

3

Metrologia, Psicometria e Harmonização da terminologia

3.1.

Metrologia

A realização de uma medição é simples, porém, a segurança de um real significado de seu resultado exige rigor metrológico. Metrologia, do grego metron: medida e logos: ciência é definida, segundo o Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia – (VIM), como a ciência da medição que engloba todos os aspectos teóricos e práticos da medição, qualquer que seja a incerteza e o campo de aplicação (VIM, 2007). É fundamental, por exemplo, que os resultados referentes às medições realizadas em diferentes laboratórios sejam comparáveis e reprodutíveis. A utilização de princípios, métodos e linguagem da Metrologia é primordial para a garantia da uniformidade das medições em todo o planeta. A inserção da ciência das medições (Metrologia) nos diferentes campos resulta em ganhos de produtividade, qualidade de produtos e serviços, redução de custos, eliminação de desperdícios, medições com resultados confiáveis e comparáveis.

3.1.1.

Histórico da Metrologia

O aparecimento de sistemas organizados de medição se confunde com a história da humanidade, pois o conhecimento sobre o mundo e a capacidade de atuar sobre ele tomando decisões depende em grande parte de nossa habilidade em fazer medições, ou seja, de quantificar atributos por meio de um processo de comparação com padrões (Instituto Euvaldo Lodi, 2005).

Medidas e regras bem definidas, provavelmente surgiram com o advento da agricultura na Síria e o Irã, por volta dos 6000 A.C. A necessidade de calcular estoques de alimentos e rações levou às primeiras medidas de volume de grãos que cabia em uma mão (Instituto Euvaldo Lodi, 2005).

Há cerca de 3000 A.C., os egípcios usavam um sistema baseado no “cúbito”, que consistia na distância do cotovelo à ponta do dedo maior da mão, com cerca de 45 centímetros. Esta unidade era subdividida em 24 “dedos”, sendo a largura de um dedo um pouco inferior a 19 milímetros (Instituto Euvaldo Lodi, 2005).

Com o domínio romano, há mil anos, o cúbito foi substituído pelo pé que era constituído de 12 polegadas, sendo esta igual ao comprimento da segunda falange do polegar da mão do homem.

A jarda, no século XII, cujo nome provavelmente foi devido ao esporte de arco e flecha popular nessa época, foi definida como sendo a distância da ponta do nariz do Rei Henrique I até o polegar. Só foi oficializada como unidade de comprimento em 1558 pela Rainha Elizabeth e materializada por uma barra de bronze. A jarda, como é hoje conhecida, foi estabelecida em 1878 como sendo a distância entre os terminais de ouro de uma barra de bronze, medida a 62° F (18° C) (Instituto Euvaldo Lodi, 2005).

Ainda neste mesmo século, fixou-se o pé que equivale a 16 “dedos”, cerca de 30 centímetros, como unidade de comprimento. As medidas eram ditas antropométricas, pois utilizava o corpo humano como referência para as medições (Instituto Euvaldo Lodi, 2005).

Como as unidades existentes eram variáveis de região para região, muitos aproveitavam para roubar ou tirar vantagem da falta de um padrão para determinar quantidades de produtos. Nesse período, na Europa Continental, especificamente na França, procurou-se uma forma de definir um padrão de comprimento que não dependesse da estatura da família real. Assim, por volta de 1790, definiu-se o metro utilizando como referência o meridiano da terra - décima milionésima parte (10⁻⁷) do quadrante do meridiano terrestre que parte do Equador e atinge o Polo Norte, passando por Paris (Instituto Euvaldo Lodi, 2005; Albertazzi e Souza, 2008).

Em 1837 foram refeitos os cálculos, obtendo-se, valores ligeiramente diferentes, por isso, a definição do metro foi alterada e passou a ser : "a distância medida, à temperatura do gelo fundente, entre dois traços gravados em uma barra de platina iridiada, depositada no Bureau Internacional des Poids et Mesures (BIPM), e considerado o protótipo do metro pela Primeira Conferência Geral de

Pesos e Medidas, e 1889, esta barra é apoiada sobre roletes nos pontos de deflexão mínima" (Figura 3.1).



Figura 2. Prototipo do Metro.

Durante o primeiro Império, foram feitas diversas tentativas de uniformização das unidades de medida brasileiras. Mas apenas em 26 de junho de 1862, Dom Pedro II promulgava a Lei Imperial nº 1157 e com ela oficializava, em todo o território nacional, o sistema métrico decimal francês. O Brasil foi uma das primeiras nações a adotar o novo sistema, que seria utilizado em todo o mundo (Instituto Euvaldo Lodi, 2005).

Com o crescimento industrial do século XX, fazia-se necessário criar no país instrumentos mais eficazes de controle que viessem a impulsionar e proteger produtores e consumidores em assuntos que envolvem as medições.

Em 20 de maio de 1875, 17 países assinaram a Convenção do Metro, criando “Bureau International des Poids et Mesures” (BIPM), que teve como objetivo inicial promover e ampliar o sistema métrico por todo o mundo (Márcio, 2008).

Em 1889, na primeira Conferência Geral de Pesos e Medidas (CGPM) foi redefinido um padrão para o metro e adotado um padrão para a massa (quilograma) e tempo (segundo), criando o sistema denominado MKS (Márcio, 2008).

Em 1960, por ocasião da décima primeira Conferência Geral de Pesos e Medidas, foi criado o Sistema Internacional de Unidades (SI) que passou a ser formado pelas unidades do MKS com a adição de três unidades, que são: a intensidade de corrente elétrica (ampere); a temperatura termodinâmica (kelvin); e a intensidade luminosa (candela). Em 1971, foi acrescentada a unidade de

quantidade de matéria (mol) ao SI, estabelecendo sete unidades de base (Inmetro, 1999).

