



doi: <http://dx.doi.org/10.7213/psicolargum.36.91.AO04>

Psicometria e Neuropsicologia: interrelações na construção e adaptação de instrumentos de medida

Psychometrics and Neuropsychology: inter-relations in the construction and adaptation of measuring instruments

RACHEL SCHLINDWEIN-ZANINI ^[a]

^[a] Psicóloga/Neuropsicóloga do Hospital Universitário – Universidade Federal de Santa Catarina (UFSC); Professora do Programa de Mestrado em Saúde Mental – CCS/UFSC, Preceptora da Residência Multiprofissional -RIMS/HU; Pesquisadora e coordenadora/líder do Núcleo de Neuropsicologia e Saude – HU/UFSC; Doutora em Ciências da Saúde/Medicina (Neurociências) - PUCRS; Pós-doutorado em Psicologia pela UFSC. Universidade Federal de Santa Catarina (UFSC). Florianópolis, Santa Catarina / Brasil. rachelsz@floripa.com.br

ROBERTO MORAES CRUZ ^[b]

^[b] Psicólogo especialista em avaliação psicológica; professor do Departamento e do Programa de Pós-Graduação em Psicologia da Universidade Federal de Santa Catarina (UFSC). Líder do Laboratório Fator Humano (UFSC) e pesquisador do Núcleo de Neuropsicologia e Saúde - HU/UFSC. Coordenador da Comissão Consultiva de Avaliação Psicológica do Conselho Federal de Psicologia. Universidade Federal de Santa Catarina (UFSC). Florianópolis, Santa Catarina / Brasil. roberto.cruz@cfp.org.br

Resumo

Estudos de parâmetros psicométricos em processos de construção, adaptação transcultural e validação de instrumentos de medida para fins clínicos são importantes procedimentos para assegurar rigor científico e eficácia operacional em situações de exame no campo neuropsicológico. OBJETIVO:

Caracterizar as interrelações entre a Psicometria e a Neuropsicologia no desenvolvimento de recursos instrumentais para a investigação de processos funcionais típicos e atípicos no contexto do diagnóstico e da intervenção neuropsicológica. MÉTODO: Realizar revisão teórica norteadora a profissionais e estudantes sobre o tema pertinente a avaliação psicológica. RESULTADOS E CONCLUSÃO: O conhecimento técnico acerca da qualidade dos instrumentos de medida, seja no campo da pesquisa ou da intervenção profissional é um importante fator na busca de evidências científicas primárias e secundárias ao fenômeno de interesse. Assim, há contribuições da área de Psicometria à construção de evidências e aprimoramento de instrumentos de medida em Neuropsicologia e na área de saúde em geral.

Palavras-chave: psicometria; neuropsicologia; validade; testes.

Abstract

Psychometrics parameters' studies in transcultural adaptation, validation, and construction processes of measuring instruments for clinical ends are important procedures to assure scientific accuracy and operational effectiveness in situations of examination in the neuropsychological field. OBJECTIVE: The objective of this article is to characterize the inter-relations between Psychometrics and Neuropsychology in the development of instrumental resources for investigating typical and atypical functional processes in the diagnosis and neuropsychological intervention processes. METHOD: Conduct a theoretical review of the professionals and students on the subject pertinent to psychological evaluation. RESULTS AND CONCLUSION: The technical knowledge concerning the quality of the measuring instruments, either in the field of research or professional intervention is an important factor in the search of primary and secondary scientific evidences to the this phenomenon. Thus, there are contributions from Psychometrics area to build evidence and improvement of measuring instruments in Neuropsychology and health in general.

Keywords: *psychometrics; neuropsychology; validity; tests.*

Introdução

A Psicometria é uma disciplina da ciência psicológica que tem como objetivo o estudo da mensuração de fenômenos psicológicos, de acordo com as condições que os definem (*behavior setting*) e demais variáveis não psicológicas.

Historicamente, o nascimento da Psicometria está diretamente relacionado o nascimento da ciência psicológica, no final do século XIX, e sua tentativa de dispor de um *corpus* de conhecimentos e de uma forma de abordagem de fenômenos psicológicos lastreada pela teoria da medida e os pressupostos do positivismo científico (Pasquali, 1997, 2003; Alchieri & Cruz, 2003). No início do século XX, tendo em vista a necessidade de desenvolver instrumentos de observação de processos subjetivos, conceitos tais como inteligência, personalidade, criatividade, emoção, percepção e memória passaram a ser enfaticamente estudados por meio de instrumentos objetivos, genericamente denominados de testes psicológicos.

Desde o início, pode-se observar que, do ponto de vista epistemológico e metodológico, a Psicometria, ou a medida de fenômenos psicológicos, assume o positivismo científico e a abordagem quantitativa como pressupostos, inserindo-se no âmbito das ciências que utilizam propriedades numéricas para o estudo das variações dos fenômenos humanos (Pasquali, 2003). Porém, esse caminho não se mostrou facilitado, como aponta Cruz (2002, p.17): “O acesso e a compreensão de fenômenos ou processos psicológicos, em quaisquer das áreas e objetos de intervenção em Psicologia, não se configuram uma tarefa fácil, pelo grau de complexidade e plurideterminação dos eventos psicológicos, mas é a condição pela qual se torna necessária uma ciência de tal magnitude conceitual”.

