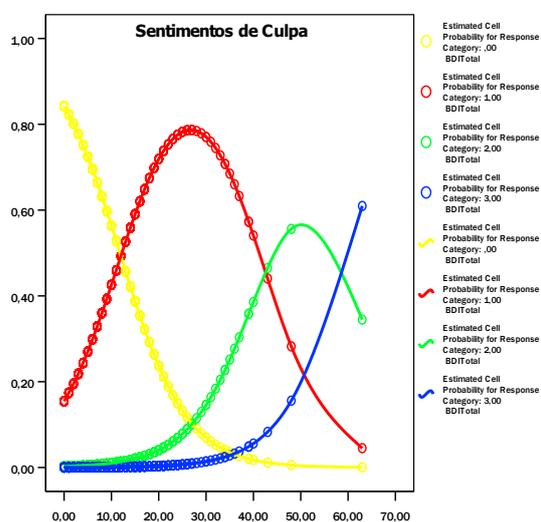
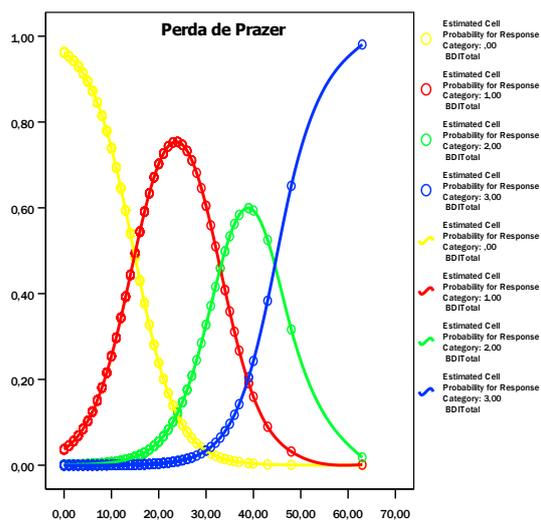


Constata-se que a segunda opção de resposta (cotada com um ponto) é mais provável do que a terceira opção (cotada com dois pontos) em indivíduos com depressão severa (com scores até aproximadamente 52). Parece haver uma cotação ordinal pouco apropriada.



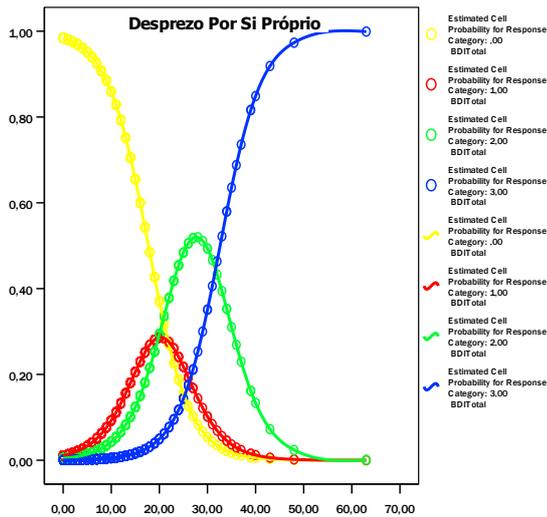
A segunda opção de resposta (cotada com um ponto) é mais provável do que a quarta (cotada com três pontos) em indivíduos com depressão severa (com scores até 50); e a terceira opção de resposta (cotada com dois pontos) é mais provável do que a quarta opção (cotada com três pontos) em indivíduos com depressão severa (com scores até 60).



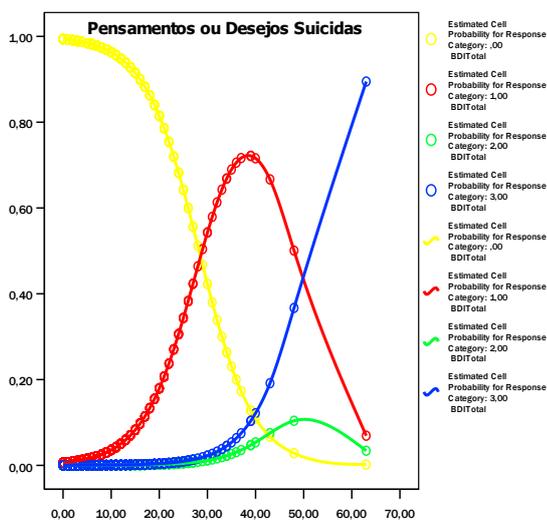
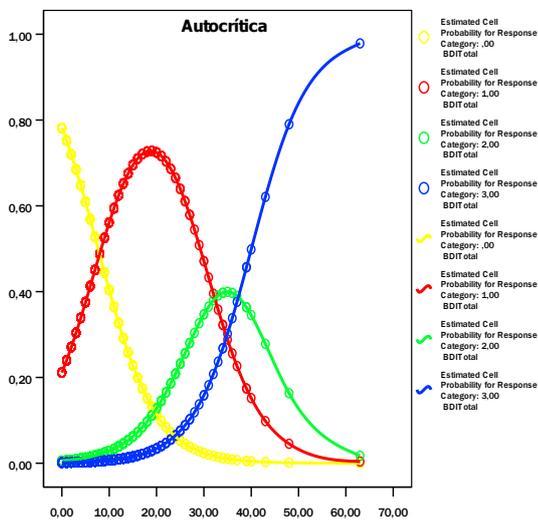


Constata-se que a segunda opção de resposta (cotada com um ponto) é mais provável do que a quarta opção (cotada com três pontos) em indivíduos com depressão severa (com scores entre 40 e 50), no entanto, a terceira opção de resposta (cotada com dois pontos) é mais provável que a segunda entre esses valores.

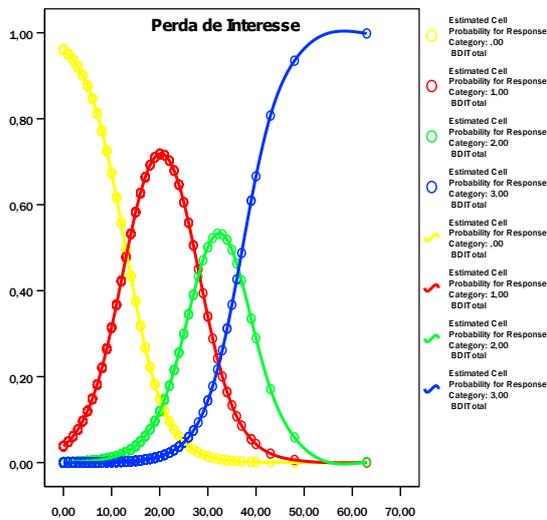
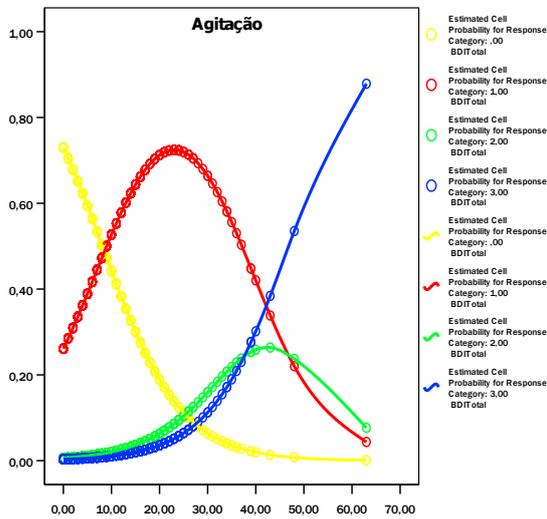
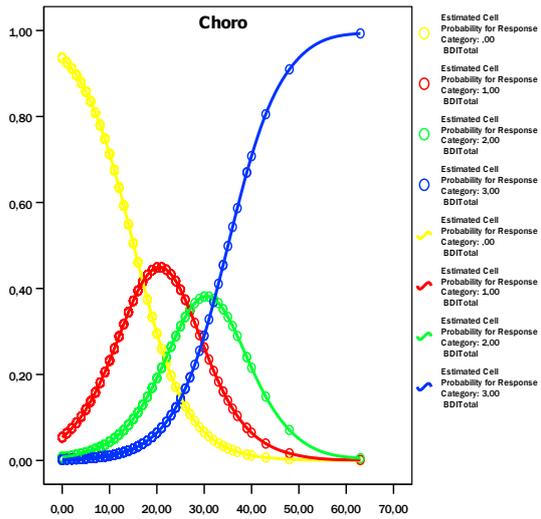
Este item reflecte claramente uma **cotação ordinal não apropriada**, na medida em que a probabilidade da terceira opção de resposta (cotada com dois pontos) é sempre zero seja qual for a severidade da depressão.



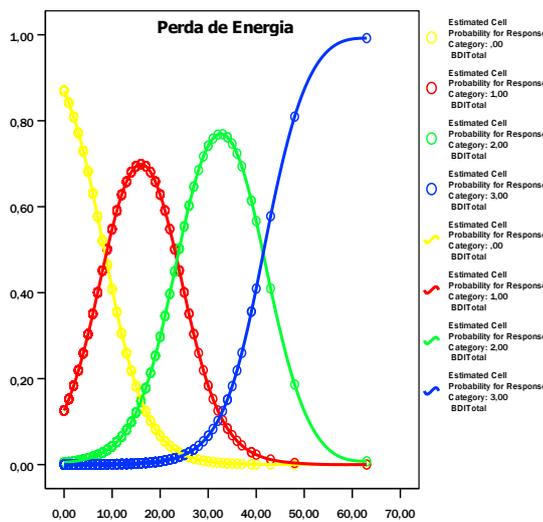
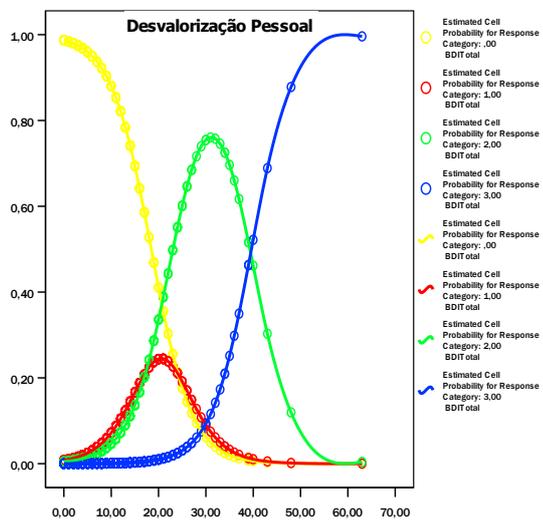
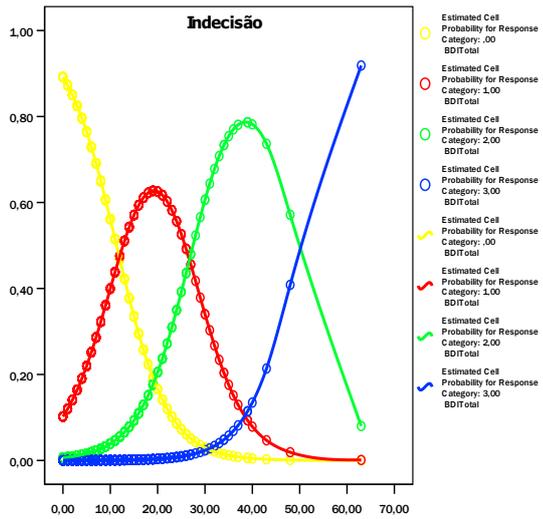
A probabilidade da segunda opção de resposta (cotada com um ponto) deveria ser um pouco mais elevada em indivíduos com depressão ligeira (com scores entre 14 e 19). Parece haver uma cotação ordinal pouco apropriada.



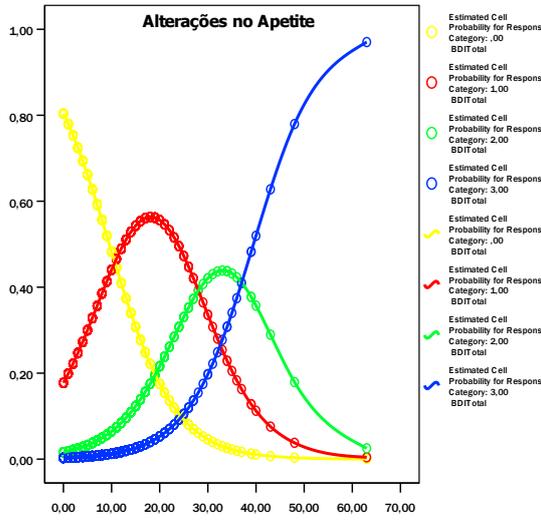
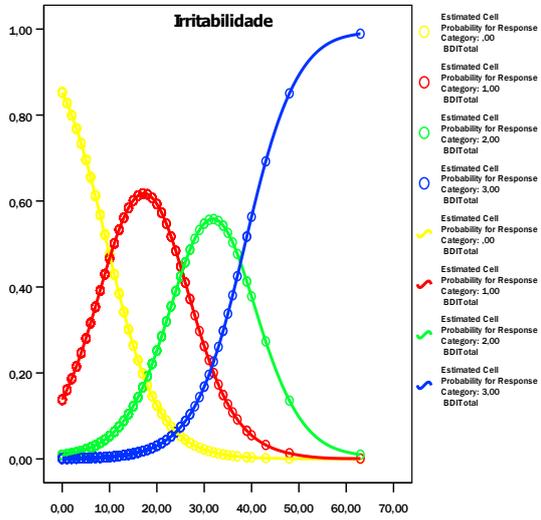
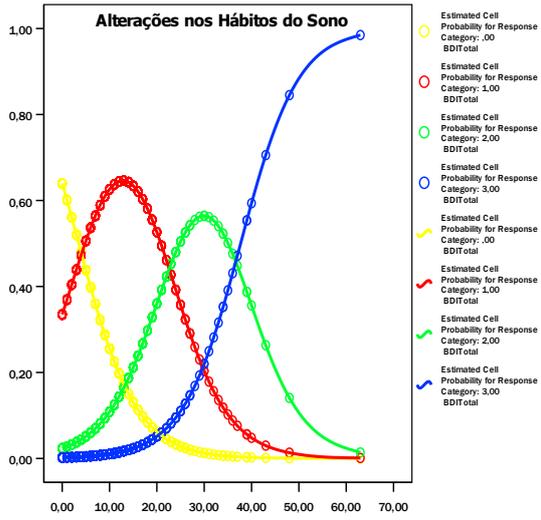
Este item reflecte claramente uma **cotação ordinal não apropriada**, na medida em que a segunda opção de resposta (cotada com 1 ponto) é mais provável que a terceira (cotada com dois pontos) e do que a quarta (cotada com três pontos) em indivíduos com depressão severa (com scores até 50), e mais provável que a terceira em indivíduos com depressão severa (com scores entre 50 e 63).