O SI compreende as unidades de base (quadro 3.1) e as unidades derivadas (como, por exemplo, o metro por segundo, o volt e o watt), sendo que cada uma das grandezas das unidades derivadas pode ser obtida pela multiplicação ou divisão entre unidades de base (Márcio, 2008).

Quadro 1: Unidades de base do Sistema Internacional de Unidades (SI).

Unidades de base	Símbolo
Metro – comprimento	m
Quilograma – massa	kg
Segundo – tempo	s
Ampere – corrente elétrica	A
Kelvin – temperatura	K
Mol – quantidade de substância	mol
Candela – intensidade luminosa	cd

Sistemas de medição são referenciados a padrões que, por sua vez, são referenciados a outros padrões de qualidade sucessivamente melhor. O elemento no topo dessa cadeia é a definição da unidade de medida que, por sua vez está relacionada com as definições das unidades de base do sistema internacional de unidades. Diz-se de um sistema de medição que se insere nessa cadeia que sua calibração está rastreada aos padrões internacionais (Albertazzi, 2008).

Rastreabilidade é a propriedade do resultado de uma medição ou do valor de um padrão estar relacionado a referências estabelecidas, geralmente padrões nacionais ou internacionais, por meio de uma cadeia contínua de comparações, todas tendo incertezas estabelecidas.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

A rastreabilidade é uma característica muito importante dos padrões e dos sistemas de medição. O referenciamento de todos os sistemas de medição calibrados aos padrões internacionais garante uniformidade nas medições efetuadas em qualquer momento e em qualquer lugar no planeta. Assim, o metro australiano, o metro russo, o metro tibetano e mesmo o metro de Itu, são equivalentes (Albertazzi e Souza, 2008).

3.1.2. Um breve histórico sobre a metrologia no Brasil

O Brasil participou da criação da Convenção do Metro em 1875, mas só em 1921 fez sua adesão à convenção. Em 1931, sem pagar as anuidades, o País teve que se desligar novamente da Convenção. Em 1933, foi criado o Instituto Nacional de Tecnologia (INT) e, em 1934, o Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT) (Dias, 1998).

No início do Estado Novo, em 1938, através do Decreto-lei nº 592, o Brasil passou a adotar como padrão o sistema de medidas definido pela CGPM.

Em 1940, foi criada a Associação Brasileira de Normas Técnicas (ABNT). No entanto, a aplicação da legislação metrológica era compulsória apenas nas capitais dos estados e no Distrito Federal (Rio de Janeiro).

Em 1953, após o pagamento das anuidades atrasadas, o Brasil retornou a Convenção do Metro. Em 1961, foi criado o Instituto Nacional de Pesos e Medidas (INPM) e extintas a Comissão de Metrologia e a Divisão de Metrologia do INT.

Em 1967, foi formulada uma Política Nacional de metrologia, expressa no decreto-lei nº 240, que mencionava a criação do Sistema Nacional de Metrologia, cujo órgão central seria o INPM. Nessa mesma época, devido à necessidade de autonomia tecnológica, iniciou-se um convênio de cooperação técnica, firmado com o governo alemão e concretizado por meio do PTB (*Physikalisch-Technische Bundesanstalt*) que auxiliou o governo brasileiro no treinamento, compra de equipamentos e definição do espaço físico para a criação de um Laboratório Nacional de Metrologia. Dessa forma, nasceu o Sistema Nacional de Metrologia, Normalização e Qualidade Industrial (SINMETRO), instituído pela Lei no 5.966,

de 11 de dezembro de 1973, com a finalidade de formular e executar a política nacional de metrologia, normalização industrial e certificação de qualidade de produtos industriais. É composto por dois órgãos: um normativo, o Conselho Nacional de Metrologia - Conmetro, e outro executivo, o Instituto Nacional de Metrologia, Normalização e Qualidade Industrial – Inmetro (Albertazzi, 2008).

No âmbito de sua ampla missão institucional, o Inmetro objetiva a promoção da qualidade de vida do cidadão e a competitividade da economia brasileira através da Metrologia e da Avaliação da Conformidade. O Inmetro promove o fortalecimento das empresas nacionais, aumentando a sua produtividade por meio da adoção de mecanismos destinados à melhoria da qualidade de produtos e serviços. O Inmetro também atua nos produtos pré-medidos (medidos sem a presença do consumidor) e nos instrumentos de medição para garantir a correção das medidas utilizadas nas transações comerciais e no setor da saúde.

O Inmetro é reconhecido internacionalmente como o organismo de acreditação brasileiro, atuando como coordenador da Rede Brasileira de Metrologia Legal e Qualidade (RBMLQ) e acredita organismos de certificação, inspeção, treinamento, laboratórios de calibração e laboratórios de ensaios. Os laboratórios acreditados são incluídos na Rede Brasileira de Calibração (RBC) ou na Rede Brasileira de Laboratórios de Ensaio (RBLE) cujas competências podem ser consultadas no *website* do Inmetro. As atividades relacionadas à acreditação e à avaliação da conformidade no Sinmetro são baseadas nas normas e guias da ABNT, ISO, IEC, Copant (Comissão Panamericana de Normalização Técnica) e Mercosul (Mercado Comum do Cone Sul), nas recomendações da OIML (*Organisation Internationale de Métrologie Légale*) e nas orientações do IAF (*International Accreditation Forum*), ILAC (*International Laboratory Accreditation Cooperation*), IATCA (*International Auditor and Training Certification Association*) e IAAC (*Interamerican Accreditation Cooperation*).

Avaliação da Conformidade é o processo sistematizado, com regras pré-estabelecidas, devidamente acompanhado e avaliado, de forma a propiciar adequado grau de confiança de que um produto, processo ou serviço, ou ainda um profissional, atende a requisitos pré-estabelecidos por normas ou regulamentos, com o menor custo possível para a sociedade.

(Conmetro, 2008)

As componentes da metrologia do Inmetro estão agrupadas em duas grandes diretorias: Diretoria de Metrologia Científica e Industrial e a Diretoria da Metrologia Legal (Albertazzi e Souza, 2008; Inmetro, 2007a, 2007b).