Durante e após a Segunda Guerra Mundial (1939-1945), especialmente nos Estados Unidos, verificou-se um avanço importante no campo da Psicometria, com o uso sistemático de testes psicológicos na avaliação de funções psicofisiológicas, desempenho psicomotor e traços de personalidade, época em que Alexander Romanovich Luria estudava, na então União Soviética, soldados com lesões cerebrais (Schlindwein-Zanini, 2015) e Hermann

Rorschach construía a técnica do psicodiagnóstico para avaliar distúrbios mentais (Cruz, 2002).

Destacou-se, também, a partir dessa época, o uso de testes objetivos no auxílio do diagnóstico de disfunções perceptivas e cognitivas, assim como na avaliação de progressos clínicos ao longo da intervenção neuropsicológica. Nessa primeira metade do século XX, destacam-se vários pesquisadores do campo da matemática e da fisiologia associando dados estatísticos à análise de funções psicofisiológicas (Pasquali, 2007), o que em muito contribuiu para o desenvolvimento futuro de instrumentos voltados ao campo neuropsicológico. Ao longo da segunda metade do século XX, com a sofisticação de técnicas estatísticas, a Psicometria revigora seus métodos de investigação de propriedades psicológicas, passando a contribuir não somente com a Psicologia, mas com diversas outras ciências na construção e adaptação de instrumento de medida, especialmente os instrumentos de autorrelato e de desempenho psicomotor (Alchieri & Cruz, 2003).

No atual cenário, parte-se do pressuposto, em Psicometria, de que pessoas diferem em suas capacidades, modos de funcionamento, repertórios cognitivos e emocionais, expressões comportamentais, dentre outros aspectos, e que essas diferenças podem ser identificadas, de forma mais ou menos compartilhadas, em grupos e populações específicas, tendo em vista as condições biopsicossociais que determinam ou contribuem para essa variabilidade.

Por sua vez, estudos acerca da avaliação neuropsicológica estão em desenvolvimento, auxiliando na construção, na validação e no aprimoramento de instrumentos. Os testes neuropsicológicos (selecionados pelo neuropsicólogo, conforme a indicação do exame) são importantes na detecção de possíveis déficits funcionais, porém, seus resultados não devem ser utilizados como única fonte de informação sobre o paciente (Schlindwein-Zanini et al, 2013).

No sentido prático, a Psicometria contribuiu na observação de processos subjetivos por meio da construção e adaptação de instrumentos de medida, verbais (oral e escrito) e não verbais, tais como escalas, questionários, inventários, protocolos de observação e registro, de diferentes formatos de estímulos ou itens (frases, desenhos, figuras, imagens). Assim, os estudos

psicométricos servem a três propósitos fundamentais e, em geral, complementares (Cozby, 2003; Alchieri & Cruz, 2004; Pasquali, 2010):

a) na aferição das variações de fenômenos psicológicos (observáveis e não observáveis), em contextos específicos e em diferentes sujeitos, por meio de representações conceituais (símbolos, categorias) ou numéricas (números relativos, percentuais);

b) na busca de evidências de validade e precisão de instrumentos de medida adaptados ou construídos para mensurar atributos específicos (atitude, atividade mental, estilos comportamentais, por exemplo), ou seja, em que medida o instrumento é útil e sensível para avaliar o que se propõe medir (validade) e quão consistentes são os resultados obtidos e suficientemente livres de erro para serem úteis à avaliação (precisão);

c) na padronização de instrumentos de medida ou procedimentos para medir as variações de fenômenos comportamentais, de forma a poder comparar, no caso concreto, os resultados obtidos em um indivíduo submetido ao instrumento, com os resultados e dados normativos de grupos ou populações pesquisadas (p.e., sexo, faixa etária, escolaridade, ocupação, tipo de agravo à saúde), que expressam a medida dessas variações, sejam elas similares ou discrepantes.

A Psicometria contemporânea possui como vertentes, a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), que contribuem no processo de validação e adaptação de instrumentos de medida.

Na TCT, os escores ou pontuações obtidos em estudos empíricos por meio de instrumentos de medida, estimam o grau ou nível do atributo (atitude, estilo de personalidade, interesse, comportamento específico) com base nas diferenças sistemáticas entre as respostas dos examinandos, que se supõem determinadas pela variação da variável de interesse na condição específica de exame, o que incluem outras fontes de variação devidas ao próprio instrumento e sua forma de aplicação ou, ainda, às condições internas dos examinandos (Primi, 2012). Por isso, a necessidade de técnicas de padronização do instrumento, a fim de controlar erros sistemáticos e permitir inferências estatísticas necessários a elaboração de normas populacionais. De qualquer forma, a limitação mais importante dos instrumentos de medida elaborados

segundo a TCT é de que não se pode separar as características dos examinandos das características de cada instrumento, ou seja, há uma relação de dependência na interpretação entre esses dois aspectos. Na prática, os escores obtidos pelos examinandos dependem das características da amostra da qual foram obtidos os escores que, por sua vez, dependem da seleção dos itens realizadas para a busca dos parâmetros psicométricos (Pasquali, 2010).