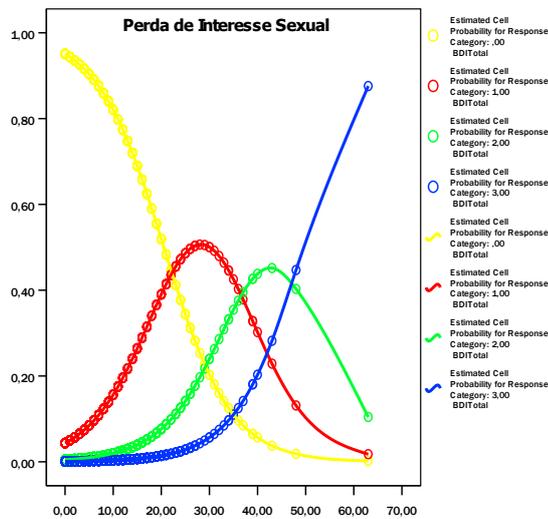
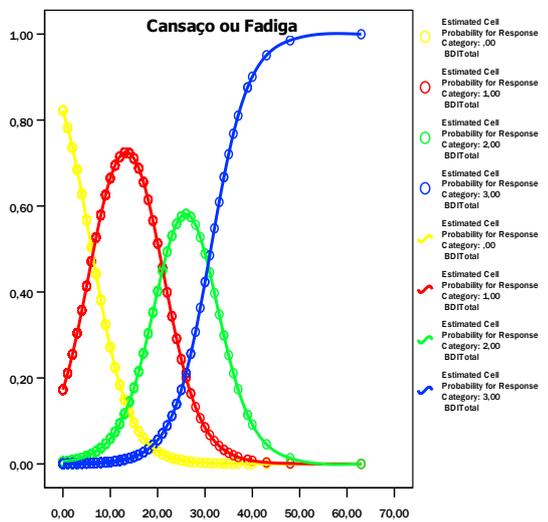
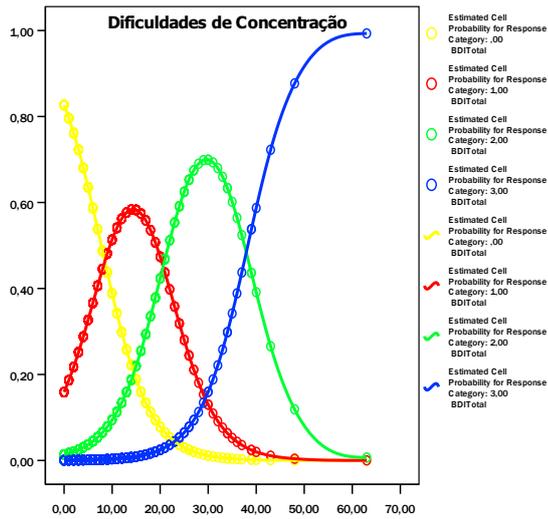


Constata-se que a segunda opção de resposta (cotada com um ponto) é mais provável do que a terceira (cotada com dois pontos) em indivíduos com depressão severa (com scores entre 30 e 45). Parece haver uma cotação ordinal pouco apropriada.



A probabilidade da segunda opção de resposta (cotada com um ponto) deveria ser um pouco superior à probabilidade da primeira opção de resposta (cotada com zero pontos) em indivíduos com depressão ligeira (com scores entre 14 e 19). Parece haver uma cotação ordinal pouco apropriada.





A probabilidade da segunda opção de resposta (cotada com um ponto) deveria ser um pouco superior à probabilidade da primeira opção de resposta (cotada com zero pontos) em indivíduos com depressão ligeira (com scores entre 14 e 19). Parece haver uma cotação ordinal pouco apropriada.

Anexo I: Análise da Validade Concorrente

Correlations (BDI-II/IACLIDE)

Descriptive Statistics

	Mean	Std. Deviation	N
BDITotal	9,4701	7,60054	670
IACLIDETotal	11,9821	11,44895	559

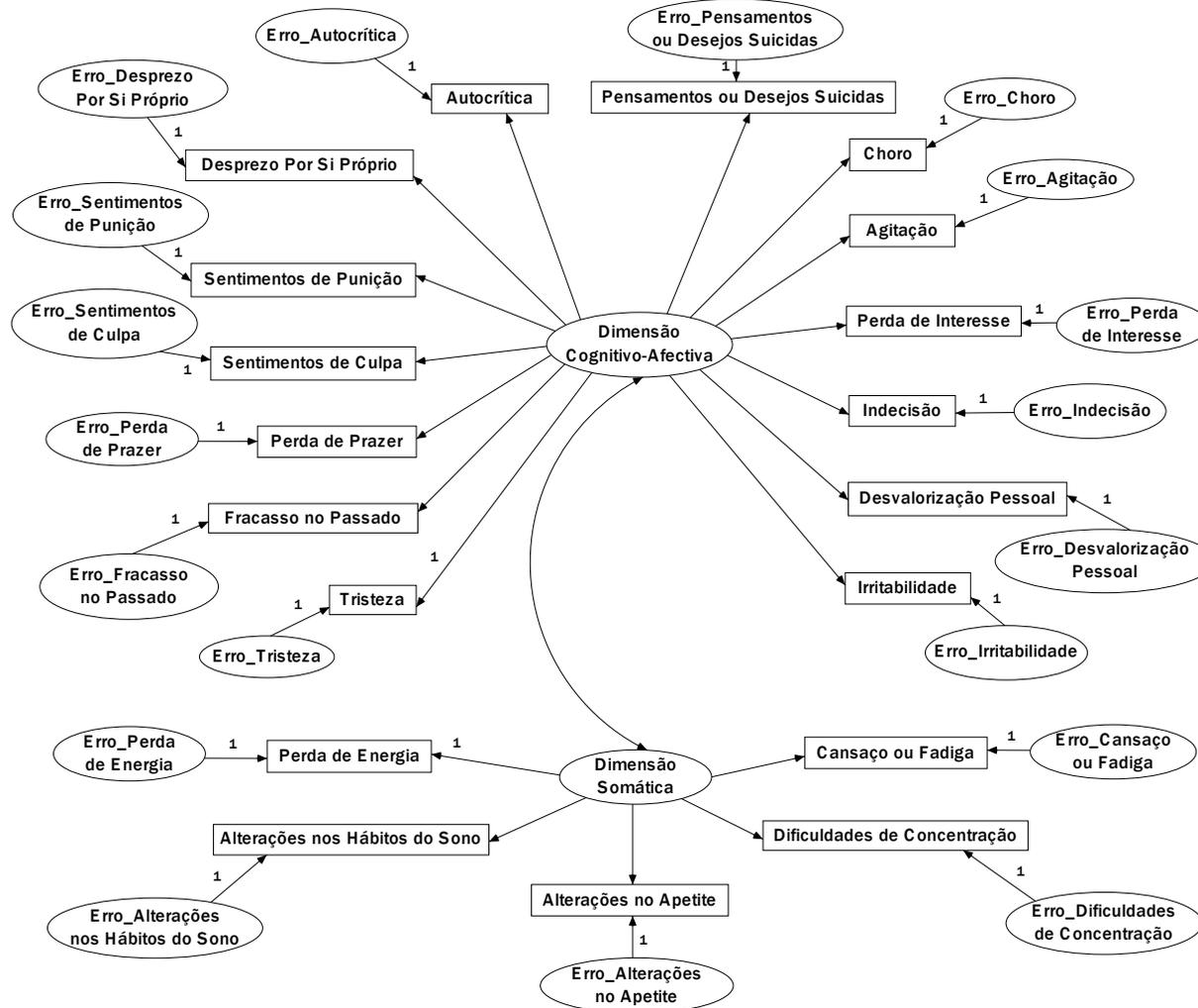
Correlations

		BDITotal	IACLIDETotal
BDITotal	Pearson Correlation	1	,913(**)
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	38646,903	45549,256
	Covariance	57,768	81,629
	N	670	559
IACLIDETotal	Pearson Correlation	,913(**)	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	45549,256	73141,821
	Covariance	81,629	131,079
	N	559	559

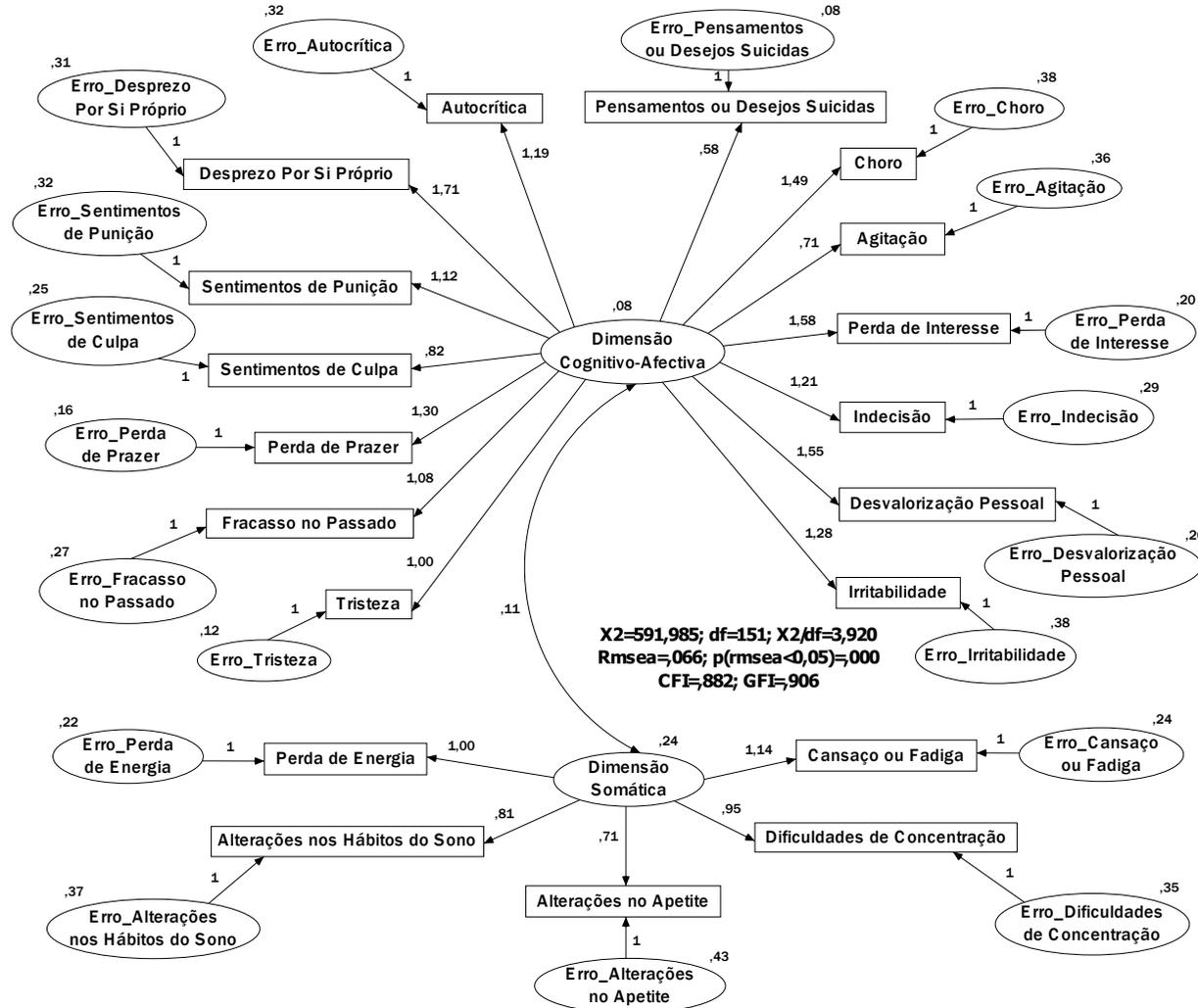
** Correlation is significant at the 0.01 level (2-tailed).

Anexo J: Análise Factorial Confirmatória do Modelo de Beck para a População Académica
do ISPA

Modelo Factorial de Beck para a População de Estudantes Universitários



Análise Factorial Confirmatória do Modelo de Beck para a População Académica do ISPA



Amos Output

Analysis Summary

Groups

Group number 1 (Group number 1)

Notes for Group (Group number 1)

The model is recursive.