Metrologia Científica – organiza, desenvolve e conserva os padrões de medida;

Metrologia Industrial – assegura o funcionamento adequado de instrumentos de medição usados na indústria, garantindo a qualidade dos produtos;

Metrologia Legal – relacionada aos sistemas de medição usados nas transações comerciais e nos setores da saúde, segurança e meio ambiente.

(Inmetro, 2007a, 2007b)

3.2. Psicometria

Psicometria, do grego *psyké* (alma) e *metron* (medida ou medição), é uma área da Psicologia que faz a ponte entre as ciências exatas, principalmente estatística, e a Psicologia (Pasquali, 1997). Segundo Erthal (2003), também pode ser definida como o conjunto de técnicas que permite a quantificação dos fenômenos psicológicos.

Os testes Psicológicos são ferramentas com características objetivas e padronizadas de uma amostra de comportamento (Anastasi, 2000). Tais testes quantificam o constructo por meio de sua especificação em itens (que correspondem a cada uma das questões de um inventário).

3.2.1. Histórico das Medições em Psicologia

A realização de medições em Psicologia ocorreu a partir do século XIX. Em 1879, em Leipzig (Alemanha) surgiu o primeiro laboratório de psicologia experimental fundado por Wilhelm Wundt. Os primeiros psicólogos

experimentais estavam interessados nas semelhanças e não nas diferenças entre os indivíduos, obtendo então descrições generalizadas do comportamento humano (Anastasi, 2000). Os problemas testados nesses laboratórios, eram em grande parte, relacionados à sensibilidade aos estímulos visuais, auditivos e outros estímulos sensoriais e, ao tempo de reação. Esta ênfase dada nos fenômenos sensoriais refletiu a natureza dos primeiros testes psicológicos (Anastasi, 2000).

Francis Galton (biólogo) no intuito de constatar as semelhanças e diferenças entre os indivíduos com e sem grau de parentesco, criou instrumentos de medição e, procurou estabelecer relação entre os resultados. Galton e seu aluno Karl Pearson criaram um método correlacional para analisar as medidas psicológicas (Anastasi, 2000).

No final do século XIX, James McKeen Cattell (psicólogo) e Galton construíram testes do tipo sensorio-motor, que tinham como objetivo a mensuração da inteligência. Esses instrumentos mediam o tempo de reação e a velocidade sensorial e acreditavam que os indivíduos mais capazes tinham um tempo de reação melhor. Mesmo sem medir o aspecto intelectual, como acreditava, este fato contribuiu para unir o movimento dos testes à psicologia experimental. Em 1890 num artigo escrito por Cattell o termo “teste mental” foi utilizado pela primeira vez na literatura psicológica (Anastasi, 2000).

Binet e Simon criticaram as escalas até então elaboradas por serem muito sensoriais. Em 1905, eles elaboraram a Escala de Binet-Simon, que ficou conhecida como “escala de 1905”, com o objetivo de investigar as reprovações escolares. Foi a primeira tentativa de estudar as diferenças entre os indivíduos referentes à inteligência (Anastasi, 2000). Em 1908, foi feita a primeira modificação, onde foi aumentado o número de testes e agrupados em níveis de idade. O escore da criança no teste completo podia ser expresso como um nível mental correspondente à idade das crianças normais cujo desempenho fosse comparado. O termo nível mental foi substituído por idade mental (Anastasi, 2000). Em 1911, houve uma outra modificação da escala, estendendo-se até a idade adulta. Neste período, Binet falece prematuramente.

Em 1916, Lewis M. Terman, desenvolveu uma versão e deu-lhe o nome de Stanford-Binet, em homenagem à universidade à qual era afiliado e, adotou o conceito de quociente de inteligência (QI) (Schultz, 2006).

Com a I Guerra Mundial, houve a necessidade de selecionar soldados com intuito de atribuir-lhes as tarefas mais adequadas e com isso a escala coletiva de desenvolvimento mental foi elaborada por Otis, incluindo questões de múltipla escolha e introduzindo também a aplicação coletiva (Schultz, 2006).

A inteligência não foi a única variável tida como objeto de investigação por parte dos psicólogos no início do século XX. A personalidade também foi estudada de modo que em 1918 foi elaborado por Woodworth o inventário de auto-descrição. Este instrumento permitiu que vários outros pudessem ser elaborados com base em técnicas mais exatas, como por exemplo a análise fatorial (Schultz, 2006).

A fim de dotar a recém-iniciada empreitada de credibilidade científica e autoridade, os psicólogos criadores dos testes de inteligência adotavam a terminologia de outras disciplinas mais antigas, como a medicina e a engenharia. O objetivo era convencer as pessoas de que a psicologia era tão legítima, científica e fundamental quanto as demais ciências há tempos estabelecidas (Schultz, 2006).

Com o fim da I Guerra Mundial houve um aumento da demanda por parte do comércio, da indústria e do governo pelos serviços dos psicólogos industriais para a reorganização dos seus procedimentos de pessoal e a preparação de testes psicológicos para a seleção de funcionários (Schultz, 2006).

Grande parte da pesquisa sobre testes, realizada nas forças armadas, baseou-se na análise fatorial e foi dirigida para a construção de baterias de aptidões múltiplas (Anastasi, 2000). Durante a Segunda Guerra Mundial, os psicólogos atuaram elaborando testes para seleção de pessoal e lidando com problemas clínicos de esgotamento provocados pelas batalhas, depressão e, assim por diante (Kantowitz, 2006).

Em oposição à ausência de especificidade das medidas fisiológicas, psicometristas têm desenvolvido escalas para avaliação de estados psicológicos específicos por meio de questionários, inventários ou escalas de auto-avaliação. Essas medidas, que podem ser denominadas de explícitas (considera uma reflexão consciente por parte do sujeito que está sendo avaliado) ou implícitas (o sujeito não possui o controle consciente sobre a medição), começaram a surgir a partir da segunda metade do século XX (Cunha, 2008; Landeira, 1998).