A TRI, por sua vez, tem por pretensão fornecer uma fundamentação probabilística a necessidade de mensurar construtos ou traços latentes, ou seja, fenômenos não observáveis. A TRI considera o item de um instrumento de medida uma unidade básica de mensuração e utiliza o padrão de resposta do item como um experimento (no sentido do controle do erro), que proporciona dados acerca da probabilidade de uma resposta em particular a um item com a variável de interesse (Hambleton, 1993).

A diferença principal entre a TCP e a TRI é de que, na primeira, a relação entre o valor esperado e o traço latente é de tipo linear, enquanto, na segunda, as relações são previstas como funções exponenciais.

Adaptação transcultural de instrumentos de medida

A adaptação de instrumentos psicológicos é um processo complexo que requer elevado rigor metodológico (Borsa, Damasio & Bandeira, 2012). A *International Test Commission* (ITC), desde 1992, tem proposto diretrizes para a tradução e a adaptação de instrumentos psicológicos entre culturas (ITC, 2010). Os termos "adaptação" e "tradução" são distintos, e tem-se preferido o uso do primeiro, uma vez que compreende todos os processos concernentes à adequação cultural do instrumento, para além da mera tradução (Hambleton, 2005). Por sua vez, a *American Psychological Association* (APA) estabeleceu os *Standards for Educational and Psychological Testing*, visando que os testes possuam os parâmetros de qualidade cientificamente determinados.

O processo de adaptação transcultural de um instrumento de medida estrangeiro é uma condição importante para o seu uso no Brasil, tendo em vista a necessidade de adaptar semanticamente os itens do instrumento para a cultura local, a fim de prosseguir no intento de obter evidências de validade e precisão relevantes à padronização da medida às normas da população-alvo. A obtenção

de normas brasileiras, com base em processo de adaptação de instrumentos não autóctones, possibilita, também, comparar valores normativos obtidos em diferentes populações, tendo em vista os parâmetros psicométricos obtidos (Camargo & Contel, 2004).

A tradução é, apenas, o primeiro passo do processo de adaptação. Ao se adaptar um instrumento, deverão ser considerados os aspectos culturais, idiomáticos, linguísticos e contextuais concernentes à sua tradução (Hambleton, 2005). Uma vez adaptado o instrumento, é possível realizar estudos com amostras populacionais comparando características de indivíduos inseridos em diferentes contextos culturais (Gjersing, Caplehorn, & Clausen, 2010; Hambleton, 2005). Atentando, também para limitações em estudos com sujeitos com alto nível de educação formal e indivíduos com sintomas psiquiátricos e neurológicos (Salgado et al, 2011).

O processo de adaptação de um instrumento de medida geralmente é concebido em protocolos de pesquisa. Inicia-se com o consentimento dos autores do instrumento original e segue um conjunto de procedimentos básicos (Dobansky e Santos (2000); Beaton e Guillemin (2000)):

a) Tradução da versão original para idioma português brasileiro, geralmente realizada por dois ou três especialistas na matéria e com domínio do idioma inglês. As versões traduzidas são comparadas e, com base na verificação de semelhanças e diferenças, consolida-se a versão em português brasileiro do questionário. Nesse processo procura-se utilizar alguns critérios, tais como: termos traduzidos igualmente nas versões são mantidos, desde que não prejudiquem o sentido da sentença original; termos ou expressões traduzidas diferentemente nas versões são comparados à escala original, mantendo-se o termo que mais se aproximava do termo original.

b) Realização de estudo piloto na população de interesse, conferindo com atenção dúvidas, impressões, sugestões e questionamentos oriundos do contato com os participantes, a fim de verificar a adequação da semântica da instrução e dos itens do instrumento ao nível de escolaridade, idade e valores culturais dos participantes do estudo. Esse procedimento permite verificar a qualidade e pertinência dos itens, permitindo realizar modificações no instrumento, quando necessárias, antes de sua fase de pesquisa extensiva;

c) Retrotradução (*back translation*), realizado por dois ou três profissionais, brasileiros e especialistas no idioma inglês que traduzem a versão em português brasileiro para o inglês. Esse procedimento busca verificar se as alterações realizadas na escala em português não prejudicaram o sentido das palavras e sentenças da escala original. As versões retrotraduzidas são comparadas com a escala original. Os termos traduzidos diferentemente da escala original são aceitos desde que apresentassem o mesmo significado.

d) Validade semântica e validade de face: tem por objetivo verificar se respondentes pertencentes à população-alvo compreendem plenamente as instruções, conteúdo dos itens, variação escalar e apresentação do instrumento de avaliação (Cozby, 2003). Esse procedimento segue o mesmo padrão do estudo piloto anteriormente referido. A investigação de evidências de validade semântica e de face procura verificar se são necessárias novas modificações associadas à linguagem, conteúdo e formato do instrumento, definindo-se a versão a ser administrada para coleta de dados do estudo.

Assim, sinteticamente, há vários passos na adaptação transcultural de instrumentos de medida, como a tradução do instrumento para o novo idioma, a síntese das versões traduzidas, avaliação da síntese por experts, avaliação pelo público-alvo, tradução reversa (*back-translation*), estudo-piloto, revisões, verificação de evidências de validade convergente e demais análises estatísticas pertinentes. Possíveis alterações que aconteçam ao longo dos estudos de validação devem ser discutidas à luz de aspectos quantitativos e qualitativos, com vistas a compreender as possíveis razões que levaram à alteração na estrutura fatorial do instrumento. Determinadas mudanças são esperadas devido a características amostrais, principalmente em instrumentos complexos, que apresentam um alto número de itens e de fatores.

