

5

Discussão, conclusões e trabalhos futuros

5.1.

Discussão

Considerando a relevância da confiabilidade dos resultados de quantificação de constructos obtidos por meio de testes psicométricos, o presente trabalho realiza amplo estudo da confiabilidade metrológica dos inventários de Hamilton e Beck para medir a intensidade dos transtornos de humor e ansiedade. Outro aspecto abordado no trabalho consiste em estudos para a construção de uma proposta de harmonização entre a terminologia utilizada em psicometria e as definições universais para os termos referentes às medições, com base no Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM). Esta proposta, que é apresentada e discutida no item 3.3 desta dissertação, tem por intuito possibilitar melhor comunicação e compreensão interdisciplinar dos resultados e abordagens para avaliação das escalas psicométricas para medição de transtornos psíquicos.

A tabela 5.1 apresenta, de forma resumida, os principais resultados comparativos que relacionam os inventários de Hamilton e Beck realizados com base nos resultados finais de escore total dos entrevistados. Com referência a avaliação dos itens de cada inventário, nesta tabela estão incluídos somente os resultados da abordagem restrita da análise de seleção de parâmetros do item (ASPI qualitativa restrita), sendo as abordagens ASPI ampla e quantitativa discutidas mais adiante. A comparação tem por base as escalas de Beck por estas já estarem validadas na população brasileira. Os estudos apresentados na tabela 5.1 incluem as seguintes abordagens analíticas:

- Validade entre os inventários Hamilton e Beck
- Índices de erro e incerteza entre escores obtidos para os inventários de Hamilton e Beck.
- Convergência de quantificação de constructo entre os inventários de Hamilton e Beck

- ASPI qualitativa restrita dos itens dos inventários de Hamilton e Beck

Tabela 23: Resumo dos resultados comparativos entre os inventários de Hamilton (HAM-A e HAM-D) e Beck (BAI e BDI).

	Amostra de Pacientes	Validade	Convergência de quantificação do construto	Abordagens Analíticas da Quantificação de Intensidade do Construto	mínimo	leve	moderado	grave
HAM-D x BDI	Todos os pacientes (100)	r = 0,81	43%	Percentual de Convergência da quantificação do construto entre os inventários	13%	48%	34%	89%
				Incerteza entre os inventários (valor médio = 0,17)	0,53	0,08	0,04	0,04
				ASPI qualitativa restrita dos itens do Ham-D	57%	76%	9%	52%
				ASPI qualitativa restrita dos itens do BDI	95%	52%	9%	48%
	Pacientes com diagnóstico de Transtorno de Humor (53) e comórbidos (3)	-	36%	Percentual de Convergência da quantificação do construto entre os inventários	0%	33%	21%	92%
				Incerteza entre os inventários (valor médio = 0,18)	0,53	0,12	0,05	0,04
				ASPI qualitativa restrita dos itens do Ham-D	0%	52%	14%	66%
				ASPI qualitativa restrita dos itens do BDI	52%	62%	14%	57%
HAM-D x BAI	Todos os pacientes (100)	r = 0,62	-	-	-	-	-	-
HAM-A x BAI	Todos os pacientes (100)	r = 0,74	45%	Percentual de Convergência da quantificação do construto entre os inventários	64%	26%	19%	83%
				Incerteza entre os inventários (valor médio = 0,16)	0,41	0,13	0,07	0,04
				ASPI qualitativa restrita dos itens do Ham-A	57%	14%	7%	71%
				ASPI qualitativa restrita dos itens do BAI	90%	9%	5%	33%
	Pacientes com diagnóstico de Transtorno de Ansiedade (44) e comórbidos (3)	-	42%	Percentual de Convergência da quantificação do construto entre os inventários	78%	20%	21%	78%
				Incerteza entre os inventários (valor médio = 0,28)	0,77	0,17	0,12	0,07
				ASPI qualitativa restrita dos itens do Ham-A	57%	14%	14%	50%
				ASPI qualitativa restrita dos itens do BAI	95%	52%	5%	9%
HAM-A x BDI	Todos os pacientes (100)	r = 0,64	-	-	-	-	-	-