3.2.2. Instrumentos de medição para quantificação dos transtornos de humor e ansiedade

3.2.2.1. Medição da intensidade de sintomas de Ansiedade

A resposta fisiológica mais utilizada para avaliar estados de ansiedade em seres humanos é o reflexo galvânico da pele (Hugdahl, 1988). Essa resposta é medida por meio de um galvanômetro e está relacionada com a ativação do sistema nervoso simpático, produzindo através das glândulas sudoríparas, uma pequena quantidade de solução iônica capaz de alterar a condutância elétrica da pele. Além da resposta galvânica da pele, outras respostas controladas pelo sistema nervoso autônomo, como por exemplo, frequência cardíaca, pressão arterial e dilatação da pupila, têm sido utilizadas como medidas de ansiedade em seres humanos. No entanto, a validade de constructo dessas medidas tem sido criticada porque essas mesmas respostas autonômicas são compartilhadas com outras emoções (por exemplo, raiva, alegria, excitação sexual) que também têm a capacidade de mobilizar o sistema nervoso autônomo (Cacioppo, et al 2000; Landeira, 1998).

Conforme já mencionado, metodologias mais específicas para avaliação de estados específicos de ansiedade por meio de questionários, inventários ou escalas de auto-avaliação foram desenvolvidas a partir da segunda metade do século XX.

Dentre essas escalas destacam-se:

- Escala de Ansiedade de Hamilton (HAM-A; Hamilton, 1959);
- Escala breve de avaliação psiquiátrica (BPRS; Overall et al., 1962);
- Inventário de ansiedade traço-estado (IDATE; Spielberger et al., 1970);
- Escala de Ansiedade de Zung (Zung, 1971);
- Subescala de ansiedade do Symptom Checklist (SCL-90; Derogatis et al., 1973);
- Escala Clínica de Ansiedade (CAS; Snaith, Baugh, Clayde, Husai, Sipple, 1982);
- Escala Hospitalar de Ansiedade e Depressão (HAS; Zigmond e Snaith, 1983);
- Escala Breve de Ansiedade (BAS; Tyrer, et al, 1984);
- POMS (Profile of Mood States—POMS; Lorr e McNair, 1984) e

- Inventário de Ansiedade de Beck (BAI; Beck, Epstein, Brown e Steer, 1988).

3.2.2.2.

Medição da intensidade de sintomas de Transtorno de Humor

O diagnóstico da depressão é feito com base no exame clínico do paciente de acordo com as queixas relatadas (sintomas) e a história pessoal e familiar constituem as principais fontes de informação que permitem firmar o diagnóstico. “É fundamental que se faça o diagnóstico diferencial, para não haver um diagnóstico equivocado que levará a um tratamento inadequado e, conseqüentemente um insucesso terapêutico” (Nardi, 2006).

Não existem exames laboratoriais ou complementares que possam dar o diagnóstico de depressão. Nem mesmo métodos mais modernos como tomografia por emissão de pósitrons (PET scan), tomografia por emissão de fóton único (SPECT) e espectroscopia fornecem informações suficientes para se definir um padrão característico da depressão. Desta forma, a pesquisa dos sintomas é fundamental para o correto diagnóstico (Nardi, 2006).

Os principais sintomas além do humor triste são também a perda do interesse e prazer (anedonia). Acrescentam-se diversos sintomas físicos e psicológicos, como perda de energia (cansaço), alterações do apetite, distúrbios do sono, dores, sensação de desconforto torácico, alterações psicomotoras, baixa auto-estima e sentimentos de culpa, dificuldade de concentração e memória, retraimento social, uso e abuso de drogas, problemas no trabalho, irritabilidade, distorção da percepção da realidade, ideação suicida e diminuição da libido (Nardi, 2006). Dentre as escalas que quantificam os sintomas da depressão destacam-se:

- Escala de Avaliação de Depressão de Hamilton (HAM-D, 1960)
- Escala de ansiedade de Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J.. 1961).
- Escala de Depressão de Zung - SDS (Zung, 1965)
- Escala Analógica de Humor – (Norris, 1971)
- Escala de Avaliação para Depressão de Montgomery & Åsberg (MADRS, 1979);

- Escala de Depressão de Montgomery-Asberg – MADRS (Montgomery, 1979)
- Escala de Depressão de Beck – BDI (Beck, A. T. & Steer, R. A., 1993);

3.2.3.

Técnicas utilizadas em psicometria para avaliar a qualidade dos instrumentos de medição utilizados para quantificação dos fenômenos psicológicos

OBSERVAÇÃO: Neste item 3.2.3 constam algumas definições de termos utilizados em psicometria que não estão de acordo com as definições universais para os termos referentes a medições, propostas pelo vocabulário internacional de metrologia. Neste sentido, serão destacados por meio da formatação em itálico e no item 3.3 é apresentada proposta para harmonização da terminologia.

Para avaliar a qualidade das escalas psicométricas, classicamente, alguns parâmetros são testados para verificar se a medida obtida com instrumento em análise fornece os mesmos resultados mediante a mesma quantidade de traço latente (estabilidade), se possui consistência com o constructo a ser medido (fidedignidade), se mede o que pretende medir (validade) (Field, 2005; Streiner e Norman, 2003; Banton e Jaccard, 2006; Kazdin, 2006). Estas abordagens estão inclusas na Teoria Clássica dos Testes (TCT), tendo por foco explicar o resultado final total do teste (soma das respostas dadas a uma série de itens).

Recentemente uma abordagem estatística denominada Teoria de Resposta ao Item (TRI) vem sendo utilizada em psicometria, tendo como característica o interesse específico por cada um dos itens dos testes, e não o escore total, como na TCT. A TRI considera a existência de fatores que afetam a probabilidade de desempenho adequado em cada item.