O processo de construção de instrumentos de medida

Medir significa atribuir valores a características ou atributos de um objeto (construto), segundo regras que assegurem a validade e a confiabilidade dos resultados da medida (Braga & Cruz, 2006). Tornar operacionalmente verificável o que se quer medir exige a estruturação de um instrumento de

medida que busque capturar variações de fenômenos humanos, na maior parte das vezes não observáveis diretamente.

A construção de instrumentos de medida de natureza psicométrica basicamente pode sintetizada nos seguintes procedimentos: a) definição do construto e seus atributos, com base no estado da arte e características do contexto; b) construção de itens do instrumento relacionados ao construto e sua dimensionalidade; c) definição da escala de medida mais apropriada à captação da variação do que se quer mensurar; d) obtenção da validade de conteúdo da relação itens-construto ou atributos com especialistas na área de investigação; e) teste piloto do instrumento para verificar o conjunto semântico do instrumento em sua relação com as normas amostrais de interesse, com possibilidade de revisão dos itens e/ou aprimoramento das definições constitutivas e operacionais; e) análise de dados do teste piloto e busca de evidências preliminares de sensibilidade do instrumento; f) definição do instrumento final; g) busca de evidências de validade e confiabilidade do instrumento e sua possibilidade de padronização para normas amostrais específicas.

Realizar um processo de construção de um instrumento de medida implica em seguir vários procedimentos, tendo em vista a necessidade de assegurar o rigor científico e ético na definição dos construtos e suas propriedades, assim como no processo empírico e analítico de investigação, análise e interpretação dos dados obtidos (Hambleton, 2005). É necessário garantir, inicialmente, que o construto a ser mensurado (e seus atributos) sejam definidos teoricamente, constituindo-se em referências conceituais no plano da análise (definição constitutiva). As definições operacionais fazem com que o construto e seus atributos possam ser identificados funcionalmente (no sentido comportamental), o que se denomina de representação empírica, comportamental, dos traços latentes (construtos), conforme indica Pasquali (2003).

As definições constitutivas situam os limites semânticos, no campo teórico, no qual o instrumento deve ser baseado, ainda que determinados instrumentos possam não cobrir, com seus itens, toda a amplitude semântica de um conceito. Assim, boas definições constitutivas vão permitir avaliar a qualidade do instrumento, que mede o construto em termos do quanto de sua

extensão semântica é coberta pelo instrumento, surgindo daí instrumentos melhores e piores à medida que medem mais ou menos da extensão conceitual do construto, extensão essa delimitada pela definição constitutiva desse mesmo construto (Pasquali, 1998).

No processo de definição constitutiva podem ocorrer dúvidas se determinado construto ou atributo constitui uma unidade semântica única ou seria uma síntese de componentes distintos ou independentes, o que se denomina, em Psicometria, de dimensionalidade, ou seja, a composição da estrutura interna (semântica) do construto/atributos. Os fatores que compõem o construto (o atributo), obtidos por meio de análise fatorial, são produtos da dimensionalidade.

Uma vez definido o construto e seus atributos é importante identificar sob qual contexto o processo de mensuração incidirá. A escolha do *locus* e do grupo de observação e informação são relevantes à legitimação da observação das variações do que se quer investigar.

Parâmetros psicométricos: validade e confiabilidade de um instrumento de medida

Instrumentos de avaliação são cada vez mais utilizados como medidas de resultados primários ou secundários em intervenção clínica em neuropsicologia e participam cada vez como variáveis dependentes chave nas decisões que influenciam na atenção ao paciente e na orientação para o desenvolvimento de pesquisas (Hobart, Cano, Zajicek & Thompson, 2007). O uso de informações relatadas por pacientes ou registradas por especialistas, por meio de instrumentos específicos, permite, dentre outras funções, a aferição de comprometimentos neurológicos específicos, de limitações funcionais em sua vida diária, bem como percepção da condição clínica autorreferida pelo paciente (Hobart, 2003)

A adequação dessas decisões depende diretamente da qualidade científica dos instrumentos de avaliação, ou seja, dos parâmetros psicométricos aferidos em estudos especializados, geralmente identificados por meio da busca de evidências de validade e confiabilidade. Ambos são cruciais para a compreensão dos dados estatísticos resultantes de estudos empíricos e afetam a

capacidade para interpretar os resultados obtidos pelos instrumentos, em termos de mensuração de atributos teoricamente definidos (construtos).

Conceito de validade

O conceito de validade se constitui no parâmetro fundamental e indispensável da teoria da medida, aplicada nas diversas disciplinas científicas que tem por objetivo investigar fenômenos humanos e da natureza por meio de processos de mensuração de suas propriedades ou características fundamentais. Na pesquisa científica, a validade assume concepções ligadas aos resultados da pesquisa e à qualidade dos instrumentos de coleta de dados. A validade se refere à semelhança entre o conceito e suas medidas, ao grau em que uma medida representa precisamente o que se espera. Considerando-se um instrumento de medida, validade se refere a propriedade intrínseca do instrumento de medir aquilo que se propõe a medir (Pasquali, 2010; Ollaik & Ziller, 2012).