O estudo dos inventários para quantificação de transtorno de humor (HAM-D e BDI) com as escalas aplicadas aos 100 pacientes entrevistados indica uma boa validade convergente ($r=0.81$), com percentual total de convergência de quantificação de constructo de 43% e incerteza média (entre diferentes níveis de quantificação) de 0,17. A boa correlação observada na validade convergente deve-se provavelmente à boa correspondência entre estes inventários para identificar pacientes graves (89%), com reduzida incerteza (0,04). No entanto, o grau de contribuição dos itens do HAM-D (52%) e do BDI (48%) para a identificação de pacientes graves não foi tão elevado, indicando que os dois inventários quantificam de modo equivalente os pacientes graves, embora não sejam individualmente excelentes para quantificar este grau de acometimento.

Para o estudo dos inventários que quantificam transtorno de humor (HAM-D e BDI) com as escalas aplicadas aos 56 pacientes com diagnóstico de transtorno de humor (incluindo comórbidos) não foi feita análise de validade. O percentual total de convergência de quantificação de constructo (36%) foi levemente inferior ao observado com o estudo para todos os pacientes e a incerteza média (entre diferentes níveis de quantificação) foi levemente superior (0,18), o que poderia ser explicado pela convergência nula na quantificação de pacientes com grau mínimo. O valor de convergência elevado observado para quantificação do nível grave (92%) está um pouco acima do observado quando considerando a amostra de todos os 100 pacientes (89%).

O estudo dos inventários para quantificação de transtorno de ansiedade (HAM-A e BAI), com as escalas aplicadas aos 100 pacientes entrevistados, indica uma validade convergente ($r=0.74$), com percentual total de convergência de quantificação de constructo de 45% e incerteza média (entre diferentes níveis de quantificação) de 0,16. A correlação observada na validade convergente deve-se provavelmente à boa correspondência entre estes inventários para identificar pacientes graves (83%), com reduzida incerteza (0,04), e pacientes com grau mínimo de acometimento (64%), embora neste caso apresente incerteza mais elevada (0,41).

Para o estudo dos inventários que quantificam transtorno de ansiedade (HAM-A e BAI) com as escalas aplicadas aos 47 pacientes com diagnóstico de transtorno de ansiedade (incluindo comórbidos) não foi feita análise de validade.

O percentual total de convergência de quantificação de constructo (42%) foi levemente inferior ao observado com o estudo para todos os pacientes e a incerteza média (entre diferentes níveis de quantificação) foi muito superior (0,28). Assim como observado para a avaliação da convergência entre os dois inventários de ansiedade aplicados aos 100 pacientes, os resultados considerando somente os pacientes com transtorno de ansiedade indicaram convergência maior para quantificação do nível grave (78%) e nível mínimo (78%).

Em todas as avaliações, os resultados de incerteza foram inferiores para a quantificação do nível grave e maiores para a quantificação do nível mínimo. Somente os inventários de Beck apresentaram elevados percentuais de contribuição dos itens para a quantificação do nível mínimo de acometimento. Para os demais graus de intensidade de constructo (leve, moderado e grave) nenhum dos inventários apresentou desempenho excelente para a quantificação do transtorno.

A avaliação dos inventários considerando somente os pacientes com o diagnóstico específico para o qual o inventário está direcionado, resultou em piores valores para o percentual total de convergência da quantificação do constructo e para os valores médios de incerteza.

Apesar dos inventários HAM-D e BAI, assim como HAM-A e BDI, avaliarem constructos diferentes, não se observou boa validade discriminante entre eles ($r=0,62$ e $r=0,64$, respectivamente). Estes resultados podem se dever à superposição de sintomas presentes tanto na ansiedade quanto na depressão (Mineka et al 1998). Alguns autores ainda indicam que a ansiedade e a depressão possam apresentar uma única dimensão básica, formando uma classe geral de distúrbios de humor (Feldman, 1993).