3.2.3.1.

Fidedignidade

Conforme a definição de Erthal (2003) “A fidedignidade de um instrumento refere-se ao fato de os resultados terem sido reproduzidos por um teste em diferentes ocasiões nas quais se mantiveram condições similares, inclusive os mesmos indivíduos ou grupos de indivíduos, variando somente de tempo de

aplicação”. Nesta publicação de Erthal (2003), são descritos três aspectos fundamentais para a fidedignidade. O primeiro é a precisão, que a autora explica o termo como “implica em medir sem erro”. O segundo aspecto destacado é a estabilidade, que a autora define como “implica reproduzir diferentes fenômenos, havendo pouco erro de mensuração em tempos diferentes (comenta que a fidedignidade é maior quanto mais estável for o traço psicológico explorado)”. O último aspecto indicado por Erthal para fidedignidade é a consistência interna ou homogeneidade, que definiu como “implica que todos os itens do teste meçam um mesmo aspecto” (Erthal, 2003). Além do uso de termos não conformes ao vocabulário utilizados no parágrafo acima, como será detalhado no item 3.3, deve-se destacar que, considerando que o erro está sempre presente em qualquer medida a afirmação “medir sem erro” de Erthal se refere à determinação de quanto o parâmetro estudado se afasta do valor unitário (considerado como medir sem erro).

Assim como Erthal (2003), Pasquali (2003) indica que a fidedignidade dos testes vem referenciada sob uma série elevada e heterogênea de nomes como: precisão, confiabilidade, consistência interna, estabilidade, constância.

Existem alguns fatores, tanto ligados ao próprio teste, quanto ligados ao examinando, que podem alterar a fidedignidade de um instrumento. Tais fatores são listados a seguir (Erthal, 2003):

- Estabilidade temporal (influência de flutuações diárias (estresse, fadiga) e flutuações referentes à aplicação do teste (ruídos, mobiliários) nos resultados obtidos em testes realizados em diferentes ocasiões).
- Generalidade da amostragem de itens (relação direta dos resultados obtidos com a seleção dos itens que compõem a escala).
- Homogeneidade dos itens (capacidade do instrumento medir uma mesma habilidade, função ou conhecimento).

3.2.3.2. Técnicas de estimativa da Fidedignidade

Existem três técnicas de estimativa da fidedignidade: teste-reteste (repetição idêntica do teste); formas paralelas (formas equivalentes de um teste) e método das duas metades (Erthal, 2003; Cunha, 2008)

a - Teste-reteste

O mesmo grupo de sujeitos é testado em duas ocasiões diferentes. Para estabelecer a correlação entre as duas situações, utiliza-se a correlação linear de Pearson.

Supõe-se que o constructo que está sendo medido se apresente relativamente estável ao longo do tempo, então, quanto maior a correlação, menos suscetíveis às mudanças aleatórias diárias nas condições dos testados ou do ambiente de testagem os escores estarão.

Uma das limitações desse método se deve à memorização de algumas respostas (efeito-memória) e também aos comentários dos examinandos no espaço entre a aplicação dos testes (efeito-prática). Ainda que o intervalo entre o teste e o reteste seja longo, o erro de mensuração pode ser confundido com mudanças reais decorrentes da aprendizagem na habilidade do examinando. Outra limitação decorrente desta técnica é a falta de motivação por parte do examinando de executar novamente o teste, o que resulta em uma medida mais pobre do que a primeira (Erthal, 2003).

b - Formas paralelas ou formas equivalentes de um teste

Consiste na aplicação, numa primeira ocasião, de um teste e na segunda, aplica-se sua forma equivalente no mesmo grupo de sujeitos. Utiliza-se a correlação linear de Pearson, para avaliar o grau de associação entre as duas formas.

Se duas formas do teste forem aplicadas numa mesma ocasião, o coeficiente de equivalência expressará uma generalidade de diferentes amostras de itens. Caso contrário (em situações diferentes), o coeficiente expressará uma estabilidade temporal.

c- Método das duas metades

Divide-se um único teste em duas metades razoavelmente equivalentes. Aplica-se o teste a um grupo de sujeitos e assim se obtém dois grupos de escores que, correlacionados, darão o índice de fidedignidade do teste, chamado de coeficiente de consistência interna (alfa de Cronbach), também conhecido como coeficiente de fidedignidade bipartida.

3.2.3.3. Validade

A validade de um teste refere-se à capacidade de o teste medir aquilo que se propõe (Tyler, 1973; Pasqualli, 2003; Erthal, 2003; Anastasi, 2000). Um instrumento é válido quando as diferenças de resultados obtidas por um instrumento refletem, necessariamente, diferenças reais entre indivíduos ou entre o mesmo indivíduo em ocasiões diferentes (Erthal, 2003). A validade diz o que podemos inferir dos escores de teste. Em relação a isso devemos ter cuidado para não aceitar o nome do teste como índice daquilo que se mede. “Os nomes dos testes são rótulos curtos e convenientes para fins de identificação (Anastasi, 2000)”.

Existem fatores que afetam a validade do teste como:

- Heterogeneidade dos itens: costuma diminuir a precisão dos testes, mas, em geral, aumenta a validade (Erthal, 2003).
- Representatividade da amostra: um teste que apresente resultados consistentes no sentido de que mede apenas características relevantes de alguma área de conteúdo, não será válido (Erthal, 2003).
- Heterogeneidade do grupo: quanto mais heterogêneo, mais alto será o coeficiente de fidedignidade (Erthal, 2003)

3.2.3.4. Tipos de validade

Segundo Anastasi (1988), na edição de 1985 do manual *Standards for Educational and Psychological Testing*, foi proposto que a nomenclatura sobre validade de medidas fosse agrupada em três categorias principais:

a – Validade de Conteúdo: consiste na análise do conteúdo do teste de maneira a avaliar se os seus itens representam o comportamento que se quer medir. O controle desse tipo de validade é realizado pela análise de itens, que é um questionamento feito a diversos juízes (especialistas) que apontam os objetivos relevantes a medir e analisam a representatividade dos itens.

b – Validade de Critério: consiste na capacidade do teste de funcionar como um preditor presente ou futuro de outra variável, chamada critério (Cunha, 2008).