Sample size = 664

Variable Summary (Group number 1)

Your model contains the following variables (Group number 1)

Observed, endogenous variables:

Tristeza
 Fracasso no Passado
 Perda de Prazer
 Sentimentos de Culpa
 Sentimentos de Punição
 Desprezo Por Si Próprio
 Autocrítica
 Pensamentos ou Desejos Suicidas
 Choro
 Agitação
 Perda de Interesse
 Indecisão
 Desvalorização Pessoal
 Irritabilidade
 Perda de Energia
 Alterações nos Hábitos do Sono
 Alterações no Apetite
 Dificuldades de Concentração
 Cansaço ou Fadiga

Unobserved, exogenous variables:

Dimensão Cognitivo-Afectiva
 Erro_Tristeza
 Erro_Fracasso no Passado
 Erro_Perda de Prazer
 Erro_Sentimentos de Culpa
 Erro_Sentimentos de Punição
 Erro_Desprezo Por Si Próprio
 Erro_Autocrítica
 Erro_Pensamentos ou Desejos Suicidas
 Erro_Choro
 Erro_Agitação
 Erro_Perda de Interesse
 Erro_Indecisão
 Erro_Desvalorização Pessoal
 Erro_Irritabilidade
 Dimensão Somática
 Erro_Perda de Energia
 Erro_Alterações nos Hábitos do Sono
 Erro_Alterações no Apetite
 Erro_Dificuldades de Concentração
 Erro_Cansaço ou Fadiga

Variable counts (Group number 1)

Number of variables in your model: 40
 Number of observed variables: 19
 Number of unobserved variables: 21
 Number of exogenous variables: 21
 Number of endogenous variables: 19

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	21	0	0	0	0	21
Labeled	0	0	0	0	0	0
Unlabeled	17	1	21	0	0	39
Total	38	1	21	0	0	60

Models

Default model (Default model)

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	190
Number of distinct parameters to be estimated:	39
Degrees of freedom (190 - 39):	151

Result (Default model)

Minimum was achieved
 Chi-square = 591,985
 Degrees of freedom = 151
 Probability level = ,000

Group number 1 (Group number 1 - Default model)

Estimates (Group number 1 - Default model)

Scalar Estimates (Group number 1 - Default model)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
Tristeza	<--- Dimensão Cognitivo-Afectiva	1,000				
Fracasso no Passado	<--- Dimensão Cognitivo-Afectiva	1,083	,095	11,439	***	
Perda de Prazer	<--- Dimensão Cognitivo-Afectiva	1,299	,089	14,585	***	
Sentimentos de Culpa	<--- Dimensão Cognitivo-Afectiva	,822	,084	9,812	***	
Sentimentos de Punição	<--- Dimensão Cognitivo-Afectiva	1,124	,101	11,105	***	
Desprezo Por Si Próprio	<--- Dimensão Cognitivo-Afectiva	1,713	,120	14,224	***	
Autocrítica	<--- Dimensão Cognitivo-Afectiva	1,186	,103	11,544	***	
Pensamentos ou Desejos Suicidas	<--- Dimensão Cognitivo-Afectiva	,585	,051	11,561	***	
Choro	<--- Dimensão Cognitivo-Afectiva	1,486	,119	12,531	***	
Agitação	<--- Dimensão Cognitivo-Afectiva	,714	,095	7,500	***	
Perda de Interesse	<--- Dimensão Cognitivo-Afectiva	1,583	,106	15,006	***	
Indecisão	<--- Dimensão Cognitivo-Afectiva	1,206	,101	11,988	***	
Desvalorização Pessoal	<--- Dimensão Cognitivo-Afectiva	1,549	,110	14,075	***	
Irritabilidade	<--- Dimensão Cognitivo-Afectiva	1,281	,111	11,497	***	
Perda de Energia	<--- Dimensão Somática	1,000				
Alterações nos Hábitos do Sono	<--- Dimensão Somática	,815	,064	12,641	***	
Alterações no Apetite	<--- Dimensão Somática	,714	,066	10,870	***	
Dificuldades de Concentração	<--- Dimensão Somática	,952	,068	14,085	***	
Cansaço ou Fadiga	<--- Dimensão Somática	1,136	,068	16,727	***	

Standardized Regression Weights: (Group number 1 - Default model)

		Estimate
Tristeza	<--- Dimensão Cognitivo-Afectiva	,628
Fracasso no Passado	<--- Dimensão Cognitivo-Afectiva	,507
Perda de Prazer	<--- Dimensão Cognitivo-Afectiva	,680
Sentimentos de Culpa	<--- Dimensão Cognitivo-Afectiva	,426
Sentimentos de Punição	<--- Dimensão Cognitivo-Afectiva	,490
Desprezo Por Si Próprio	<--- Dimensão Cognitivo-Afectiva	,659
Autocrítica	<--- Dimensão Cognitivo-Afectiva	,512
Pensamentos ou Desejos Suicidas	<--- Dimensão Cognitivo-Afectiva	,513
Choro	<--- Dimensão Cognitivo-Afectiva	,564
Agitação	<--- Dimensão Cognitivo-Afectiva	,319
Perda de Interesse	<--- Dimensão Cognitivo-Afectiva	,706
Indecisão	<--- Dimensão Cognitivo-Afectiva	,535
Desvalorização Pessoal	<--- Dimensão Cognitivo-Afectiva	,650
Irritabilidade	<--- Dimensão Cognitivo-Afectiva	,510
Perda de Energia	<--- Dimensão Somática	,723
Alterações nos Hábitos do Sono	<--- Dimensão Somática	,550
Alterações no Apetite	<--- Dimensão Somática	,470
Dificuldades de Concentração	<--- Dimensão Somática	,616
Cansaço ou Fadiga	<--- Dimensão Somática	,753

Covariances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Dimensão Somática <--> Dimensão Cognitivo-Afectiva	,108	,010	10,593	***	

Correlations: (Group number 1 - Default model)

	Estimate
Dimensão Somática <--> Dimensão Cognitivo-Afectiva	,774

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Dimensão Somática	,240	,024	9,850	***	
Dimensão Cognitivo-Afectiva	,081	,009	8,572	***	
Erro_Tristeza	,124	,007	16,683	***	
Erro_Fracasso no Passado	,274	,016	17,402	***	
Erro_Perda de Prazer	,158	,010	16,193	***	
Erro_Sentimentos de Culpa	,246	,014	17,691	***	
Erro_Sentimentos de Punição	,323	,018	17,472	***	
Erro_Desprezo Por Si Próprio	,308	,019	16,415	***	
Erro_Autocrítica	,319	,018	17,379	***	
Erro_Pensamentos ou Desejos Suicidas	,077	,004	17,375	***	
Erro_Choro	,382	,022	17,120	***	
Erro_Agitação	,364	,020	17,944	***	
Erro_Perda de Interesse	,203	,013	15,882	***	
Erro_Indecisão	,292	,017	17,272	***	
Erro_Desvalorização Pessoal	,264	,016	16,497	***	
Erro_Irritabilidade	,377	,022	17,389	***	
Erro_Perda de Energia	,219	,016	13,943	***	
Erro_Alterações nos Hábitos do Sono	,368	,022	16,554	***	
Erro_Alterações no Apetite	,431	,025	17,130	***	
Erro_Dificuldades de Concentração	,355	,022	15,852	***	
Erro_Cansaço ou Fadiga	,237	,018	13,113	***	

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
Tristeza	,395
Fracasso no Passado	,257
Perda de Prazer	,463
Sentimentos de Culpa	,181
Sentimentos de Punição	,240
Desprezo Por Si Próprio	,434
Autocrítica	,262
Pensamentos ou Desejos Suicidas	,263
Choro	,318
Agitação	,101
Perda de Interesse	,499
Indecisão	,286
Desvalorização Pessoal	,423
Irritabilidade	,260
Perda de Energia	,523
Alterações nos Hábitos do Sono	,302
Alterações no Apetite	,221
Dificuldades de Concentração	,380
Cansaço ou Fadiga	,566

Modification Indices (Group number 1 - Default model)**Covariances: (Group number 1 - Default model)**

		M.I.	Par Change
Erro_Alterações nos Hábitos do Sono	<--> Erro_Alterações no Apetite	13,916	,061
Erro_Irritabilidade	<--> Dimensão Cognitivo-Afectiva	30,607	-,031
Erro_Irritabilidade	<--> Dimensão Somática	57,012	,077
Erro_Irritabilidade	<--> Erro_Cansaço ou Fadiga	15,774	,053
Erro_Agitação	<--> Dimensão Cognitivo-Afectiva	12,262	-,019
Erro_Agitação	<--> Dimensão Somática	22,738	,048
Erro_Agitação	<--> Erro_Cansaço ou Fadiga	14,082	,049
Erro_Agitação	<--> Erro_Alterações nos Hábitos do Sono	12,879	,053
Erro_Agitação	<--> Erro_Irritabilidade	27,947	,078
Erro_Pensamentos ou Desejos Suicidas	<--> Dimensão Somática	11,494	-,016
Erro_Autocrítica	<--> Erro_Desvalorização Pessoal	11,229	,040
Erro_Autocrítica	<--> Erro_Perda de Interesse	10,575	-,035
Erro_Desprezo Por Si Próprio	<--> Dimensão Somática	15,527	-,037
Erro_Desprezo Por Si Próprio	<--> Erro_Irritabilidade	27,428	-,074
Erro_Desprezo Por Si Próprio	<--> Erro_Desvalorização Pessoal	36,404	,073
Erro_Sentimentos de Punição	<--> Erro_Pensamentos ou Desejos Suicidas	15,140	,025
Erro_Sentimentos de Culpa	<--> Erro_Perda de Interesse	12,310	-,033
Erro_Sentimentos de Culpa	<--> Erro_Sentimentos de Punição	14,667	,043
Erro_Perda de Prazer	<--> Erro_Perda de Interesse	80,936	,069
Erro_Perda de Prazer	<--> Erro_Autocrítica	16,934	-,038
Erro_Perda de Prazer	<--> Erro_Sentimentos de Culpa	13,888	-,030
Erro_Fracasso no Passado	<--> Erro_Sentimentos de Culpa	25,449	,053

Variances: (Group number 1 - Default model)

	M.I.	Par Change
--	------	------------

Regression Weights: (Group number 1 - Default model)

		M.I.	Par Change
Cansaço ou Fadiga	<--- Agitação	10,454	,109
Alterações no Apetite	<--- Irritabilidade	10,432	,118
Alterações nos Hábitos do Sono	<--- Alterações no Apetite	10,359	,106
Irritabilidade	<--- Dimensão Somática	14,925	,211
Irritabilidade	<--- Cansaço ou Fadiga	24,517	,163
Irritabilidade	<--- Alterações no Apetite	17,936	,139
Irritabilidade	<--- Alterações nos Hábitos do Sono	19,636	,148
Irritabilidade	<--- Agitação	24,804	,190
Irritabilidade	<--- Desprezo por Si Próprio	14,228	-,124
Desvalorização Pessoal	<--- Desprezo por Si Próprio	18,971	,123
Perda de Interesse	<--- Perda de Prazer	39,769	,216
Agitação	<--- Cansaço ou Fadiga	14,791	,123
Agitação	<--- Alterações nos Hábitos do Sono	16,840	,133
Agitação	<--- Irritabilidade	19,875	,147
Pensamentos ou Desejos Suicidas	<--- Sentimentos de Punição	11,116	,056
Desprezo por Si Próprio	<--- Irritabilidade	19,572	-,140
Desprezo por Si Próprio	<--- Desvalorização Pessoal	19,439	,147
Sentimentos de Punição	<--- Pensamentos ou Desejos Suicidas	10,722	,227
Sentimentos de Punição	<--- Sentimentos de Culpa	11,715	,140
Sentimentos de Culpa	<--- Sentimentos de Punição	10,762	,098
Sentimentos de Culpa	<--- Fracasso no Passado	18,194	,137
Perda de Prazer	<--- Perda de Interesse	36,506	,154
Perda de Prazer	<--- Autocrítica	12,044	-,086
Perda de Prazer	<--- Sentimentos de Culpa	11,113	-,099
Fracasso no Passado	<--- Sentimentos de Culpa	20,329	,171

Minimization History (Default model)

Iteration	Negative eigenvalues	Condition #	Smallest eigenvalue	Diameter	F	NTries	Ratio
0	e 4		-,837	9999,000	3657,555	0	9999,000
1	e 2		-,170	2,222	1530,109	19	,365
2	e* 1		,000	,749	940,339	4	,862
3	e 0	145,702		,673	686,151	5	,826
4	e 0	85,065		,508	627,982	3	,000
5	e 0	215,811		,517	596,366	1	1,089
6	e 0	362,967		,201	592,164	1	1,096
7	e 0	416,949		,068	591,986	1	1,044
8	e 0	419,457		,005	591,985	1	1,004
9	e 0	419,415		,000	591,985	1	1,000

Model Fit Summary**CMIN**

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	39	591,985	151	,000	3,920
Saturated model	190	,000	0		
Independence model	19	3914,009	171	,000	22,889

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	,024	,906	,882	,720
Saturated model	,000	1,000		
Independence model	,126	,368	,298	,331

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,849	,829	,883	,867	,882
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,883	,749	,779
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

NCP

Model	NCP	LO 90	HI 90
Default model	440,985	369,966	519,569
Saturated model	,000	,000	,000
Independence model	3743,009	3543,026	3950,281

FMN

Model	FMN	F0	LO 90	HI 90
Default model	,893	,665	,558	,784
Saturated model	,000	,000	,000	,000
Independence model	5,903	5,646	5,344	5,958

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,066	,061	,072	,000
Independence model	,182	,177	,187	,000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	669,985	672,411	845,418	884,418
Saturated model	380,000	391,820	1234,674	1424,674
Independence model	3952,009	3953,191	4037,476	4056,476

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	1,011	,903	1,129	1,014
Saturated model	,573	,573	,573	,591
Independence model	5,961	5,659	6,273	5,963

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	203	218
Independence model	35	37

Anexo K: Análise Factorial Exploratória

Factor Analysis

Correlation Matrix^a

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.		
1. Tristeza	1,000																						
2. Pessimismo	,310	1,000																					
3. Fracasso no Passado	,336	,231	1,000																				
4. Perda de Prazer	,440	,313	,335	1,000																			
5. Sentimentos de Culpa	,212	,224	,364	,200	1,000																		
6. Sentimentos de Punição	,326	,228	,258	,330	,322	1,000																	
7. Desprezo Por Si Próprio	,424	,375	,379	,444	,315	,368	1,000																
8. Autocrítica	,278	,268	,277	,255	,266	,239	,379	1,000															
9. Pensamentos ou Desejos Suicidas	,330	,288	,266	,360	,283	,360	,340	,295	1,000														
10. Choro	,361	,241	,226	,352	,240	,315	,355	,305	,318	1,000													
11. Agitação	,183	,159	,153	,173	,166	,166	,160	,185	,127	,186	1,000												
12. Perda de Interesse	,461	,358	,312	,644	,219	,332	,447	,291	,369	,367	,163	1,000											
13. Indecisão	,286	,334	,297	,357	,257	,201	,338	,329	,196	,293	,213	,363	1,000										
14. Desvalorização Pessoal	,391	,352	,363	,384	,270	,269	,551	,413	,352	,380	,125	,434	,383	1,000									
15. Perda de Energia	,352	,291	,235	,393	,232	,204	,286	,262	,197	,324	,195	,427	,380	,370	1,000								
16. Alterações nos Hábitos do Sono	,294	,189	,152	,238	,210	,111	,189	,240	,154	,271	,277	,256	,242	,181	,365	1,000							
17. Irritabilidade	,350	,245	,189	,346	,148	,216	,212	,277	,208	,324	,325	,347	,284	,282	,385	,348	1,000						
18. Alterações no Apetite	,319	,141	,147	,272	,161	,222	,238	,244	,206	,263	,224	,304	,208	,214	,319	,359	,347	1,000					
19. Dificuldades de Concentração	,307	,305	,279	,339	,267	,199	,313	,246	,196	,307	,223	,415	,339	,339	,442	,310	,345	,234	1,000				
20. Cansaço ou Fadiga	,350	,280	,264	,367	,234	,222	,314	,290	,243	,347	,319	,377	,305	,321	,573	,454	,448	,307	,457	1,000			
21. Perda de Interesse Sexual	,300	,226	,213	,405	,174	,235	,330	,219	,251	,233	,187	,410	,242	,323	,288	,206	,281	,206	,256	,291	1,000		

a. Determinant = ,002

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,934
Bartlett's Test of Sphericity	Approx. Chi-Square	4248,100
	df	210,000
	Sig.	,000