Atualmente, o conceito de validade é entendido como um processo, que se inicia com a definição detalhada do construto a ser mensurado, ou de sua dimensionalidade, prossegue com a observação e análise sistemáticas do domínio comportamental relevante e persiste até a validação dos escores e combinações interpretativas dos escores obtidos por meio de análises estatísticas em comparação com critérios externos, da vida real (Anastasi & Urbina, 2000).

Essencialmente, a pesquisa é válida se as evidências fornecem o apoio necessário às suas conclusões. Nessa perspectiva, a validade estaria relacionada com a coerência interna da pesquisa (Martins, 2004). Ou seja, há concepções que dão mais ênfase à validade dos resultados, também denominada validade externa, e há concepções que dão mais ênfase à validade do processo, do método, também denominada validade interna. As definições de validade nas pesquisas de abordagem quantitativa e qualitativa são distintas. Whittemore, Chase e Mandle (2001) ressaltam que isso ocorre porque os princípios epistemológicos e ontológicos são diferentes entre essas duas perspectivas de pesquisa, o que torna inapropriado o uso de medidas de validade de uma perspectiva em outra. Embora as diferenças sejam explícitas, isso não implica em afirmar que todos os enfoques de validade sejam incompatíveis entre as perspectivas qualitativas e

quantitativas de pesquisa, mas sim que uma simples tradução é inapropriada e inadequada (Ghunter, 2006).

Acerca da busca de evidências de validade baseadas nas relações com variáveis externas, observa-se que os estudos mais frequentemente desenvolvidos pelos neuropsicólogos são os de validade convergente ou discriminante. No caso da validade convergente, há a busca de um padrão de correlação entre os escores do teste-alvo e de outras variáveis que medem o mesmo construto ou construtos relacionados (por exemplo: desempenho em um teste de avaliação de síndrome disexecutiva *versus* desempenho em subtestes da escala de inteligência Wechsler para adultos (*Wechsler adult intelligence scale – WAIS III*) ou em um teste de fluência verbal. No entanto, nota-se que, muitas vezes, os testes escolhidos como padrão-ouro para as análises de correlação são testes de rastreio cognitivo, formados por poucos itens específicos de cada domínio e por fatores pouco sensíveis para identificação de algumas situações clínicas (por exemplo: MiniExame do Estado Mental). Também há de se considerar que, pela complexidade do funcionamento cognitivo, ainda existe certa dificuldade para estabelecer que magnitudes mínimas de correlação deveriam ser esperadas em cada avaliação para que se pudesse considerá-las, de fato, como evidências de validade (REPPOLD et al, 2015).

Investigar evidências de validade de instrumentos de medida é um processo pelo qual se coletam dados sobre o fenômeno ou variável que se pretende medir, com base na adequação e resultados do instrumento de medida aplicado ao contexto de estudo (Anastasi & Urbina, 1997; Cozby, 2003; Urbina, 2007). Há diferentes técnicas para aferição da validade de uma pesquisa científica, conforme a concepção de validade que esteja sendo utilizada. Para evidenciar a validade de instrumentos de medida, Pasquali (2003) apresenta três principais classificações:

a) Validade de conteúdo, que se refere à avaliação subjetiva dos itens e indicadores de uma escala ou instrumento, verificando, qualitativamente, se o conteúdo de tais é congruente à mensuração do que se propõem. Na análise dos itens, a compreensão de leitura do examinando ao responder o item, sua capacidade de avaliar um determinado atributo (comportamento) e a capacidade

dos itens de abarcarem a maior parte das manifestações comportamentais do fenômeno sob análise são investigados (Cruz, 2003).

b) Validade de critério, na qual se verifica o funcionamento das medidas em relação a outras variáveis denominadas critérios. A validade de critério de um instrumento é na verdade um teste de hipótese de que está medindo aquilo que se propõe a medir pela relação com um critério ou padrão externo de avaliação (Pasquali, 2003), isto é, a escolha de uma variável (condição clínica, por exemplo) ou um outro instrumento que possa manifestar uma associação estatística com as medidas originais examinadas: o grau de associação estatística assim determinado configurará o nível de validade empírica do instrumento de mensuração (Souza, 2005).

c) Validade de construto, que avalia o construto ou fenômeno que a escala ou instrumento está medindo. Para tanto, a validade de construto subdivide-se em validade convergente (extensão pela qual os indicadores relacionam-se entre si dentro de um mesmo construto), discriminante (extensão pela qual a escala difere-se de outras escalas) e nomológica (extensão pela qual a escala se relaciona com as definições teóricas do construto).

Na verificação das evidências de validade do instrumento, é usual a realização das análises de variância ANOVA e análises de Correlação de Pearson. O coeficiente de variação também é importante neste contexto, pois é uma medida de dispersão que descreve a quantidade de variabilidade relativa à média., analisa a dispersão em termos relativos. O coeficiente de variação é obtido pela razão entre o desvio-padrão e a média, e pode ser expresso em porcentagem. O Qui Quadrado (χ^2) é igualmente útil neste contexto, já que consiste em um teste não-paramétrico (isto é, não depende dos parâmetros populacionais, como variância e média) de hipóteses que visa encontrar um valor da dispersão para duas variáveis nominais, avaliando a associação entre variáveis qualitativas.