Para sua avaliação utiliza-se a correlação entre os resultados do teste e as condutas subseqüentes tomadas como critério (Erthal, 2003). Esta validade é subdividida em validade concorrente (preditor presente) e validade preditiva (preditor futuro).

c – Validade de Conceito ou Constructo ou Teórica: trata-se do grau de capacidade de um teste para medir o constructo teórico para o qual ele foi designado. Segundo Anastasi (1988), a validade está relacionada ao constructo e o conhecimento sobre os atributos do mesmo requer a acumulação gradual de informação a partir de várias fontes (Cunha, 2008; Anastasi, 1988). Subdividi-se em validade convergente (quando a medida em questão está relacionada a outras formas de medir o mesmo constructo), validade discriminante (a medida em questão não está relacionada com outros constructos) e validade fatorial (verifica a unidimensionalidade do constructo que está sendo medido).

3.3. Harmonização entre os termos da Psicometria e da Metrologia

Para que seja possível a comunicação entre os vários ambientes da ciência que realizam medições é importante a uniformização de suas terminologias. A metrologia possui um vocabulário estabelecido internacionalmente (Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia - VIM) contendo definições universais para os termos referentes a medições. Sem a uniformização dos termos, existe a possibilidade de um entendimento equivocado, podendo gerar ambigüidade, dificultando com isso a interpretação dos resultados de uma pesquisa.

A Psicometria tem uma terminologia própria, mas seria interessante que os termos por esta utilizados fossem harmonizados às definições do VIM para facilitar o entendimento do vocabulário existente, em nível interdisciplinar e universal.

No Brasil, a última edição do VIM foi publicada em 2007, tendo por base a segunda edição internacional do vocabulário que fora publicada em 1993. Em 2008 foi publicada a 3ª edição internacional do VIM que ainda não adotada no Brasil até abril de 2009. Os comentários apresentados a seguir têm por base o VIM publicado em 2007 no Brasil (quando houver a terminologia específica em

seu conteúdo) ou a 3ª edição internacional do VIM, publicada em 2008 e ainda não adotada no Brasil (quando a terminologia não estiver contida no VIM 2007).

A seguir são re-apresentadas algumas definições em psicometria, já abordadas no item 3.2 do presente capítulo, destacando-se os termos utilizados em psicometria e apresentando sua harmonização com a terminologia convencionada internacionalmente no VIM.

Conforme descrito no item 3.2.3, para maior confiabilidade dos questionários que quantificam os constructos psicológicos, estes são avaliados quanto a sua “fidedignidade” e “validade”. Uma avaliação crítica da terminologia utilizada pela psicometria para a definição de tais termos é apresentada a seguir por meio de comparação com as definições dos termos correspondentes segundo o VIM (cujas definições estão destacadas para consulta). Ao final do presente item, apresenta-se uma proposta de harmonização dos termos usados para conceituar “fidedignidade” e “validade” segundo as definições universais do VIM.

3.3.1.

Harmonização da terminologia utilizada para definir Fidedignidade

Conforme descrito no item 3.2.3, a publicação em psicometria de Erthal (2003) definiu a avaliação da denominada fidedignidade de um instrumento de medição em psicologia como referindo-se ao fato de os resultados de um teste psicométrico terem sido reproduzidos em diferentes ocasiões nas quais se mantiveram as condições de aplicação (Anastasi, 2000; Erthal, 2003; Pasquali, 2003; Cunha, 2008).

O termo “fidedignidade”, que significa confiabilidade, poderia ser entendido como o aspecto geral a ser avaliado por todas as abordagens que estudam os inventários, e portanto não apropriado para denominar uma das avaliações em particular. Considerando a terminologia do VIM, a característica descrita pela psicometria para definir o termo “fidedignidade” é denominada de Reprodutibilidade.

Reprodutibilidade

Grau de concordância entre os resultados das medições de um mesmo mensurando efetuadas sob condições variadas de medição.

Observações:

1) Para que uma expressão da reprodutibilidade seja válida, é necessário que sejam especificadas as condições alteradas.

2) As condições alteradas podem incluir:

- princípio de medição;
- método de medição;
- observador;
- instrumento de medição;
- padrão de referência;
- local;
- condições de utilização;
- tempo.

3) Reprodutibilidade pode ser expressa, quantitativamente, em função das características da dispersão dos resultados.

4) Os resultados aqui mencionados referem-se, usualmente, a resultados corrigidos.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

No entanto, ao realizar maior detalhamento da definição de *fidedignidade*, a publicação de Erthal descreve três aspectos fundamentais. Comparando as definições para estes aspectos apontados pelas publicações em psicometria, os conflitos de terminologia ficam mais evidentes e suas harmonizações são discutidas a seguir:

Aspecto 1 - *precisão* – implica em “medir sem erro” (Erthal, 2003; Pasquali, 2003; Cunha, 2008).

Correlação com a terminologia da metrologia: Segundo o VIM o termo correspondente para a aptidão de um instrumento de medição em dar respostas próximas a um **valor verdadeiro** (“medir sem erro”) é denominado exatidão. O termo precisão não se encontra na 2ª edição internacional do VIM (1993) que corresponde à edição adotada no Brasil (VIM 2007), sendo que neste documento ressalta-se que “o termo precisão não deve ser utilizado como exatidão”. O VIM 2008, que ainda não foi adotado na língua portuguesa pelo INMETRO, introduz o termo precisão, mas com o significado de proximidade dos valores obtidos por meio de medições sob condições especificadas (o que corresponde à repetitividade e/ou reprodutibilidade (VIM 1993 e VIM 2008), e não à exatidão como definido

neste primeiro aspecto. Apesar da denominação do primeiro aspecto da fidedignidade ser precisão, sua definição psicométrica não apresenta características de repetitividade e/ou reprodutibilidade, e sim de exatidão.