Na avaliação clássica da *fidedignidade* realizada para o inventário HAM-A (Tabela 4.6), o único item que se destaca como apresentando um baixo coeficiente de correlação item-total é o item 14 (0,06), referente ao “comportamento na entrevista”. Este item que, pela análise de *fidedignidade* conclui-se que contribui para a redução da consistência interna do questionário, em uma análise detalhada apresentada na avaliação qualitativa dos itens (Tabelas 4.18 e apêndice 2), observa-se que a ASPI qualitativa ampla o classifica como incapaz de discernir

entre os diferentes graus de acometimento do constructo e a ASPI quantitativa apresenta todos coeficientes insuficientes.

Provavelmente este baixo desempenho do item 14 para a quantificação do constructo ansiedade se deva à grande dificuldade de aplicação do mesmo, que inclui identificação de sintomas fisiológicos, por profissionais de psicologia (profissionais habilitados para o uso do inventário).

Por meio da análise fatorial do HAM-A, segundo rotação varimax, foram extraídos dois fatores que foram denominados de “sintomas somáticos”, composto por nove itens que são: 10-Sintomas respiratórios; 2-Tensão; 13-Sistema Nervoso Autônomo (SNA); 9-sistema cardiovascular; 1-humor ansioso; 11-sintomas gastrointestinais; 8-sintoma somático (sensorial); 3-medo e 7-sintoma somático (muscular) e “sintomas psíquicos” composto por cinco itens, dos quais fazem parte os itens 6-humor depressivo; 4-insônia; 12-sintomas genitourinários; 5-dificuldades intelectuais e 14-comportamento na entrevista.

Em um estudo realizado em 1989 nos Estados Unidos, para avaliar os sintomas de ansiedade generalizada e os efeitos da buspirona, assim como observado na presente dissertação, foram encontrados dois fatores compondo os itens do HAM-A (Feighner e Cohn, 1989). Em outro estudo realizado nos Estados Unidos em 2008, objetivando avaliar a tolerância e a eficácia da duloxetine em pacientes idosos com transtorno de ansiedade generalizada, também foram extraídos dois fatores do HAM-A (Davidson J e col, 2008). Um terceiro estudo realizado na Suíça em 2004 para avaliar a eficácia da Sertralina em pacientes com transtorno de ansiedade generalizada, também extraiu o mesmo numero de fatores (Allgulander C, 2004). Em todos os três estudos mencionados, os dois fatores encontrados foram denominados “sintomas somáticos” e “sintomas psíquicos”.

A unidimensionalidade dos itens de uma escala aditiva (itens correlacionados uns com os outros) é uma característica desejável em um inventário, pois indica que este representa um único conceito (Cunha, 2008). No entanto, o fato de terem sido extraídos dois fatores para o HAM-A (tanto no presente trabalho quanto nos outros trabalhos referidos no parágrafo anterior), ainda poderia indicar estar-se medindo um único constructo (transtorno de ansiedade), cujas distintas características (somáticas e psíquicas) presentes na definição do atributo são medidas por meio de dois conjuntos de itens.

Recentes estudos, porém, argumentam contra o uso do coeficiente Alfa de Cronbach, considerando-o como incapaz de avaliar a consistência interna (Sijtsma, 2009; Green e Jang, 2009; Bentler, 2009).

5.2. Análise de Seleção de Parâmetros dos Itens

A Análise de Seleção de Parâmetros dos Itens (ASPI), proposta no presente trabalho, detalha a avaliação do inventário no nível de seu componente fundamental, que é o item, objetivando avaliar a capacidade de cada item em discernir entre os diferentes níveis de acometimento do constructo. A ASPI avalia o item por meio de abordagens qualitativas e quantitativas, sendo que a associação dos resultados permite detectar o quanto os índices quantitativos propostos podem ser efetivamente preditivos do desempenho do item.