Uma ferramenta eficiente para identificação do índice mais adequado para o ponto de corte entre dois grupos é a relação entre a sensibilidade e a especificidade do instrumento. Os resultados obtidos por essa relação podem ser chamados de Curvas de Características de Operação do Receptor (Curvas ROC). A Curva ROC é uma ferramenta de medida de critérios diagnósticos por permitir

o estudo da variação entre a sensibilidade e a especificidade de um teste para identificação de diferentes características entre grupos discrepantes (Margotto, 2010). Na área da saúde, é comum que um desses grupos seja um grupo clínico e o outro um grupo de pessoas híidas.

Para Margotto (2010), muitos exames realizados no âmbito da investigação de processos clínicos utilizam escala numérica e, assim, a identificação da sensibilidade e a especificidade do instrumento dependem do ponto de corte (*cut off*) atribuído com base nos resultados da Curva ROC, em função dos resultados positivos e negativos relacionados ao objeto de investigação. São vários os instrumentos expostos a esses processos na área de Neuropsicologia, tais como como a Escala de Inteligência Wechsler para Adultos - WAIS (Wechsler, 2004), o Inventário de Alterações Neuropsicológicas para Adultos - NEUROPSZC (Schlindwein-Zanini e Cruz, 2013), escalas de avaliação funcional de portadores de doenças cerebrovasculares (NIH Stroke Scale, Glasgow Outcome Scale, Modified Rankin Scale, Barthel Index e Functional Independence Measure (Guimarães & Guimarães, 2004; Schlindwein-Zanini, 2016), dentre outros.

O conceito de confiabilidade

Confiabilidade, por sua vez, é um parâmetro psicométrico que implica em consistência e precisão. Falta de confiabilidade indica inconsistência e imprecisão, sendo que ambos são equiparados a erros de medição. No contexto dos instrumentos de medida, erros de medição pode ser definido como qualquer flutuação na pontuação, relacionados ao processo de medição, que são irrelevantes para o que está sendo medido (Pasquali, 1998; Hambleton, 2005).

A confiabilidade é uma condição necessária, mas não suficiente, para verificar a qualidade de escalas de avaliação. Portanto, confiabilidade indica uma qualidade de resultados de testes, que sugere que eles são suficientemente consistentes e livre de erros relevantes de medição para ser útil. Ou seja, se os resultados de um instrumento, uma vez replicados, mostram-se de forma consistente eles podem ser considerados confiáveis. Um coeficiente de correlação pode ser usado para avaliar o grau de fiabilidade. Se um instrumento é confiável ele deve mostrar uma alta correlação positiva. Assim, o instrumento

mais confiável é aquele em que um escore de um participante será aproximadamente o mesmo, ao longo do tempo ou entre diferentes avaliadores (no caso de escalas de autorrelato e clínico-administrado, por exemplo).

A confiabilidade de instrumentos pode ser investigada basicamente por meio de três métodos: *consistência interna*, *confiabilidade teste-reteste* e *concordância entre avaliadores*.

A *consistência interna* reflete o grau em que os itens de um questionário ou escala medem a propriedade ou dimensão teórica ao qual os itens se referem. Tipicamente os instrumentos que são utilizados clinicamente devem ter coeficientes de consistência interna (expressa em geral por meio do "coeficiente alfa") superior a 0,75. Instrumentos com pouco itens (10 ou menos), sem adequados estudos de calibração dos itens, tendem a apresentar baixos índices de consistência interna. Por isso, o cuidado que se deve ter em utilizar escalas ou questionários reduzidos sem estudos comparativos sólidos em diferentes normas populacionais.

A *confiabilidade teste-reteste* (*test-retest reliability*) verifica a estabilidade dos resultados de testes repetidos. Nesse caso, o período de tempo utilizados entre as repetições é uma variável importante na verificação de coeficientes de estabilidade, que são considerados relevantes com valores acima de 0,75. Um intervalo de tempo de 2 ou 3 semanas, por exemplo, pode ser adequado para ser aplicado um inventário de sintomas depressivos (condição de estado), enquanto um intervalo de 5-6 meses seria apropriado para o uso de uma medida de traço de personalidade (condição característica).

A *concordância entre avaliadores* é outro método importante da verificação da confiabilidade de um instrumento. Para as escalas clínicas é importante saber em que medida diferentes profissionais concordam/discordam ao atribuir pontuações em um exame específico. A variação da concordância entre avaliadores é geralmente expressa por um coeficiente Kappa (para resultados com duas ou três categorias) ou um coeficiente de correlação intra-classe (para desfechos contínuos). Coeficientes Kappa muitas vezes são ponderados para corrigir a oportunidade de acordo (Kappa ponderado). Embora não existam regras firmes para interpretar valores de Kappa, coeficiente de Kappa $\leq 0,40$ são geralmente considerados "pobres", valores entre 0,60 e 0,70

são considerados "bons" e coeficientes $> 0,70$ são considerados "ótimos". Os valores de Kappa podem ser afetados pelo número de opções avaliadores são obrigados a fazer (por exemplo, " presente ou ausente " versus " leve, moderada ou forte) ou pela taxa de prevalência da condição alvo.