Valor Verdadeiro (de uma grandeza)

Valor consistente com a definição de uma dada grandeza específica.

Observações:

- 1) É um valor que seria obtido por uma medição perfeita.
- 2) Valores verdadeiros são, por natureza, indeterminados.
- 3) O artigo indefinido “um” é usado, preferivelmente ao artigo definido “o” em conjunto com “valor verdadeiro”, porque pode haver muitos valores consistentes com a definição de uma dada grandeza específica.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Valor Convencional (de uma grandeza)

Valor atribuído a uma grandeza específica e VERDADEIRO aceito, às vezes por convenção, como tendo uma incerteza apropriada para uma dada finalidade.

Exemplos:

- a) Em um determinado local, o valor atribuído a uma grandeza, por meio de um padrão de referência, pode ser tomado como um valor verdadeiro convencional;
- b) O CODATA (1986) recomendou o valor para a constante de Avogadro como sendo $A = 6,022\ 136\ 7 \times 10^{23} \text{ mol}^{-1}$.

Observações:

- 1) “Valor verdadeiro convencional” é, às vezes, denominado valor designado, melhor estimativa do valor, valor convencional ou valor de referência.
“Valor de referência”, neste sentido, não deve ser confundido com “valor de referência” no sentido usado na observação do item 5.7.
- 2) Frequentemente, um grande número de resultados de medições de uma grandeza é utilizado para estabelecer um valor verdadeiro convencional.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Exatidão de Medição

Grau de concordância entre o resultado de uma medição e um valor verdadeiro do mensurando.

Observações:

- 1) Exatidão é um conceito qualitativo.

2) O termo precisão não deve ser utilizado como exatidão.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Exatidão de um instrumento de medição

Aptidão de um instrumento de medição para dar respostas próximas a um valor verdadeiro.

Observação: exatidão é um conceito qualitativo.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Repetitividade (de resultados de medições)

Grau de concordância entre os resultados de medições sucessivas de um mesmo mensurando efetuadas sob as mesmas condições de medição.

Observações:

1) Estas condições são denominadas condições de repetitividade.

2) Condições de repetitividade incluem:

- mesmo procedimento de medição;
- mesmo observador;
- mesmo instrumento de medição, utilizado nas mesmas condições;
- mesmo local;
- repetição em curto período de tempo.

3) Repetitividade pode ser expressa, quantitativamente, em função das características da dispersão dos resultados

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Repetitividade (de um instrumento de medição)

Aptidão de um instrumento de medição em fornecer indicações muito próximas, em repetidas aplicações do mesmo mensurando, sob as mesmas condições de medição.

Observações:

1) Estas condições incluem:

- redução ao mínimo das variações devidas ao observador;
- mesmo procedimento de medição;
- mesmo observador;
- mesmo equipamento de medição, utilizado nas mesmas condições;
- mesmo local;
- repetições em um curto período de tempo.

2) Repetitividade pode ser expressa quantitativamente em termos das características da dispersão das indicações.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Measurement precision

Closeness of agreement between **indications** or **measured quantify values** obtained by replicate **measurements** on the same or similar objects under specified conditions

NOTE 1 – Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance or coefficient of variation under the specified conditions of measurement.

*NOTE 2 – The “specified conditions” can be, for example, **repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement** (see ISO 5725-3:1994)*

*NOTE 3 – Measurement precision is used to define **measurement repeatability, intermediate measurement precision, and measurement reproducibility.***

*NOTE 4 – Sometimes “measurement precision” is erroneously used to mean **measurement accuracy.***

International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM 2008)

Aspecto 2 – **estabilidade** - implica reproduzir diferentes fenômenos, em tempos diferentes, havendo pouco erro de mensuração (Erthal, 2003).

Correlação com a terminologia da metrologia: Segundo o VIM, esta definição do aspecto 2 corresponde à reprodutibilidade e à exatidão. Para este segundo aspecto, a definição em psicometria repete a idéia de exatidão (referida no primeiro aspecto) e introduz a reprodutibilidade que fora mencionada na definição geral de fidedignidade. Segundo o VIM (2007) o termo **estabilidade** tem definição diferente da utilizada pela psicometria.

Estabilidade

Aptidão de um instrumento de medição em conservar constantes suas características metrológicas ao longo do tempo.

Observações:

1) Quando a estabilidade for estabelecida em relação a uma outra grandeza que não o tempo, isto deve ser explicitamente mencionado.

2) A estabilidade pode ser quantificada de várias maneiras, por exemplo:

- pelo tempo no qual a característica metrológica varia de um valor determinado; ou
- em termos da variação de uma característica em um determinado período de tempo.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Aspecto 3 – **consistência interna ou homogeneidade** – implica que todos os itens do teste meçam o mesmo aspecto.

3.3.2.

Harmonização da terminologia utilizada para definir validade

Segundo Erthal (2003), a “validade” refere-se à capacidade do teste de medir aquilo que se propõe. Diz-se que um instrumento é válido quando as diferenças de resultados obtidas com o instrumento refletem, necessariamente, diferenças reais entre indivíduos ou entre o mesmo indivíduo em ocasiões diferentes (Erthal, 2003).

Correlação com a terminologia da metrologia: Segundo a 3ª edição internacional do VIM (2008), a validade dos resultados de medição é altamente dependente das propriedades metrológicas do instrumento, determinadas 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. O intervalo dos valores atribuídos ao mensurando é o intervalo dos valores dos padrões que teriam fornecido as mesmas indicações. Segundo a mais recente edição do VIM (2008), a calibração consiste na operação que estabelece, numa primeira etapa e sob condições especificadas, uma relação entre os **valores** e as **incertezas de medição** fornecidos por **padrões** e as **indicações** correspondentes com as incertezas associadas; numa segunda etapa, utiliza esta informação para estabelecer uma relação visando a obtenção de um **resultado de medição** a partir de uma indicação.