Communalities

	Initial	Extraction
1. Tristeza	1,000	,475
2. Pessimismo	1,000	,378
3. Fracasso no Passado	1,000	,442
4. Perda de Prazer	1,000	,643
5. Sentimentos de Culpa	1,000	,584
6. Sentimentos de Punição	1,000	,550
7. Desprezo Por Si Próprio	1,000	,567
8. Autocrítica	1,000	,398
9. Pensamentos ou Desejos Suicidas	1,000	,498
10. Choro	1,000	,374
11. Agitação	1,000	,427
12. Perda de Interesse	1,000	,671
13. Indecisão	1,000	,481
14. Desvalorização Pessoal	1,000	,554
15. Perda de Energia	1,000	,583
16. Alterações nos Hábitos do Sono	1,000	,530
17. Irritabilidade	1,000	,507
18. Alterações no Apetite	1,000	,484
19. Dificuldades de Concentração	1,000	,502
20. Cansaço ou Fadiga	1,000	,602
21. Perda de Interesse Sexual	1,000	,387

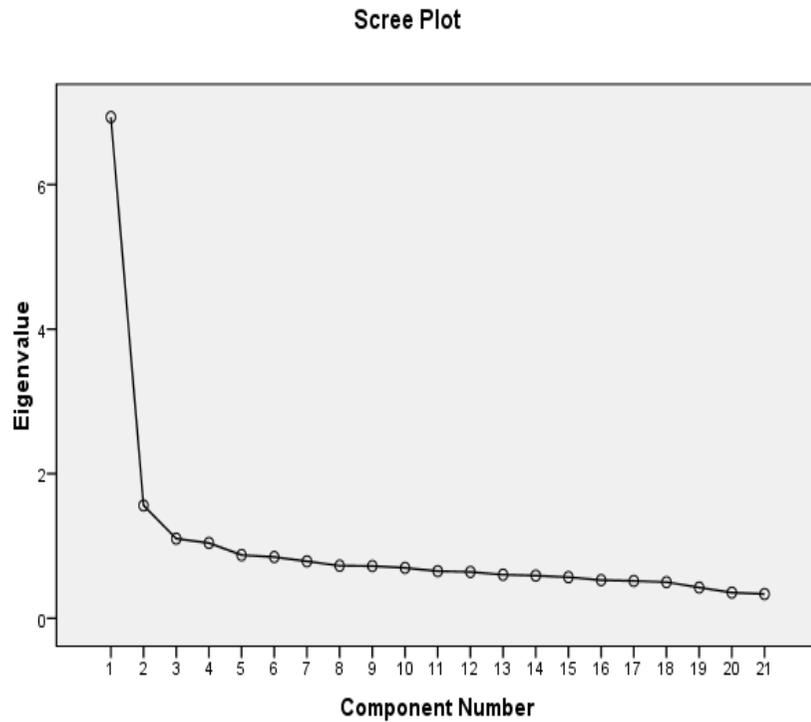
Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation ^a
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6,934	33,020	33,020	6,934	33,020	33,020	5,366
2	1,560	7,428	40,448	1,560	7,428	40,448	3,745
3	1,101	5,241	45,689	1,101	5,241	45,689	4,202
4	1,042	4,961	50,650	1,042	4,961	50,650	3,628
5	,874	4,163	54,813				
6	,846	4,029	58,842				
7	,788	3,753	62,595				
8	,726	3,459	66,054				
9	,722	3,439	69,493				
10	,696	3,316	72,809				
11	,652	3,106	75,915				
12	,641	3,051	78,966				
13	,602	2,868	81,834				
14	,590	2,809	84,643				
15	,568	2,707	87,350				
16	,526	2,507	89,857				
17	,515	2,454	92,310				
18	,500	2,380	94,690				
19	,425	2,024	96,714				
20	,355	1,690	98,404				
21	,335	1,596	100,000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

**Component Matrix^a**

	Component			
	1	2	3	4
1. Tristeza	,645	-,062	-,164	,168
2. Pessimismo	,531	-,149	-,013	-,272
3. Fracasso no Passado	,519	-,285	,262	-,151
4. Perda de Prazer	,682	-,119	-,401	,060
5. Sentimentos de Culpa	,463	-,212	,570	-,008
6. Sentimentos de Punição	,502	-,304	,129	,434
7. Desprezo Por Si Próprio	,659	-,360	,005	-,055
8. Autocrítica	,542	-,112	,298	-,053
9. Pensamentos ou Desejos Suicidas	,525	-,332	,036	,333
10. Choro	,588	-,017	,029	,164
11. Agitação	,380	,372	,325	,196
12. Perda de Interesse	,713	-,105	-,391	,008
13. Indecisão	,573	,000	,085	-,380
14. Desvalorização Pessoal	,656	-,280	-,019	-,210
15. Perda de Energia	,639	,319	-,111	-,247
16. Alterações nos Hábitos do Sono	,492	,507	,160	,071
17. Irritabilidade	,574	,396	-,053	,131
18. Alterações no Apetite	,484	,311	,011	,391
19. Dificuldades de Concentração	,600	,205	,008	-,316
20. Cansaço ou Fadiga	,653	,405	,033	-,103
21. Perda de Interesse Sexual	,524	-,038	-,323	,084

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Pattern Matrix^a

	Component			
	1	2	3	4
1. Tristeza	,563	,136	,038	,077
2. Pessimismo	,176	-,125	,474	,179
3. Fracasso no Passado	-,011	-,094	,339	,531
4. Perda de Prazer	,796	-,043	,151	-,126
5. Sentimentos de Culpa	-,318	,138	,186	,791
6. Sentimentos de Punição	,404	,110	-,311	,488
7. Desprezo Por Si Próprio	,392	-,174	,267	,376
8. Autocrítica	-,054	,130	,258	,491
9. Pensamentos ou Desejos Suicidas	,474	,008	-,200	,406
10. Choro	,317	,232	,035	,222
11. Agitação	-,210	,655	-,025	,241
12. Perda de Interesse	,769	-,042	,221	-,121
13. Indecisão	-,012	,008	,625	,189
14. Desvalorização Pessoal	,315	-,184	,440	,287
15. Perda de Energia	,171	,308	,523	-,138
16. Alterações nos Hábitos do Sono	-,095	,689	,155	,023
17. Irritabilidade	,233	,561	,098	-,091
18. Alterações no Apetite	,278	,604	-,221	,028
19. Dificuldades de Concentração	,038	,205	,579	,018
20. Cansaço ou Fadiga	,064	,511	,386	-,023
21. Perda de Interesse Sexual	,626	,028	,079	-,134

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 11 iterations.

Structure Matrix

	Component			
	1	2	3	4
1. Tristeza	,672	,393	,357	,391
2. Pessimismo	,423	,161	,549	,356
3. Fracasso no Passado	,363	,140	,431	,590
4. Perda de Prazer	,784	,308	,456	,295
5. Sentimentos de Culpa	,209	,245	,294	,710
6. Sentimentos de Punição	,551	,257	,030	,633
7. Desprezo Por Si Próprio	,625	,162	,468	,599
8. Autocrítica	,354	,307	,405	,555
9. Pensamentos ou Desejos Suicidas	,589	,208	,114	,591
10. Choro	,534	,419	,317	,435
11. Agitação	,161	,612	,188	,268
12. Perda de Interesse	,790	,325	,515	,303
13. Indecisão	,361	,277	,670	,339
14. Desvalorização Pessoal	,577	,168	,581	,513
15. Perda de Energia	,458	,545	,681	,140
16. Alterações nos Hábitos do Sono	,262	,714	,378	,158
17. Irritabilidade	,457	,672	,390	,165
18. Alterações no Apetite	,437	,639	,136	,236
19. Dificuldades de Concentração	,386	,441	,678	,223
20. Cansaço ou Fadiga	,430	,678	,602	,211
21. Perda de Interesse Sexual	,606	,281	,333	,199

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

Component Correlation Matrix

Component	1	2	3	4
1	1,000			
2	,402	1,000		
3	,442	,376	1,000	
4	,492	,208	,248	1,000

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

Component Score Covariance Matrix

Component	1	2	3	4
1	2,042			
2	1,449	1,497		
3	2,742	1,674	3,700	
4	1,795	1,144	2,216	2,173

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

Component Score Coefficient Matrix

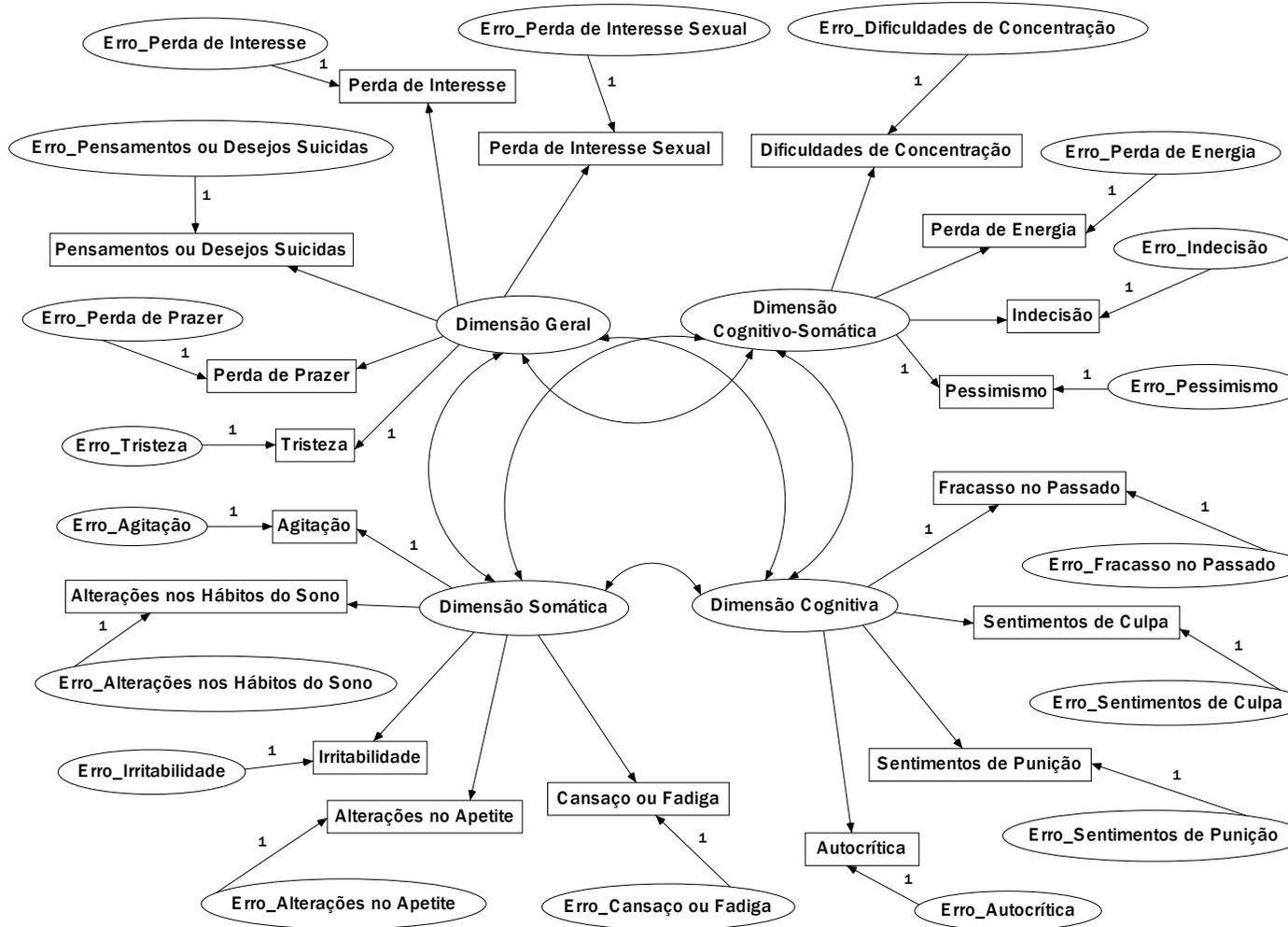
	Component			
	1	2	3	4
1. Tristeza	,188	0,53	-,039	,032
2. Pessimismo	,026	-,092	,217	,040
3. Fracasso no Passado	-,028	-,079	,117	,240
4. Perda de Prazer	,262	-,034	,036	-,083
5. Sentimentos de Culpa	-,124	,033	,016	,389
6. Sentimentos de Punição	,161	,048	-,257	,291
7. Desprezo Por Si Próprio	,116	-,112	,072	,171
8. Autocrítica	-,039	,032	,071	,223
9. Pensamentos ou Desejos Suicidas	,177	-,003	-,189	,237
10. Choro	,103	,096	-,046	,105
11. Agitação	-,078	,302	-,062	,109
12. Perda de Interesse	,247	-,037	,074	-,089
13. Indecisão	-,051	-,036	,301	,023
14. Desvalorização Pessoal	,077	-,123	,179	,103
15. Perda de Energia	,015	,118	,258	-,146
16. Alterações nos Hábitos do Sono	-,052	,314	,047	-,028
17. Irritabilidade	,066	,256	,014	-,077
18. Alterações no Apetite	,105	,290	-,172	,023
19. Dificuldades de Concentração	-,033	,063	,284	-,067
20. Cansaço ou Fadiga	-,013	,217	,170	-,075
21. Perda de Interesse Sexual	,208	,005	,009	-,081