A avaliação de confiabilidade (*assessing reability*) pode ser de dois tipos, interna e externa, realizadas de forma complementar ou não.

A *confiabilidade interna* pressupõe que o instrumento ou procedimento será avaliado com base nos resultados obtidos por ele mesmo, tendo em vista seus propósitos. O método de separação das metades (*split-half*) avalia a consistência interna de um instrumento por meio da comparação dos resultados de uma metade de um teste com os resultados a partir da outra metade. Um teste, por exemplo, pode ser dividido ao meio em várias formas: ser aplicada uma parte em dois períodos distintos ou pela divisão entre itens pares e ímpares. Se as duas metades do teste fornecem resultados semelhantes indica confiabilidade interna. Por outro lado, resultados de itens, em metades separadas de um instrumento, que apresentam baixa correlação ($r < 0,3$), sugerem inconsistência da medida e a necessidade de revisão dos itens e sua dimensionalidade. O método da separação das metades é uma maneira rápida e fácil para estabelecer a confiabilidade, embora tenda a ser mais eficaz com questionários e escalas com muitos itens que medem o mesmo construto. Em geral, é menos recomendado para testes que medem diferentes construtos (subescalas, por exemplo).

A *confiabilidade externa* avalia o instrumento com base em uma referência comparativa e é realizada segundo dois procedimentos:

O procedimento teste-reteste avalia a consistência externa de um instrumento. Uma avaliação típica envolveria fazer eu os examinandos respondam ao mesmo teste em duas ocasiões distintas. Por exemplo, estudar a resposta de pacientes ambulatoriais em um instrumento de rastreio de sintomas em duas etapas. Se for encontrado alto índice de correlação em ambos, ao comparar os resultados nas duas etapas, há evidências de validade externa. Assim, o procedimento teste-reteste mensura, na verdade, a estabilidade dos resultados de um teste ao longo do tempo, o que pode ser considerada uma restrição ou limitação ao seu emprego.

O procedimento de verificação de *confiabilidade entre avaliadores* refere-se ao grau em que diferentes profissionais atribuem estimativas consistentes ao mesmo comportamento, com base na resposta a um instrumento específico. A confiabilidade entre avaliadores pode ser um procedimento útil para calibrar entrevistas, por exemplo.

Estatisticamente, para realizar análise da confiabilidade de instrumentos de medida é possível utilizar a correlação de Pearson (entre teste e reteste), a correlação de Spearman-Brown para o método das metades e o Alfa de Cronbach para a verificação da consistência interna, quantificando-o em uma escala entre zero e um, sendo que o valor mínimo aceitável para considerar que um questionário seja confiável é, em geral, de 0,7 ou superior.

Uma vez contempladas as condições psicométricas do instrumento (como a análise dos itens, compreensão de leitura, elaboração e análise dos itens, verificação da validade/confiabilidade, normatização, padronização, classificação dos resultados respondidos pelos participantes); então o instrumento pode ser considerado um instrumento efetivo e útil à comunidade de profissionais, de forma que pode ser utilizado como critério primário ou auxiliar nos exames pretendidos.

Conclusões

A Neuropsicologia investiga a relação entre comportamento e cérebro, e sua decorrente atividade mental, interagindo com a Psicometria, que por sua vez, contribui na observação de processos subjetivos por meio da construção e adaptação de instrumentos de medida.

Inegavelmente, há contribuições da área de Psicometria à construção de evidências e aprimoramento de instrumentos de medida em Neuropsicologia e na área de saúde em geral. Nesse sentido, é válido salientar que o uso de instrumentos de medida constitui um aspecto importante no processo de investigação clínica, diagnóstico e controle da qualidade da intervenção de profissionais de saúde.

Seja qual for o processo de investigação com uso de instrumentos de medida, é consenso na literatura que a verificação dos parâmetros psicométricos de validade e confiabilidade é um empreendimento relevante para o controle

científico rigoroso da qualidade técnicas dos instrumentos construídos e adaptados à realidade brasileira. Assim, um instrumento de medida válido e confiável (testes, questionários, escalas, inventários, protocolos de registro) deve apresentar boas condições psicométricas, tais como, evidências satisfatórias de validade e confiabilidade, consistência interna, capacidade analítica de descrever e produzir inferências sobre as variações do fenômeno investigado e suas implicações no contexto investigado (características amostrais) e possibilidade efetiva de padronização para uso por profissionais em situações de trabalho.