Pela ASPI qualitativa ampla foram detectados 5 possíveis diagnósticos de desempenho do item (quadro 4.2), sendo um deles a incapacidade de discernimento (diagnóstico E). Nenhum dos itens dos inventários foi capaz de realizar excelente grau de discernimento entre diferentes níveis de classificação do constructo. Conforme observável na tabela 4.19, em sua maioria, o tipo de discernimento realizado pelos itens foi entre o nível grave em relação aos demais ou em relação ao mínimo (não sendo capazes de identificar os graus intermediários), que correspondem aos diagnósticos A e B da ASPI qualitativa. Este resultado, que indica melhor capacidade de identificar o grau grave do que outros níveis de acometimento, observado em todos os quatro inventários, pode explicar a obtenção de menores valores de incertezas entre os inventários obtidos para a identificação do grau grave, assim como a maior convergência de classificação de constructo, que também ocorreu para este grau de acometimento. O discernimento que se poderia considerar mais próximo do desejável correspondente ao diagnóstico D, que indica que o item seria capaz de discernir “um pouco” entre todos graus de acometimento. Este diagnóstico só foi atendido por um item do inventário BAI e dois itens do BDI. A tabela 5.2 resume o diagnóstico final da ASPI qualitativa ampla quanto à capacidade dos itens de discernimento ou não entre os níveis de constructo. Os resultados indicam um melhor desempenho dos itens dos inventários de Beck em relação aos inventários

de Hamilton, mas em todos os casos, a maioria dos itens dos inventários não atende a critérios de capacidade de discernimento entre os níveis do constructo.

Tabela 24: Diagnóstico da ASPI qualitativa ampla quanto à capacidade dos itens de realizar algum discernimento entre os níveis de constructo ou não, apresentada individualmente para cada um dos quatro inventários e para todos os itens avaliados (de todos inventários, sem discrimina-los).

ASPI qualitativa ampla	BDI	BAI	HAM-A	HAM-D	Todos Inventários
Itens capazes de discernimento	(10) 48%	(9) 43%	(4) 29%	(5) 24%	(28) 36%
Itens incapazes de discernimento	(11) 52%	(12) 57%	(10) 71%	(16) 76%	(49) 63%

Por meio da avaliação dos resultados da ASPI quantitativa apresentados em associação aos resultados da ASPI qualitativa (tabela 4.21), uma análise de correspondência pode ser efetuada e as aparentes incoerências discutidas.

O inventário HAM-D, que classificou apenas 3 pacientes como mínimos, possui 9 de seus 21 itens com apenas 3 parâmetros. Tais questões com apenas 3 opções apresentaram elevado índice de coeficientes inadequados, sendo que 100% destas apresentaram coeficientes angulares insuficientes. Do ponto de vista da ASPI qualitativa, 4 destes itens do HAM-D apresentaram algum grau de discernimento, tendo sido estes os únicos com coeficientes lineares dentro dos limites aceitáveis. O item 19 do inventário BDI, embora possuísse número maior de parâmetros, o último não foi selecionado pelos respondentes, tendo sido classificado pela ASPI qualitativa como incapaz de discernimento, apesar de um coeficiente angular bom. Os resultados de baixo desempenho ou duvidosos apresentados nos itens com apenas 3 parâmetros podem indicar que para classificar os entrevistados em 4 níveis de acometimento é necessário pelo menos 4 parâmetros para cada item. O item 19 do BDI que, apesar de 4 parâmetros selecionáveis, não obteve bom resultado devido ao último parâmetro não ter sido selecionado por ninguém, indica que mais um parâmetro com conteúdo intermediário de gravidade deve ser incluído anteriormente ao parâmetro 2.

As maiores divergências de quantificação de constructo foram observadas para a quantificação do grau mínimo entre HAM-D e BDI. O mesmo não foi

observado, porém, ao avaliarem-se os questionários de avaliação do transtorno de Ansiedade, os quais convergiram a quantificação para o nível grave de acometimento e para o grau mínimo. A divergência de quantificação para o nível mínimo observada entre inventários para transtorno de humor deveu-se, provavelmente, ao fato de o HAM-D ter quantificado apenas 3 respondentes como mínimos.