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

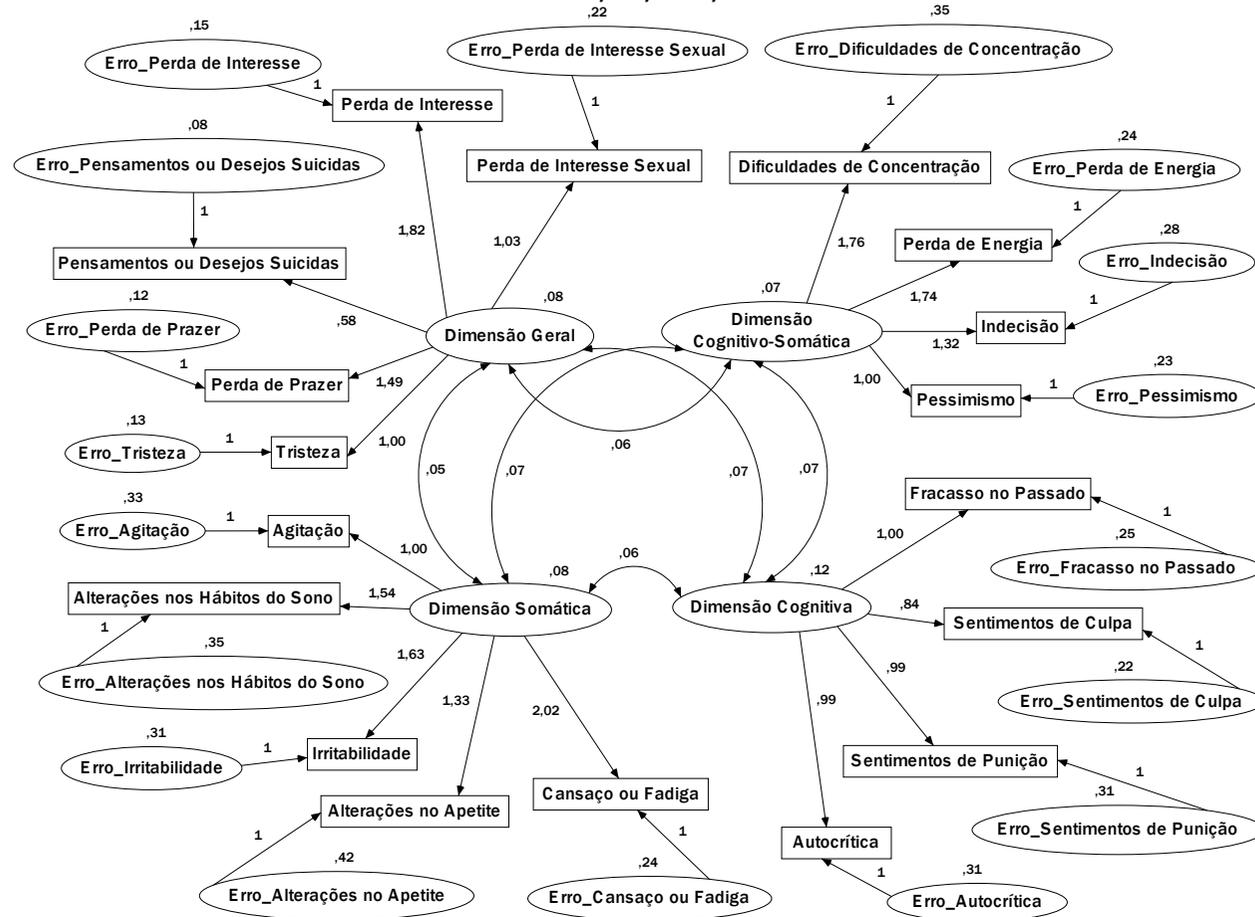
Anexo L: Análise Factorial Confirmatória

Modelo Factorial da População Académica do ISPA



Análise Factorial Confirmatória do Modelo da População Académica do ISPA

X²=336,621; df=129; X²/df=2,609
 Rmse=,049; p(rmse<0,05)=,563
 CFI=,935; GFI=,947



Amos Output

Analysis Summary

Groups

Group number 1 (Group number 1)

Notes for Group (Group number 1)

The model is recursive.

Sample size = 664

Variable Summary (Group number 1)

Your model contains the following variables (Group number 1)

Observed, endogenous variables:

Tristeza
 Perda de Prazer
 Pensamentos ou Desejos Suicidas
 Perda de Interesse
 Perda de Interesse Sexual
 Agitação
 Alterações nos Hábitos do Sono
 Irritabilidade
 Alterações no Apetite
 Cansaço ou Fadiga
 Pessimismo
 Indecisão
 Perda de Energia
 Dificuldades de Concentração
 Fracasso no Passado
 Sentimentos de Culpa
 Sentimentos de Punição
 Autocrítica

Unobserved, exogenous variables:

Dimensão Geral
 Erro_Tristeza
 Erro_Perda de Prazer
 Erro_Pensamentos ou Desejos Suicidas
 Erro_Perda de Interesse
 Erro_Perda de Interesse Sexual
 Dimensão Somática
 Erro_Agitação
 Erro_Alterações nos Hábitos do Sono
 Erro_Irritabilidade
 Erro_Alterações no Apetite
 Erro_Cansaço ou Fadiga
 Dimensão Cognitivo-Somática
 Erro_Pessimismo
 Erro_Indecisão
 Erro_Perda de Energia
 Erro_Dificuldades de Concentração
 Dimensão Cognitiva
 Erro_Fracasso no Passado
 Erro_Sentimentos de Culpa
 Erro_Sentimentos de Punição
 Erro_Autocrítica

Variable counts (Group number 1)

Number of variables in your model: 40
 Number of observed variables: 18
 Number of unobserved variables: 22
 Number of exogenous variables: 22
 Number of endogenous variables: 18

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	22	0	0	0	0	22
Labeled	0	0	0	0	0	0
Unlabeled	14	6	22	0	0	42
Total	36	6	22	0	0	64

Models

Default model (Default model)

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	171
Number of distinct parameters to be estimated:	42
Degrees of freedom (171 - 42):	129

Result (Default model)

Minimum was achieved

Chi-square = 336,621

Degrees of freedom = 129

Probability level = ,000

Group number 1 (Group number 1 - Default model)

Estimates (Group number 1 - Default model)

Scalar Estimates (Group number 1 - Default model)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
Tristeza	<--- Dimensão Geral	1,000				
Perda de Prazer	<--- Dimensão Geral	1,486	,100	14,889	***	
Pensamentos ou Desejos Suicidas	<--- Dimensão Geral	,577	,054	10,777	***	
Perda de Interesse	<--- Dimensão Geral	1,816	,119	15,237	***	
Perda de Interesse Sexual	<--- Dimensão Geral	1,034	,092	11,277	***	
Agitação	<--- Dimensão Somática	1,000				
Alterações nos Hábitos do Sono	<--- Dimensão Somática	1,541	,169	9,139	***	
Irritabilidade	<--- Dimensão Somática	1,630	,173	9,431	***	
Alterações no Apetite	<--- Dimensão Somática	1,332	,158	8,410	***	
Cansaço ou Fadiga	<--- Dimensão Somática	2,017	,201	10,023	***	
Pessimismo	<--- Dimensão Cognitivo-Somática	1,000				
Indecisão	<--- Dimensão Cognitivo-Somática	1,316	,132	9,966	***	
Perda de Energia	<--- Dimensão Cognitivo-Somática	1,739	,156	11,176	***	
Dificuldades de Concentração	<--- Dimensão Cognitivo-Somática	1,756	,165	10,649	***	
Fracasso no Passado	<--- Dimensão Cognitiva	1,000				
Sentimentos de Culpa	<--- Dimensão Cognitiva	,836	,086	9,714	***	
Sentimentos de Punição	<--- Dimensão Cognitiva	,988	,102	9,669	***	
Autocrítica	<--- Dimensão Cognitiva	,990	,103	9,630	***	

Standardized Regression Weights: (Group number 1 - Default model)

		Estimate
Tristeza	<--- Dimensão Geral	,615
Perda de Prazer	<--- Dimensão Geral	,761
Pensamentos ou Desejos Suicidas	<--- Dimensão Geral	,495
Perda de Interesse	<--- Dimensão Geral	,792
Perda de Interesse Sexual	<--- Dimensão Geral	,523
Agitação	<--- Dimensão Somática	,432
Alterações nos Hábitos do Sono	<--- Dimensão Somática	,584
Irritabilidade	<--- Dimensão Somática	,628
Alterações no Apetite	<--- Dimensão Somática	,493
Cansaço ou Fadiga	<--- Dimensão Somática	,751
Pessimismo	<--- Dimensão Cognitivo-Somática	,488
Indecisão	<--- Dimensão Cognitivo-Somática	,554
Perda de Energia	<--- Dimensão Cognitivo-Somática	,691
Dificuldades de Concentração	<--- Dimensão Cognitivo-Somática	,625
Fracasso no Passado	<--- Dimensão Cognitiva	,571
Sentimentos de Culpa	<--- Dimensão Cognitiva	,529
Sentimentos de Punição	<--- Dimensão Cognitiva	,525
Autocrítica	<--- Dimensão Cognitiva	,522

Covariances: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
Dimensão Geral	<--> Dimensão Somática	,052	,007	7,628	***	
Dimensão Somática	<--> Dimensão Cognitivo-Somática	,065	,009	7,459	***	
Dimensão Cognitivo-Somática	<--> Dimensão Cognitiva	,070	,009	7,867	***	
Dimensão Geral	<--> Dimensão Cognitiva	,072	,008	8,819	***	
Dimensão Somática	<--> Dimensão Cognitiva	,060	,009	6,949	***	
Dimensão Geral	<--> Dimensão Cognitivo-Somática	,060	,007	8,636	***	

Correlations: (Group number 1 - Default model)

		Estimate
Dimensão Geral	<--> Dimensão Somática	,677
Dimensão Somática	<--> Dimensão Cognitivo-Somática.	,880
Dimensão Cognitivo-Somática.	<--> Dimensão Cognitiva	,750
Dimensão Geral	<--> Dimensão Cognitiva	,748
Dimensão Somática	<--> Dimensão Cognitiva	,628
Dimensão Geral	<--> Dimensão Cognitivo-Somática	,801

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Dimensão Geral	,077	,009	8,142	***	
Dimensão Somática	,076	,014	5,292	***	
Dimensão Cognitivo-Somática.	,072	,012	6,130	***	
Dimensão Cognitiva	,120	,018	6,764	***	
Erro_Tristeza	,127	,008	16,215	***	
Erro_Perda de Prazer	,124	,009	13,596	***	
Erro_Pensamentos ou Desejos Suicidas	,079	,005	17,150	***	
Erro_Perda de Interesse	,151	,012	12,558	***	
Erro_Perda de Interesse Sexual	,219	,013	16,980	***	
Erro_Agitação	,330	,019	17,261	***	
Erro_Alterações nos Hábitos do Sono	,348	,022	16,050	***	
Erro_Irritabilidade	,308	,020	15,464	***	
Erro_Alterações no Apetite	,419	,025	16,881	***	
Erro_Cansaço ou Fadiga	,238	,019	12,645	***	
Erro_Pessimismo	,232	,014	17,071	***	
Erro_Indecisão	,284	,017	16,562	***	
Erro_Perda de Energia	,240	,017	14,487	***	
Erro_Dificuldades de Concentração	,348	,022	15,732	***	
Erro_Fracasso no Passado	,248	,017	14,896	***	
Erro_Sentimentos de Culpa	,216	,014	15,578	***	
Erro_Sentimentos de Punição	,308	,020	15,630	***	
Erro_Autocrítica	,314	,020	15,676	***	

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
Tristeza	,378
Perda de Prazer	,579
Pensamentos ou Desejos Suicidas	,245
Perda de Interesse	,628
Perda de Interesse Sexual	,274
Agitação	,187
Alterações nos Hábitos do Sono	,341
Irritabilidade	,395
Alterações no Apetite	,243
Cansaço ou Fadiga	,564
Pessimismo	,238
Indecisão	,306
Perda de Energia	,478
Dificuldades de Concentração	,391
Fracasso no Passado	,326
Sentimentos de Culpa	,280
Sentimentos de Punição	,276
Autocrítica	,273

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

		M.I.	Par Change
Erro_Sentimentos de Punição	<--> Dimensão Geral	14,246	,019
Erro_Sentimentos de Culpa	<--> Dimensão Geral	14,112	-,016
Erro_Perda de Energia	<--> Dimensão Somática	11,933	,016
Erro_Cansaço ou Fadiga	<--> Dimensão Cognitivo-Somática	11,583	,016
Erro_Cansaço ou Fadiga	<--> Erro_Perda de Energia	33,781	,066
Erro_Alterações no Apetite	<--> Erro_Cansaço ou Fadiga	10,978	-,048
Erro_Pensamentos ou Desejos Suicidas	<--> Dimensão Cognitiva	35,993	,024
Erro_Pensamentos ou Desejos Suicidas	<--> Erro_Sentimentos de Punição	17,236	,027
Erro_Perda de Prazer	<--> Erro_Perda de Interesse	11,000	,022

Variances: (Group number 1 - Default model)

	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

		M.I.	Par Change
Sentimentos de Punição	<--- Pensamentos ou Desejos Suicidas	17,002	,289
Perda de Energia	<--- Cansaço ou Fadiga	17,331	,116
Cansaço ou Fadiga	<--- Perda de Energia	18,486	,138
Pensamentos ou Desejos Suicidas	<--- Sentimentos de Punição	19,822	,076
Pensamentos ou Desejos Suicidas	<--- Sentimentos de Culpa	11,942	,071

Minimization History (Default model)

Iteration		Negative eigenvalues	Condition #	Smallest eigenvalue	Diameter	F	NTries	Ratio
0	e	10		-,422	9999,000	3528,926	0	9999,000
1	e	5		-,163	2,342	1493,512	20	,546
2	e*	0	1054,149		,808	931,658	5	,719
3	e	0	290,923		,949	586,038	3	,000
4	e	0	595,137		1,117	383,682	1	,920
5	e	0	1095,047		,608	351,877	1	,744
6	e	0	2495,129		,246	337,120	1	1,076
7	e	0	3319,671		,100	336,628	1	1,062
8	e	0	3577,398		,019	336,621	1	1,019
9	e	0	3599,371		,001	336,621	1	1,001

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	42	336,621	129	,000	2,609
Saturated model	171	,000	0		
Independence model	18	3330,579	153	,000	21,768

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	,016	,947	,929	,714

Model	RMR	GFI	AGFI	PGFI
Saturated model	,000	1,000		
Independence model	,111	,407	,337	,364

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,899	,880	,935	,923	,935
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,843	,758	,788
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

NCP

Model	NCP	LO 90	HI 90
Default model	207,621	157,145	265,770
Saturated model	,000	,000	,000
Independence model	3177,579	2993,508	3368,964