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.

Calibração

Conjunto de operações que estabelece, sob condições especificadas, a relação entre os valores indicados por um instrumento de medição ou sistema de medição ou valores representados por uma medida materializada ou um material de referência, e os valores correspondentes das grandezas estabelecidos por **padrões**.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

Calibration

*Operation that, under specified conditions, in a first step, establishes a relation between the quantity values with **measurement uncertainties** provided by **measurement standards** and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication*

NOTE 1 A calibration may be expressed by a statement, calibration function, calibration diagram, calibration curve, or calibration table. In some cases, it may consist of an additive or multiplicative correction of the indication with associated measurement uncertainty.

NOTE 2 Calibration should not be confused with adjustment of a measuring system, often mistakenly called “self-calibration”, nor with verification of calibration.

NOTE 3 Often, the first step alone in the above definition is perceived as being calibration.

International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM 2008)

Metrological Compatibility of Measurement Results

*Property of a set of measurement results for a specified measurand, such that the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than some chosen multiple of the **standard measurement uncertainty** of that difference*

*NOTE 1 Metrological compatibility of measurement results replaces the traditional concept of ‘staying within the error’, as it represents the criterion for deciding whether two measurement results refer to the same measurand or not. If in a set of measurements of a measurand, thought to be constant, a measurement result is not compatible with the others, either the measurement was not correct (e.g. its **measurement uncertainty** was assessed as being too small) or the measured quantity changed between measurements.*

NOTE 2 Correlation between the measurements influences metrological compatibility of measurement results. If the measurements are completely uncorrelated, the standard measurement uncertainty of their difference is equal to the root mean square sum of their standard measurement uncertainties, while it is lower for positive covariance or higher for negative covariance.

International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM 2008)

3.3.3.**Proposta de Harmonização de Termos entre Psicometria e Metrologia**

No quadro 3.3.1 é apresentado um resumo da correlação de terminologia entre Psicometria e Metrologia, tomando-se por base os textos discutidos nos itens 3.3.1 e 3.3.2.

Quadro 3.2 – Correlação entre terminologias em Psicometria e Metrologia, tomando por base as definições dos termos utilizada em Psicometria (Erthal 2003).

Psicometria	Definições segundo publicações em Psicometria	Correspondência com a terminologia em Metrologia (VIM)
Fidedignidade	Precisão – Implica <u>medir sem erro</u>	<u>Exatidão</u> - Aptidão de um instrumento de medição para dar respostas próximas a um valor verdadeiro
	Estabilidade – <u>Reproduzir</u> diferentes fenômenos, em tempos diferentes, havendo <u>pouco erro</u> de mensuração	<u>Exatidão e Reprodutibilidade</u> <u>Exatidão</u> – Aptidão de um instrumento de medição para dar respostas próximas a um valor verdadeiro <u>Reprodutibilidade</u> – Grau de concordância entre os resultados das medições de um mesmo mensurando efetuadas sob condições variadas de medição.
	Homogeneidade – implica que todos os itens do teste meçam o mesmo aspecto.	
Validade	Medir aquilo que se propõe.	Calibração e Compatibilidade metrológica dos resultados de medição

Considerando a harmonização da terminologia, como exemplo, poder-se-ia propor a definição de fidedignidade e validade de um instrumento de medição em psicologia da seguinte forma:

“Para que um instrumento de medição em psicologia seja confiável são necessárias características como reprodutibilidade, exatidão e homogeneidade dos itens. Para que seja exato é necessária a calibração rastreada a um padrão, implicando assim, na análise da incerteza de medição”.

Incerteza de medição

Parâmetro, associado ao resultado de uma medição, que caracteriza a dispersão dos valores que podem ser fundamentadamente atribuídos a um mensurando.

Observações:

- 1) O parâmetro pode ser, por exemplo, um desvio padrão (ou um múltiplo dele), ou a metade de um intervalo correspondente a um nível de confiança estabelecido.
- 2) A incerteza de medição compreende, em geral, muitos componentes. Alguns destes componentes podem ser estimados com base na distribuição estatística dos resultados das séries de medições e podem ser caracterizados por desvios padrão experimentais. Os outros componentes, que também podem ser caracterizados por desvios padrão, são avaliados por meio de distribuição de probabilidades assumidas, baseadas na experiência ou em outras informações.
- 3) Entende-se que o resultado da medição é a melhor estimativa do valor do mensurando, e que todos os componentes da incerteza, incluindo aqueles resultantes dos efeitos sistemáticos, como os componentes associados com correções e padrões de referência, contribuem para a dispersão.

Vocabulário Internacional de Termos Fundamentais e Gerais de Metrologia (VIM 2007)

measurement uncertainty

*non-negative parameter characterizing the dispersion of the **quantity values** being attributed to a **measurand**, based on the information used*

*NOTE 1 Measurement uncertainty includes components arising from systematic effects, such as components associated with **corrections** and the assigned quantity values of **measurement standards**, as well as the **definitional uncertainty**. Sometimes estimated systematic effects are not corrected for but, instead, associated measurement uncertainty components are incorporated.*

*NOTE 2 The parameter may be, for example, a standard deviation called **standard measurement uncertainty** (or a specified multiple of it), or the half-width of an interval, having a stated **coverage probability**.*

*NOTE 3 Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by **Type A evaluation of measurement uncertainty** from the statistical distribution of the quantity values from series of **measurements** and can be characterized by standard deviations. The other components, which may be evaluated by **Type B evaluation of measurement uncertainty**, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information.*

NOTE 4 In general, for a given set of information, it is understood that the measurement uncertainty is associated with a stated quantity value attributed to the measurand. A modification of this value results in a modification of the associated uncertainty.

International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM 2008)