A tabela 5.3 apresenta um resumo da análise dos índices quantitativos considerados mais importantes e o percentual destes índices que estão associados com o diagnóstico ASPI qualitativo de discernimento ou não. Espera-se que os índices com valores insuficientes preponderem na classificação qualitativa de item como incapaz de discernimento, e os índices quantitativos bem qualificados segundo o quadro 4.2 (indicadores de bom desempenho), estejam presentes em itens que tenham se classificado qualitativamente como capazes de discernir o grau do constructo. A análise dos resultados apresentada na tabela 5.3 indica que o coeficiente linear (CL), obtido pela ASPI quantitativa, é fator decisivo para a identificação do desempenho do item. Todos os itens com $CL \leq 0,20$ (indicador de bom desempenho) apresentaram capacidade de discernimento. Embora poucas discrepâncias entre ASPI quantitativa e qualitativa tenham sido encontradas, verifica-se que todas se caracterizaram por itens com coeficientes angulares ou diferença de coeficientes incoerentes com a classificação qualitativa, associadas a coeficientes lineares coerentes com a classificação. Outro índice da ASPI quantitativa que foi totalmente coerente com a classificação qualitativa foi a diferença entre os coeficientes com valores acima de 0,25, pois todos os itens com esse desempenho se classificaram como capazes de algum discernimento entre os níveis de constructo.

Tabela 25: Associação entre itens (de quaisquer dos 4 inventários estudados) que apresentaram determinadas faixas de valores dos índices da ASPI quantitativa e o diagnóstico da ASPI qualitativa ampla quanto à capacidade dos itens de realizar algum discernimento entre os níveis de constructo ou não. Entre parênteses se apresenta o número de itens correspondentes a cada associação. Nos resultados que demonstraram discordância entre ASPI qtt e qit, é inserida também a informação do valor de CL (sempre coerente com a classificação qualitativa).

ASPI quantitativa	ASPI qualitativa	
	Incapaz de discernimento	Discernimento

CL \leq 0,20 (19)	0%	(19) 100%
CL \geq 0,27 (29)	(29) 100%	0%
D \leq 0,13 (46)	(43) 93%	(3) 7% (CL \leq 0,20)
D \geq 0,25 (46)	0%	(6) 100%
CA \geq 0,42 (12)	(4) 33% (CL \geq 0,27)	(8) 67%
CA \leq 0,30 (22)	(20) 91%	(2) 9% (CL \leq 0,20)

Recentemente, os testes de avaliação psicológica vêm também sendo avaliados por meio da Teoria de Respostas ao Item (TRI), abordagem que, assim como a proposta no presente trabalho, avaliam a capacidade do item de discriminar a gravidade do transtorno. Curí (2006), utilizando TRI para avaliação do inventário BDI, indicou seus itens 3, 5 e 7 como os mais discriminantes para quantificação do transtorno. Estes resultados não estão de acordo com os obtidos no presente trabalho, sendo que tanto o item 5 quanto o item 7 do BDI foram classificados como incapazes de discriminar entre diferentes graus de acometimento do transtorno pela ASPI qualitativa ampla, e segundo ASPI quantitativa, ambos apresentaram índice insuficiente de CL, mas o item 7 obteve bom resultado para o coeficiente angular. É necessário, porém, aprofundar a compreensão das diferenças de informação obtidas pelas abordagens, para realizar-se melhor análise comparativa entre ambas.

5.3. Harmonização do Vocabulário

Segundo o VIM 2008, a validade dos resultados de medição é altamente dependente das propriedades metrológicas do instrumento conforme demonstrado pela sua calibração. Neste sentido, do ponto de vista metrológico, para que haja validade nos estudos dos questionários em psicometria, seria necessário definir um padrão clínico. O intervalo dos valores atribuídos ao mensurando é o intervalo dos valores dos padrões que teriam fornecido as mesmas indicações (VIM 2008). Outras abordagens metrológicas evitam completamente o conceito de valor

verdadeiro e avaliam a validade dos resultados de medição com auxílio do conceito de compatibilidade metrológica (VIM 2008). Esta última consiste na propriedade de um conjunto de resultados de medição correspondentes a um mensurando especificado, tal que o valor absoluto da diferença dos valores medidos de todos os pares de resultados de medição seja menor que um certo múltiplo escolhido da incerteza-padrão desta diferença, representando o critério de decisão se dois resultados de medição referem-se a um mesmo mensurando ou não.