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	,508	,313	,237	,401
Saturated model	,000	,000	,000	,000
Independence model	5,023	4,793	4,515	5,081

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,049	,043	,056	,563
Independence model	,177	,172	,182	,000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	420,621	423,099	609,549	651,549
Saturated model	342,000	352,090	1111,206	1282,206
Independence model	3366,579	3367,641	3447,548	3465,548

ECVI

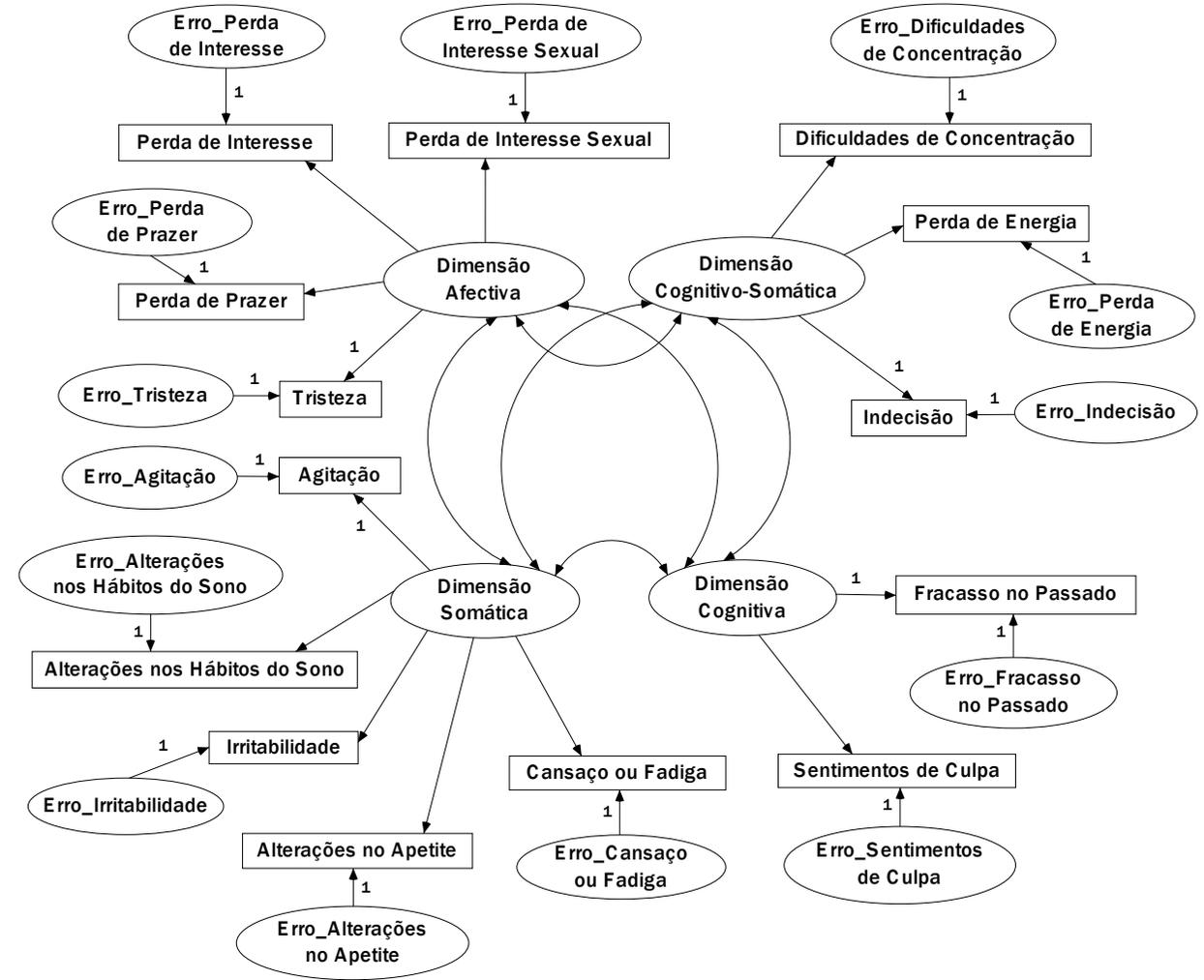
Model	ECVI	LO 90	HI 90	MECVI
Default model	,634	,558	,722	,638
Saturated model	,516	,516	,516	,531
Independence model	5,078	4,800	5,366	5,079

HOELTER

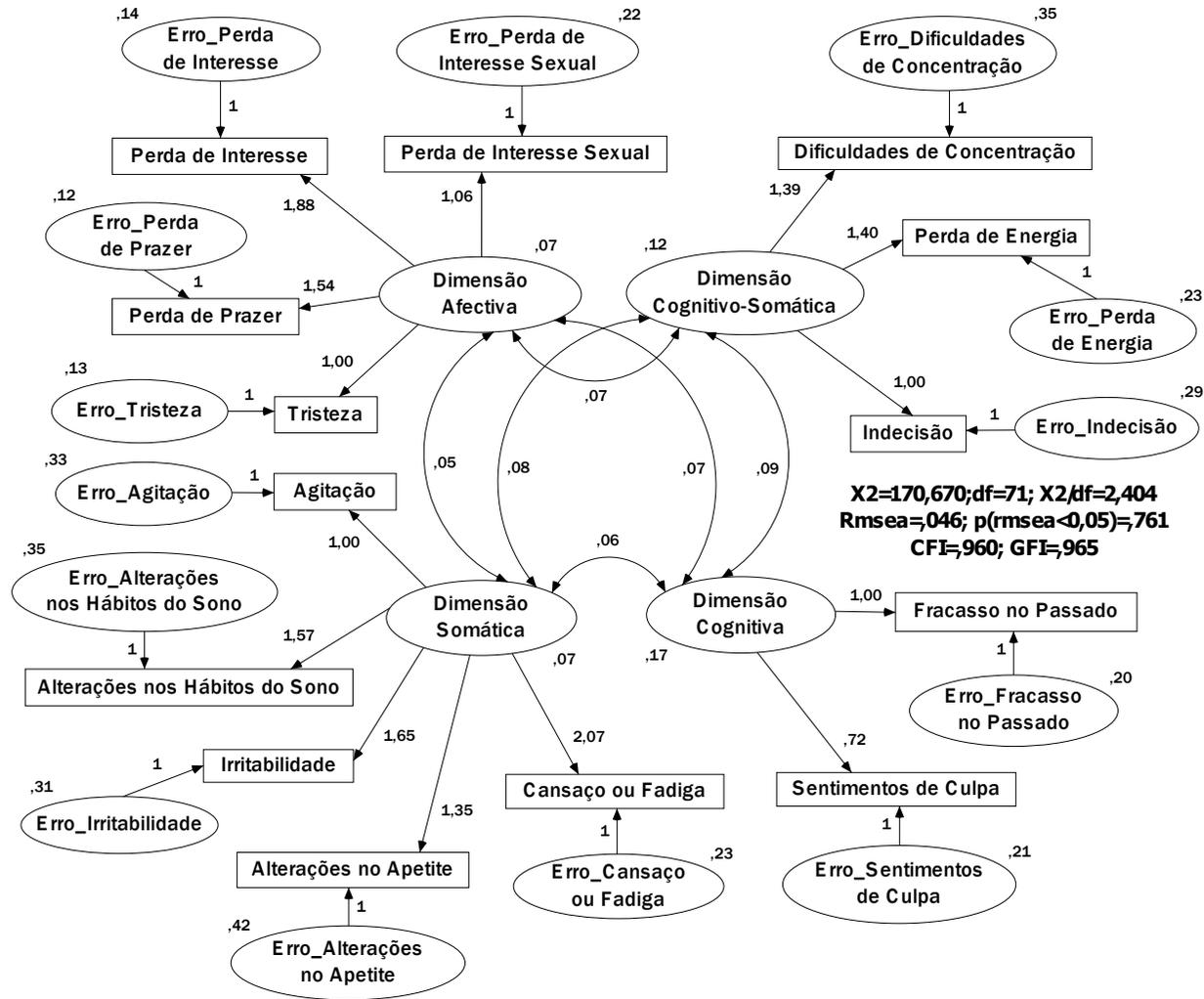
Model	HOELTER .05	HOELTER .01
Default model	309	334
Independence model	37	40

Anexo M: Análise Factorial Confirmatória com Itens com Pesos Factoriais nos Factores
Iguais ou Superiores a 0,50

Modelo Factorial da População Académica do ISPA



Análise Factorial Confirmatória do Modelo da População Académica do ISPA



Amos Output

Analysis Summary

Groups

Group number 1 (Group number 1)

Notes for Group (Group number 1)

The model is recursive.

Sample size = 664

Variable Summary (Group number 1)

Your model contains the following variables (Group number 1)

Observed, endogenous variables:

Tristeza
Perda de Prazer
Perda de Interesse
Perda de Interesse Sexual
Agitação
Alterações nos Hábitos do Sono
Irritabilidade
Alterações no Apetite
Cansaço ou Fadiga
Indecisão
Perda de Energia
Dificuldades de Concentração
Fracasso no Passado
Sentimentos de Culpa

Unobserved, exogenous variables :

Dimensão Afectiva
Erro_Tristeza
Erro_Perda de Prazer
Erro_Perda de Interesse
Erro_Perda de Interesse Sexual
Dimensão Somática
Erro_Agitação
Erro_Alterações nos Hábitos do Sono
Erro_Irritabilidade
Erro_Alterações no Apetite
Erro_Cansaço ou Fadiga
Dimensão Cognitivo-Somática
Erro_Indecisão
Erro_Perda de Energia
Erro_Dificuldades de Concentração
Dimensão Cognitiva
Erro_Fracasso no Passado
Erro_Sentimentos de Culpa

Variable counts (Group number 1)

Number of variables in your model: 32
Number of observed variables: 14
Number of unobserved variables: 18
Number of exogenous variables: 18
Number of endogenous variables: 14

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	18	0	0	0	0	18
Labeled	0	0	0	0	0	0
Unlabeled	10	6	18	0	0	34
Total	28	6	18	0	0	52

Models

Default model (Default model)

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	105
Number of distinct parameters to be estimated:	34
Degrees of freedom (105 - 34):	71

Result (Default model)

Minimum was achieved
 Chi-square = 170,670
 Degrees of freedom = 71
 Probability level = ,000

Group number 1 (Group number 1 - Default model)

Estimates (Group number 1 - Default model)

Scalar Estimates (Group number 1 - Default model)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
Tristeza	<--- Dimensão Afectiva	1,000				
Perda de Prazer	<--- Dimensão Afectiva	1,542	,107	14,436	***	
Perda de Interesse	<--- Dimensão Afectiva	1,885	,128	14,700	***	
Perda de Interesse Sexual	<--- Dimensão Afectiva	1,057	,096	11,005	***	
Agitação	<--- Dimensão Somática	1,000				
Alterações nos Hábitos do Sono	<--- Dimensão Somática	1,568	,174	9,034	***	
Irritabilidade	<--- Dimensão Somática	1,646	,177	9,285	***	
Alterações no Apetite	<--- Dimensão Somática	1,349	,162	8,308	***	
Cansaço ou Fadiga	<--- Dimensão Somática	2,067	,209	9,898	***	
Indecisão	<--- Dimensão Cognitivo-Somática	1,000				
Perda de Energia	<--- Dimensão Cognitivo-Somática	1,399	,114	12,276	***	
Dificuldades de Concentração	<--- Dimensão Cognitivo-Somática	1,387	,121	11,504	***	
Fracasso no Passado	<--- Dimensão Cognitiva	1,000				
Sentimentos de Culpa	<--- Dimensão Cognitiva	,723	,088	8,184	***	

Standardized Regression Weights: (Group number 1 - Default model)

		Estimate
Tristeza	<--- Dimensão Afectiva	,600
Perda de Prazer	<--- Dimensão Afectiva	,771
Perda de Interesse	<--- Dimensão Afectiva	,803
Perda de Interesse Sexual	<--- Dimensão Afectiva	,522
Agitação	<--- Dimensão Somática	,425
Alterações nos Hábitos do Sono	<--- Dimensão Somática	,585
Irritabilidade	<--- Dimensão Somática	,624
Alterações no Apetite	<--- Dimensão Somática	,491

		Estimate
Cansaço ou Fadiga	<--- Dimensão Somática	,757
Indecisão	<--- Dimensão Cognitivo-Somática	,535
Perda de Energia	<--- Dimensão Cognitivo-Somática	,707
Dificuldades de Concentração	<--- Dimensão Cognitivo-Somática	,628
Fracasso no Passado	<--- Dimensão Cognitiva	,674
Sentimentos de Culpa	<--- Dimensão Cognitiva	,540

Covariances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Dimensão Cognitivo-Somática <--> Dimensão Cognitiva	,091	,012	7,761	***	
Dimensão Somática <--> Dimensão Cognitivo-Somática	,084	,011	7,692	***	
Dimensão Afectiva <--> Dimensão Somática	,049	,007	7,469	***	
Dimensão Afectiva <--> Dimensão Cognitiva	,067	,008	8,050	***	
Dimensão Afectiva <--> Dimensão Cognitivo-Somática	,073	,008	8,850	***	
Dimensão Somática <--> Dimensão Cognitiva	,056	,009	6,304	***	

Correlations: (Group number 1 - Default model)

		Estimate
Dimensão Cognitivo-Somática <--> Dimensão Cognitiva		,646
Dimensão Somática <--> Dimensão Cognitivo-Somática		,901
Dimensão Afectiva <--> Dimensão Somática		,671
Dimensão Afectiva <--> Dimensão Cognitiva		,600
Dimensão Afectiva <--> Dimensão Cognitivo-Somática		,782
Dimensão Somática <--> Dimensão Cognitiva		,507

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Dimensão Afectiva	,074	,009	7,844	***	
Dimensão Somática	,073	,014	5,203	***	
Dimensão Cognitivo-Somática	,117	,017	6,780	***	
Dimensão Cognitiva	,167	,026	6,438	***	
Tristeza	,131	,008	16,190	***	
Perda de Prazer	,119	,009	12,735	***	
Perda de Interesse	,144	,013	11,514	***	
Perda de Interesse Sexual	,219	,013	16,874	***	
Agitação	,332	,019	17,308	***	
Alterações nos Hábitos do Sono	,347	,022	16,063	***	
Irritabilidade	,311	,020	15,552	***	
Alterações no Apetite	,420	,025	16,908	***	
Cansaço ou Fadiga	,233	,019	12,482	***	
Indecisão	,292	,018	16,623	***	
Perda de Energia	,229	,017	13,611	***	
Dificuldades de Concentração	,346	,022	15,470	***	
Fracasso no Passado	,201	,023	8,907	***	
Sentimentos de Culpa	,213	,016	13,655	***	

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
Tristeza	,360
Perda de Prazer	,595
Perda de Interesse	,644
Perda de Interesse Sexual	,273
Agitação	,181
Alterações nos Hábitos do Sono	,342

	Estimate
Irritabilidade	,390
Alterações no Apetite	,241
Cansaço ou Fadiga	,573
Indecisão	,287
Perda de Energia	,500
Dificuldades de Concentração	,395
Fracasso no Passado	,454
Sentimentos de Culpa	,292

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

	M.I.	Par Change
Erro_Cansaço ou Fadiga <--> Dimensão Cognitivo-Somática	11,155	,021
Erro_Cansaço ou Fadiga <--> Erro_Perda de Energia	24,228	,055
Erro_Alterações no Apetite <--> Erro_Cansaço ou Fadiga	11,831	-,049
Erro_Agitação <--> Erro_Perda de Energia	10,007	-,038
Erro_Tristeza <--> Dimensão Somática	12,615	,012

Variances: (Group number 1 - Default model)

	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

	M.I.	Par Change
Perda de Energia <--- Cansaço ou Fadiga	10,874	,091
Cansaço ou Fadiga <--- Perda de Energia	12,820	,114
Tristeza <--- Alterações no Apetite	10,974	,066

Minimization History (Default model)

Iteration	Negative eigenvalues	Condition #	Smallest eigenvalue	Diameter	F	NTries	Ratio
0	e 10		-,367	9999,000	2884,593	0	9999,000
1	e 3		-,143	2,113	1149,711	20	,606
2	e 0	368,990		,764	623,453	5	,773
3	e 0	161,972		,817	379,571	3	,000
4	e 0	523,329		,973	206,061	1	,935
5	e 0	732,485		,366	182,176	2	,000
6	e 0	1897,869		,294	171,443	1	1,098
7	e 0	2799,005		,169	170,705	1	1,061
8	e 0	3253,310		,033	170,670	1	1,029
9	e 0	3346,712		,003	170,670	1	1,003
10	e 0	3353,541		,000	170,670	1	1,000

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	34	170,670	71	,000	2,404
Saturated model	105	,000	0		
Independence model	14	2607,114	91	,000	28,650

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	,015	,965	,948	,652
Saturated model	,000	1,000		
Independence model	,124	,439	,353	,381

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,935	,916	,961	,949	,960
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,780	,729	,749
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

NCP

Model	NCP	LO 90	HI 90
Default model	99,670	65,214	141,830
Saturated model	,000	,000	,000
Independence model	2516,114	2353,223	2686,340

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	,257	,150	,098	,214
Saturated model	,000	,000	,000	,000
Independence model	3,932	3,795	3,549	4,052

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,046	,037	,055	,761
Independence model	,204	,197	,211	,000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	238,670	240,244	391,612	425,612
Saturated model	210,000	214,861	682,320	787,320
Independence model	2635,114	2635,763	2698,090	2712,090

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	,360	,308	,424	,362
Saturated model	,317	,317	,317	,324
Independence model	3,975	3,729	4,231	3,976

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	357	395
Independence model	30	32

Anexo N: Correlações Inter-Itens dos Factores

Nonparametric Correlations (Inter-Items do Factor 1 e dos Restantes Factores)

		Correlations					
		Factor 1					
		1.	4.	9.	12.	21.	
Spearman's rho	Factor 1	1. Tristeza	1,000				
		4. Perda de Prazer	,389(**)	1,000			
		9. Pensamentos ou Desejos Suicidas	,255(**)	,241(**)	1,000		
		12. Perda de Interesse	,431(**)	,600(**)	,265(**)	1,000	
		21. Perda de Interesse Sexual	,263(**)	,302(**)	,161(**)	,296(**)	1,000
	Factor 2	11. Agitação	,154(**)	,138(**)	,090(*)	,146(**)	,124(**)
		16. Alterações nos Hábitos do Sono	,263(**)	,192(**)	,103(**)	,236(**)	,176(**)
		17. Irritabilidade	,320(**)	,266(**)	,170(**)	,307(**)	,221(**)
		18. Alterações no Apetite	,258(**)	,243(**)	,176(**)	,266(**)	,181(**)
		20. Cansaço ou Fadiga	,309(**)	,295(**)	,204(**)	,347(**)	,230(**)
	Factor 3	2. Pessimismo	,298(**)	,267(**)	,228(**)	,318(**)	,190(**)
		13. Indecisão	,240(**)	,286(**)	,147(**)	,306(**)	,164(**)
		15. Perda de Energia	,324(**)	,378(**)	,162(**)	,426(**)	,234(**)
		19. Dificuldades de Concentração	,306(**)	,297(**)	,177(**)	,380(**)	,238(**)
	Factor 4	3. Fracasso no Passado	,293(**)	,221(**)	,227(**)	,224(**)	,114(**)
		5. Sentimentos de Culpa	,188(**)	,153(**)	,249(**)	,145(**)	,105(**)
		6. Sentimentos de Punição	,295(**)	,245(**)	,236(**)	,231(**)	,156(**)
		8. Autocrítica	,224(**)	,181(**)	,235(**)	,231(**)	,137(**)

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Nonparametric Correlations (Inter-Items do Factor 2 e dos Restantes Factores)

		Correlations				
		Factor 2				
		11.	16.	17.	18.	20.
Spearman's rho	Factor 2					
	11. Agitação	1,000				
	16. Alterações nos Hábitos do Sono	,282(**)	1,000			
	17. Irritabilidade	,337(**)	,326(**)	1,000		
	18. Alterações no Apetite	,222(**)	,343(**)	,315(**)	1,000	
	20. Cansaço ou Fadiga	,310(**)	,428(**)	,423(**)	,301(**)	1,000
	Factor 1					
	1. Tristeza	,154(**)	,263(**)	,320(**)	,258(**)	,309(**)
	4. Perda de Prazer	,138(**)	,192(**)	,266(**)	,243(**)	,295(**)
	9. Pensamentos ou Desejos Suicidas	,090(*)	,103(**)	,170(**)	,176(**)	,204(**)
	12. Perda de Interesse	,146(**)	,236(**)	,307(**)	,266(**)	,347(**)
	21. Perda de Interesse Sexual	,124(**)	,176(**)	,221(**)	,181(**)	,230(**)
	Factor 3					
	2. Pessimismo	,151(**)	,168(**)	,244(**)	,129(**)	,251(**)
	13. Indecisão	,171(**)	,226(**)	,274(**)	,196(**)	,253(**)
	15. Perda de Energia	,211(**)	,358(**)	,374(**)	,288(**)	,570(**)
	19. Dificuldades de Concentração	,216(**)	,310(**)	,322(**)	,232(**)	,435(**)
	Factor 4					
	3. Fracasso no Passado	,107(**)	,103(**)	,153(**)	,129(**)	,209(**)
	5. Sentimentos de Culpa	,107(**)	,169(**)	,117(**)	,128(**)	,192(**)
6. Sentimentos de Punição	,157(**)	,091(*)	,194(**)	,189(**)	,177(**)	
8. Autocrítica	,159(**)	,193(**)	,267(**)	,197(**)	,258(**)	

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Nonparametric Correlations (Inter-Items do Factor 3 e dos Restantes Factores)

		Correlations					
		Factor 3					
		2.	13.	15.	19.		
Spearman's rho	Factor 3	2. Pessimismo	1,000				
		13. Indecisão	,321(**)	1,000			
		15. Perda de Energia	,269(**)	,329(**)	1,000		
		19. Dificuldades de Concentração	,300(**)	,307(**)	,437(**)	1,000	
	Factor 1	1. Tristeza	,298(**)	,240(**)	,324(**)	,306(**)	
		4. Perda de Prazer	,267(**)	,286(**)	,378(**)	,297(**)	
		9. Pensamentos ou Desejos Suicidas	,228(**)	,147(**)	,162(**)	,177(**)	
		12. Perda de Interesse	,318(**)	,306(**)	,426(**)	,380(**)	
		21. Perda de Interesse Sexual	,190(**)	,164(**)	,234(**)	,238(**)	
	Factor 2	11. Agitação	,151(**)	,171(**)	,211(**)	,216(**)	
		16. Alterações nos Hábitos do Sono	,168(**)	,226(**)	,358(**)	,310(**)	
		17. Irritabilidade	,244(**)	,274(**)	,374(**)	,322(**)	
		18. Alterações no Apetite	,129(**)	,196(**)	,288(**)	,232(**)	
	Factor 4	20. Cansaço ou Fadiga	,251(**)	,253(**)	,570(**)	,435(**)	
		3. Fracasso no Passado	,177(**)	,234(**)	,202(**)	,252(**)	
		5. Sentimentos de Culpa	,169(**)	,197(**)	,193(**)	,245(**)	
		6. Sentimentos de Punição	,190(**)	,200(**)	,202(**)	,211(**)	
			8. Autocrítica	,207(**)	,294(**)	,204(**)	,225(**)

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Nonparametric Correlations (Inter-Items do Factor 4 e dos Restantes Factores)

		Correlations				
		Factor 4				
		3.	5.	6.	8.	
Spearman's rho	Factor 4	3. Fracasso no Passado	1,000			
		5. Sentimentos de Culpa	,316(**)	1,000		
		6. Sentimentos de Punição	,194(**)	,266(**)	1,000	
		8. Autocrítica	,220(**)	,201(**)	,177(**)	1,000
	Factor 1	1. Tristeza	,293(**)	,188(**)	,295(**)	,224(**)
		4. Perda de Prazer	,221(**)	,153(**)	,245(**)	,181(**)
		9. Pensamentos ou Desejos Suicidas	,227(**)	,249(**)	,236(**)	,235(**)
		12. Perda de Interesse	,224(**)	,145(**)	,231(**)	,231(**)
		21. Perda de Interesse Sexual	,114(**)	,105(**)	,156(**)	,137(**)
	Factor 2	11. Agitação	,107(**)	,107(**)	,157(**)	,159(**)
		16. Alterações nos Hábitos do Sono	,103(**)	,169(**)	,091(*)	,193(**)
		17. Irritabilidade	,153(**)	,117(**)	,194(**)	,267(**)
		18. Alterações no Apetite	,129(**)	,128(**)	,189(**)	,197(**)
		20. Cansaço ou Fadiga	,209(**)	,192(**)	,177(**)	,258(**)
	Factor 3	2. Pessimismo	,177(**)	,169(**)	,190(**)	,207(**)
		13. Indecisão	,234(**)	,197(**)	,200(**)	,294(**)
		15. Perda de Energia	,202(**)	,193(**)	,202(**)	,204(**)
		19. Dificuldades de Concentração	,252(**)	,245(**)	,211(**)	,225(**)

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Anexo O: Consistência Interna dos Factores

Reliability (Factor 1)

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	666	99,4
	Excluded(a)	4	,6
	Total	670	100,0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,764	,766	5

Item Statistics

	Mean	Std. Deviation	N
1. Tristeza	,2027	,45162	666
4. Perda de Prazer	,3168	,54587	666
9. Pensamentos ou Desejos Suicidas	,0811	,32356	666
12. Perda de Interesse	,3979	,63667	666
21. Perda de Interesse Sexual	,2477	,54858	666

Inter-Item Correlation Matrix

	1.	4.	9.	12.	21.
1. Tristeza	1,000				
4. Perda de Prazer	,434	1,000			
9. Pensamentos ou Desejos Suicidas	,330	,357	1,000		
12. Perda de Interesse	,462	,636	,369	1,000	
21. Perda de Interesse Sexual	,301	,400	,251	,411	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1. Tristeza	1,0435	2,445	,513	,273	,729
4. Perda de Prazer	,9294	2,054	,653	,456	,674
9. Pensamentos ou Desejos Suicidas	1,1652	2,806	,428	,188	,759
12. Perda de Interesse	,8483	1,822	,667	,476	,669
21. Perda de Interesse Sexual	,9985	2,308	,459	,213	,749

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
------	----------	----------------	------------

200

1,2462	3,374	1,83681	5
--------	-------	---------	---

Reliability (Factor 2)

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	668	99,7
	Excluded(a)	2	,3
	Total	670	100,0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,722	,721	5

Item Statistics

	Mean	Std. Deviation	N
11. Agitação	,5853	,63640	668
16. Alterações nos Hábitos do Sono	,8503	,72752	668
17. Irritabilidade	,5913	,71337	668
18. Alterações no Apetite	,6123	,74337	668
20. Cansaço ou Fadiga	,7650	,73884	668

Inter-Item Correlation Matrix

	11.	16.	17.	18.	20.
11. Agitação	1,000				
16. Alterações nos Hábitos do Sono	,280	1,000			
17. Irritabilidade	,326	,347	1,000		
18. Alterações no Apetite	,224	,358	,345	1,000	
20. Cansaço ou Fadiga	,319	,453	,446	,309	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
11. Agitação	2,8189	4,547	,393	,161	,707
16. Alterações nos Hábitos do Sono	2,5539	3,996	,514	,279	,662
17. Irritabilidade	2,8129	4,017	,522	,282	,659
18. Alterações no Apetite	2,7919	4,162	,430	,195	,696
20. Cansaço ou Fadiga	2,6392	3,877	,549	,321	,647

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
------	----------	----------------	------------

202

3,4042	6,019	2,45342	5
--------	-------	---------	---

Reliability (Factor 3)

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	668	99,7
	Excluded(a)	2	,3
	Total	670	100,0

a Listwise deletion based on all variables in the procedure.

Item Statistics

	Mean	Std. Deviation	N
2. Pessimismo	,3653	,55151	668
13. Indecisão	,4805	,63947	668
15. Perda de Energia	,6168	,67721	668
19. Dificuldades de Concentração	,7051	,75569	668

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,681	,682	4

Inter-Item Correlation Matrix

	2.	13.	15.	19.
2. Pessimismo	1,000			
13. Indecisão	,335	1,000		
15. Perda de Energia	,291	,381	1,000	
19. Dificuldades de Concentração	,306	,340	,444	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
2. Pessimismo	1,8024	2,552	,402	,166	,654
13. Indecisão	1,6871	2,260	,465	,221	,615
15. Perda de Energia	1,5509	2,104	,510	,268	,584
19. Dificuldades de Concentração	1,4626	1,955	,491	,251	,601

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
------	----------	----------------	------------

2,1677	3,564	1,88787	4
--------	-------	---------	---

Reliability (Factor 4)

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	667	99,6
	Excluded(a)	3	,4
	Total	670	100,0

a Listwise deletion based on all variables in the procedure.