Sugere-se que a determinação de um padrão clínico por meio de um rigoroso estudo e identificação de pacientes agrupados conforme transtorno psíquico e grau de acometimento do mesmo, considerando o parecer de uma equipe de pesquisadores formada por especialistas. A partir da elaboração desta referência, poder-se-ia aplicar os inventários objetivando a avaliação da confiabilidade metrológica dos mesmos por meio de calibração e conseqüente identificação de sua exatidão (denominada por validade pela psicometria), além da reprodutibilidade (denominada fidedignidade pela psicometria).

5.4. Conclusões

Os estudos realizados no presente trabalho objetivando a harmonização da terminologia psicométrica e as definições universais referentes às medições pelo VIM identificou os seguintes aspectos:

- Uma vez que o termo “fidedignidade” possui um significado amplo de confiabilidade, o mesmo não seria apropriado para denominar um dos tipos de avaliação da confiabilidade em particular. Neste sentido, considerando as características estudadas por este procedimento, indicar-se-ia denominá-lo como estudo de “reprodutibilidade” dos inventários.

- Para avaliação da confiabilidade de um instrumento de medição em psicologia e considerando a terminologia apresentada pelo VIM, pode-se indicar que são necessárias características como reprodutibilidade, exatidão e homogeneidade dos itens de um inventário, sendo importante a calibração rastreada a um padrão, implicando assim, na análise da incerteza de medição”.

Na presente dissertação foi avaliada a confiabilidade metrológica dos inventários de Hamilton, com base nas escalas de Beck, tendo-se proposto abordagem ASPI, voltada à avaliação dos itens dos inventários. As principais conclusões obtidas na avaliação dos inventários são apresentadas a seguir.

- Os dois inventários (Hamilton e Beck) quantificam de modo equivalente os pacientes graves, embora não sejam individualmente excelentes para quantificar este grau de acometimento. Estes resultados são corroborados pelo estudo das incertezas de medição que apresentaram seus menores valores para a quantificação dos níveis graves de acometimento do constructo, e pelos maiores percentuais de discernimento do nível grave observados na ASPI qualitativa restrita e ampla.

- Apesar da validade convergente observada entre HAM-D e BDI e HAM-A e BAI, os mesmos quantificam de modo equivalente somente os pacientes graves para transtorno de humor, e os pacientes graves e com acometimento mínimo para o transtorno de ansiedade.

- Os inventários de Beck (BAI e BDI) apresentaram o maior percentual de itens capazes de discernimento, comparado com os de Hamilton. Porém, embora com baixo percentual de itens capazes de discernir (29% para HAM-A e 24% para HAM-D), nestes inventários foram encontrados três itens com valores quantitativos excelentes (dois no HAM-A e um no HAM-D), sendo que o único outro item com tal desempenho foi observado no BDI.

- Os resultados observados nos itens do HAM-D com apenas 3 parâmetros indicam a importância da correspondência entre o número de opções para seleção e o número de níveis de acometimento do constructo a ser avaliado.

- A boa correspondência obtida entre as abordagens ASPI quantitativa e a ASPI qualitativa indica a capacidade preditiva para a identificação de desempenho do item pelos indicadores quantitativos propostos.

Com o presente trabalho espera-se contribuir para a disseminação da cultura metrológica ao ambiente das ciências humanas, especificamente no campo da psicologia.

5.5. Trabalhos Futuros

- Elaboração de um padrão clínico para cada nível de acometimento de determinado transtorno, com base em uma rigorosa avaliação clínica de qualificada equipe de especialistas, objetivando a avaliação dos inventários mediante a medição de um grupo de referência.

- Estender o estudo de critérios objetivos para o desenvolvimento de análise automática dos itens dos inventários.