Item Statistics

	Mean	Std. Deviation	N
3. Fracasso no Passado	,3373	,60665	667
5. Sentimentos de Culpa	,4378	,54823	667
6. Sentimentos de Punição	,2309	,65122	667
8. Autocrítica	,6162	,65730	667

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,612	,618	4

Inter-Item Correlation Matrix

	3.	5.	6.	8.
3. Fracasso no Passado	1,000			
5. Sentimentos de Culpa	,363	1,000		
6. Sentimentos de Punição	,259	,322	1,000	
8. Autocrítica	,276	,267	,239	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
3. Fracasso no Passado	1,2849	1,784	,411	,181	,527
5. Sentimentos de Culpa	1,1844	1,853	,445	,205	,509
6. Sentimentos de Punição	1,3913	1,755	,370	,144	,559
8. Autocrítica	1,0060	1,769	,353	,126	,573

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
------	----------	----------------	------------

206

1,622	2,818	1,67869	4
-------	-------	---------	---

Anexo P: Análise das Validades Convergente e Discriminante

Correlations (BDI-II/IACLIDE)

Descriptive Statistics

	Mean	Std. Deviation	N
BDITotal	9,4701	7,60054	670
IACLIDETotal	11,9821	11,44895	559

Correlations

		BDITotal	IACLIDETotal
BDITotal	Pearson Correlation	1	,913(**)
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	38646,903	45549,256
	Covariance	57,768	81,629
	N	670	559
IACLIDETotal	Pearson Correlation	,913(**)	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	45549,256	73141,821
	Covariance	81,629	131,079
	N	559	559

** Correlation is significant at the 0.01 level (2-tailed).

Correlations (BDI-II/STAI)

Descriptive Statistics

	Mean	Std. Deviation	N
BDITotal	9,4701	7,60054	670
STAITotal	38,2746	10,86261	670

Correlations

		BDITotal	STAITotal
BDITotal	Pearson Correlation	1	,728(**)
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	38646,903	40205,493
	Covariance	57,768	60,098
	N	670	670
STAITotal	Pearson Correlation	,728(**)	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	40205,493	78939,469
	Covariance	60,098	117,996
	N	670	670

** Correlation is significant at the 0.01 level (2-tailed).

Correlations (STAI/IACLIDE)

Descriptive Statistics

	Mean	Std. Deviation	N
STAITotal	38,2746	10,86261	670
IACLIDETotal	11,9821	11,44895	559

Correlations

		BDITotal	IACLIDETotal
STAITotal	Pearson Correlation	1	,777(**)
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	78939,469	53856,923
	Covariance	117,996	96,518
	N	670	559
IACLIDETotal	Pearson Correlation	,777(**)	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	53856,923	73141,821
	Covariance	96,518	131,079
	N	559	559

** Correlation is significant at the 0.01 level (2-tailed).

Anexo Q: Datos Sócio-Demográficos

T-Test (Será que Existem Diferenças de Género?!?)

Group Statistics

	Género	N	Mean	Std. Deviation	Std. Error Mean
BDI Score Total	Feminino	572	9,5035	7,20855	,30140
	Masculino	98	9,2755	9,61719	,97148

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BDI Score Total	Equal variances assumed	3,867	,050	,274	668	,784²¹	,22799	,83152	-1,40471	1,86069
	Equal variances not assumed			,224	116,389	,823	,22799	1,01716	-1,78657	2,24254

²¹ Sendo $p\text{-value}=0,784 > 0,05 = \alpha$, não se rejeita a hipótese de que as médias dos resultados do BDI-II dos estudantes do género feminino não diferem das médias dos estudantes do género masculino, ou seja, não existem diferenças significativas no score total do BDI-II quanto ao género.

Oneway (Será que Existem Diferenças entre Cursos?!?)

ANOVA

BDI Score Total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	105,551	2	52,776	,913	,402²²
Within Groups	38541,352	667	57,783		
Total	38646,903	669			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: BDI Score Total

Tukey HSD

(I) Curso	(J) Curso	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
RIS	DCSM	-2,32131	1,98032	,470	-6,9730	2,3304
	PSIC	-,16204	1,22666	,990	-3,0434	2,7193
DCSM	RIS	2,32131	1,98032	,470	-2,3304	6,9730
	PSIC	2,15928	1,61483	,375	-1,6339	5,9524
PSIC	RIS	,16204	1,22666	,990	-2,7193	3,0434
	DCSM	-2,15928	1,61483	,375	-5,9524	1,6339

Homogeneous Subsets

BDIScore Total

Tukey HSD ^{a, b}

Curso	N	Subset for alpha = .05
		1
RIS	41	9,2439
PSIC	606	9,4059
DCSM	23	11,5652
Sig.		,332

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 43,154.

²² Sendo p-value=0,402>0,05=α, não se rejeita a hipótese de que as médias dos resultados do BDI-II dos estudantes de Psicologia, RIS e DCSM não diferem umas das outras, ou seja, não existem diferenças significativas no score total do BDI-II quanto ao curso.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Oneway (Será que Existem Diferenças entre Anos?!?)

ANOVA

BDI Score Total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	204,839	3	68,280	1,183	,315²³
Within Groups	38442,064	666	57,721		
Total	38646,903	669			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: BDI Score Total

Tukey HSD

(I) Ano	(J) Ano	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	,89257	,89741	,753	-1,4187	3,2039
	3,00	1,61900	,86624	,242	-,6120	3,8500
	4,00	1,08547	,88762	,612	-1,2006	3,3716
2,00	1,00	-,89257	,89741	,753	-3,2039	1,4187
	3,00	,72643	,79510	,798	-1,3214	2,7742
	4,00	-,19290	,81834	,995	-1,9148	2,3006
3,00	1,00	-1,61900	,86624	,242	-3,8500	,6120
	2,00	-,72643	,79510	,798	-2,7742	1,3214
	4,00	-,53353	,78403	,905	-2,5528	1,4858
4,00	1,00	-1,08547	,88762	,612	-3,3716	1,2006
	2,00	-,19290	,81834	,995	-2,3006	1,9148
	3,00	,53353	,78403	,905	-1,4858	2,5528

Homogeneous Subsets

BDIScore Total

Tukey HSD ^{a, b}

Ano	N	Subset for alpha = .05
		1
3,00	200	8,8450
4,00	177	9,3785
2,00	168	9,5714
1,00	125	10,4640
Sig.		,220

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 162,588.

²³ Sendo $p\text{-value}=0,315 > 0,05 = \alpha$, não se rejeita a hipótese de que as médias dos resultados do BDI-II dos estudantes do 1º, 2º, 3º, e 4º anos não diferem umas das outras, ou seja, não existem diferenças significativas no score total do BDI-II quanto ao ano frequentado.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Frequencies

Statistics

Grupos Etários

N	Valid	670
	Missing	0

Grupos Etários

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 18-20	205	30,6	30,6	30,6
21-23	341	50,9	50,9	81,5
24-28	83	12,4	12,4	93,9
29-33	22	3,3	3,3	97,2
34-38	4	,6	,6	97,8
39-43	6	,9	,9	98,7
44-48	9	1,3	1,3	100,0
Total	670	100,0	100,0	

Oneway (Será que Existem Diferenças entre Grupos Etários?!?)

ANOVA

BDI Score Total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	278,282	6	46,380	,801	,569 ²⁴
Within Groups	38368,621	663	57,871		
Total	38646,903	669			

²⁴ Sendo $p\text{-value}=0,569 > 0,05 = \alpha$, não se rejeita a hipótese de que as médias dos resultados do BDI-II dos estudantes dos vários grupos etários não diferem umas das outras, ou seja, não existem diferenças significativas no score total do BDI-II quanto aos grupos etários.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: BDI Score Total
Tukey HSD

(I) Grupos Etários	(J) Grupos Etários	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
18-20	21-23	-,49504	,67232	,990	-2,4834	1,4933
	24-28	,11825	,98972	1,000	-2,8088	3,0453
	29-33	-2,10355	1,70670	,881	-7,1510	2,9439
	34-38	2,46463	3,84059	,995	-8,8936	13,8229
	39-43	4,38130	3,15079	,807	-4,9369	13,6995
	44-48	-,22981	2,59084	1,000	-7,8920	7,4324
21-23	18-20	,49504	,67232	,990	-1,4933	2,4834
	24-28	,61329	,93110	,995	-2,1404	3,3670
	29-33	-1,60850	1,67339	,962	-6,5574	3,3404
	34-38	2,95968	3,82590	,987	-8,3551	14,2745
	39-43	4,87634	3,13288	,710	-4,3889	14,1416
	44-48	,26523	2,56902	1,000	-7,3324	7,8629
24-28	18-20	-,11825	,98972	1,000	-3,0453	2,8088
	21-23	-,61329	,93110	,995	-3,3670	2,1404
	29-33	-2,22180	1,82421	,887	-7,6168	3,1732
	34-38	2,34639	3,89423	,997	-9,1705	13,8633
	39-43	4,26305	3,21597	,840	-5,2479	13,7740
	44-48	-,34806	2,66972	1,000	-8,2435	7,5474
29-33	18-20	2,10355	1,70670	,881	-2,9439	7,1510
	21-23	1,60850	1,67339	,962	-3,3404	6,5574
	24-28	2,22180	1,82421	,887	-3,1732	7,6168
	34-38	4,56818	4,13501	,927	-7,6608	16,7971
	39-43	6,48485	3,50367	,514	-3,8770	16,8467
	44-48	1,87374	3,01009	,996	-7,0284	10,7758
34-38	18-20	-2,46463	3,84059	,995	-13,8229	8,8936
	21-23	-2,95968	3,82590	,987	-14,2745	8,3551
	24-28	-2,34639	3,89423	,997	-13,8633	9,1705
	29-33	-4,56818	4,13501	,927	-16,7971	7,6608
	39-43	1,91667	4,91050	1,000	-12,6057	16,4391
	44-48	-2,69444	4,57143	,997	-16,2141	10,8252
39-43	18-20	-4,38130	3,15079	,807	-13,6995	4,9369
	21-23	-4,87634	3,13288	,710	-14,1416	4,3889
	24-28	-4,26305	3,21597	,840	-13,7740	5,2479
	29-33	-6,48485	3,50367	,514	-16,8467	3,8770
	34-38	-1,91667	4,91050	1,000	-16,4391	12,6057
	44-48	-4,61111	4,00941	,912	-16,4686	7,2464
44-48	18-20	,22981	2,59084	1,000	-7,4324	7,8920
	21-23	-,26523	2,56902	1,000	-7,8629	7,3324
	24-28	,34806	2,66972	1,000	-7,5474	8,2435

29-33	-1,87374	3,01009	,996	-10,7758	7,0284
34-38	2,69444	4,57143	,997	-10,8252	16,2141
39-43	4,61111	4,00941	,912	-7,2464	16,4686

Homogeneous Subsets

BDIScore Total

Tukey HSD ^{a, b}

Grupos Etários	N	Subset for alpha = .05
		1
39-43	6	4,8333
34-38	4	6,7500
24-28	83	9,0964
18-20	205	9,2146
44-48	9	9,4444
21-23	341	9,7097
29-33	22	11,3182
Sig.		,371

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 11,803.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type error levels are not guaranteed.

Anexo R: Análise da Sensibilidade dos Itens do IACLIDE

Summarize

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
IACLIDE1	559	83,4%	111	16,6%	670	100,0%
IACLIDE2	559	83,4%	111	16,6%	670	100,0%
IACLIDE3	559	83,4%	111	16,6%	670	100,0%
IACLIDE4	557	83,1%	113	16,9%	670	100,0%
IACLIDE5	558	83,3%	112	16,7%	670	100,0%
IACLIDE6	558	83,3%	112	16,7%	670	100,0%
IACLIDE7	558	83,3%	112	16,7%	670	100,0%
IACLIDE8	558	83,3%	112	16,7%	670	100,0%
IACLIDE9	558	83,3%	112	16,7%	670	100,0%
IACLIDE10	558	83,3%	112	16,7%	670	100,0%
IACLIDE11	558	83,3%	112	16,7%	670	100,0%
IACLIDE12	558	83,3%	112	16,7%	670	100,0%
IACLIDE13	557	83,1%	113	16,9%	670	100,0%
IACLIDE14	556	83,0%	114	17,0%	670	100,0%
IACLIDE15	556	83,0%	114	17,0%	670	100,0%
IACLIDE16	557	83,1%	113	16,9%	670	100,0%
IACLIDE17	557	83,1%	113	16,9%	670	100,0%
IACLIDE18	557	83,1%	113	16,9%	670	100,0%
IACLIDE19	556	83,0%	114	17,0%	670	100,0%
IACLIDE20	556	83,0%	114	17,0%	670	100,0%
IACLIDE21	557	83,1%	113	16,9%	670	100,0%
IACLIDE22	557	83,1%	113	16,9%	670	100,0%
IACLIDE23	556	83,0%	114	17,0%	670	100,0%
IACLIDETotal	559	83,4%	111	16,6%	670	100,0%

Case Summaries

	Skewness	Kurtosis
IACLIDE1	1,374	1,496
IACLIDE2	2,717	9,996
IACLIDE3	1,113	1,769
IACLIDE4	1,994	5,116
IACLIDE5	1,257	,514
IACLIDE6	1,250	1,480
IACLIDE7	1,311	1,502
IACLIDE8	1,582	2,296
IACLIDE9	3,049	10,904
IACLIDE10	1,220	1,688
IACLIDE11	2,815	9,036
IACLIDE12	2,230	7,171
IACLIDE13	1,437	3,895
IACLIDE14	2,348	6,904
IACLIDE15	1,207	2,145
IACLIDE16	1,696	3,028
IACLIDE17	1,988	5,053
IACLIDE18	1,822	4,124
IACLIDE19	5,242	29,418
IACLIDE20	1,459	2,955
IACLIDE21	2,354	7,519
IACLIDE22	1,670	4,733
IACLIDE23	1,898	4,983
IACLIDETotal	2,341	9,045