Referências

- ALCHIERI, J. C. & CRUZ, R. M. *Avaliação psicológica: conceito, métodos e instrumentos*. São Paulo: Casa do Psicólogo, 2003.
- ANASTASI, A., & URBINA, S. (2000). *Testagem Psicológica*. 7. ed. Porto Alegre: Artes Médicas.
- BORSA, J. C., DAMASIO, B. F. & BANDEIRA, D. R. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: algumas considerações. *Paidéia*, Ribeirão Preto, v. 22, n. 53, p. 423-432.
- BRAGA, C. G., & CRUZ, D. A. L. M. (2006). Contribuições da psicometria para a avaliação de respostas psicossociais na enfermagem. *Rev Esc Enferm US*, 40(1), 98-104.
- CAMARGO, I. DE B., & CONTEL, J. O. B. (2004). Tradução e adaptação de questionários norte-americanos para a avaliação de habilidades e conhecimentos na prática psiquiátrica brasileira. *Revista de Psiquiatria do Rio Grande do Sul*, 26(3), p. 288-299.
- CRUZ, R. M. (2002). O processo de conhecer em avaliação psicológica. In R. M. Cruz, J. C. Alchieri & J. J. Sarda Junior (Orgs.). *Avaliação e medidas psicológicas: produção do conhecimento e da intervenção profissional*. São Paulo: Casa do Psicólogo. (pp. 15-24).
- GUIMARÃES, R. B. & GUIMARÃES, R. B. (2004). Validação e adaptação cultural para a língua portuguesa de escalas de avaliação funcional em doenças cerebrovasculares: uma tentativa de padronização e melhora da qualidade de vida *Rev. bras. neurol*, 40(3): 5-13, jul.-set.

- GUIMARÃES, S. E. R. & BZUNECK, J. A. (2008). Propriedades psicométricas de um instrumento para avaliação da motivação de universitários. *Ciências & Cognição*, 13(1), 101-113.
- GÜNTHER, H. (2006). Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão? *Psicologia: Teoria e Pesquisa*, Brasília, v. 22, n. 2, p. 201-210, maio/ago.
- HAMBLETON, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, 9(1), 57-68.
- HAMBLETON, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10(3), 229-244.
- HAMBLETON, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum. pp. 3-38.
- HAMBLETON, R. K., & PATSULA, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45(1-3), 153-171.
- HOBART, J. C. (2003). Rating scales for neurologists. *J Neurol Neurosurg Psychiatry*, 74 (Suppl. IV): iv22-iv26.
- HOBART, J., CANO, S., ZAJICEK, J., & THOMPSON, A. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations, *Lancet Neurol*, 6: 1094–1105.
- MARGOTTO, P. R. (2010). Curva ROC: como fazer e interpretar no SPSS. Disponível em http://www.paulomargotto.com.br/documentos/Curva_ROC_SPSS.pdf> Acesso em 18/03/2016.
- MARTINS, H. T. de S. (2004). Metodologia de pesquisa. *Educação e Pesquisa*, São Paulo, v. 30, n. 2, maio/ago. p. 289- 300.
- MINAYO, M. C. S. (1994). *O desafio do conhecimento: pesquisa qualitativa em saúde*. (3ª ed.) São Paulo: Hucitec-Abrasco.

- OLLAIK, L. G. & Ziller, H. M. (2012). Concepções de validade em pesquisas qualitativas. *Educação e Pesquisa*, São Paulo, v. 38, n.1, p. 229-241.
- PASQUALI, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Ed. UnB.
- PASQUALI, L. (1998). Princípios de elaboração de escalas psicológicas. *Rev. Psiq. Clin.*, 25(5), 206-213.
- PASQUALI, L. (2003). *Psicometria: teoria dos testes na Psicologia e na Educação*. Petrópolis: Vozes.
- PASQUALI, L. (2007). Validade dos testes psicológicos: será possível reencontrar o caminho? *Psicologia: Teoria e Pesquisa*. Brasília, v. 23, n. esp., Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-37722007000500019&lng=pt&nrm=iso. doi: 10.1590/S0102-37722007000500019.
- PASQUALI, L. (ed.). *Instrumentação psicológica: fundamentos e práticas*. Porto Alegre: Artmed, 2010.
- PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. 2. ed. Petrópolis: Vozes, 2003.
- PRIMI, R. (2012). Psicometria: fundamentos matemáticos da Teoria Clássica dos Testes. *Aval. Psicol.*, Itatiba, v. 11, n. 2.
- REPPOLD, C. Tozzi et al. (2015) Contribuições da psicometria para os estudos em neuropsicologia cognitiva. *Psicol. teor. prat.*, São Paulo, v. 17, n. 2, p. 94-106, ago. Disponível em <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872015000200008&lng=pt&nrm=iso>. Acesso em 16 out. 2016.
- SALGADO, J.V. et al. (2011). Applicability of the Rey Auditory-Verbal Learning Test to an adult sample in Brazil. *Rev. Bras. Psiquiatr.*, São Paulo, v. 33, n. 3, p. 234-237, set. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-44462011000300005&lng=pt&nrm=iso>. Acesso em 17 dez. 2017.
- SCHLINDWEIN-ZANINI, R. (2015). Programa de reabilitação neuropsicológica do Núcleo de Neuropsicologia do Hospital Univeristário/Universidade federal de Santa Catarina (UFSC): aspectos teóricos. *Anais Congresso Sul Brasileiro de Cognição*. Florianópolis/SC.

Disponível em <<http://www.hu.ufsc.br/setores/neuropsicologia/wp-content/uploads/sites/25/2015/02/Anais-neurosono.pdf>> Acesso em 15out2016

SCHLINDWEIN-ZANINI, R. (2016). Neuropsicologia Forense. In Peres, C. & Schlindwein-Zanini, R. *Neuropsicologia em ação: Entendendo a prática*. Rio de Janeiro: Wak, p. 107 – 126.

SCHLINDWEIN-ZANINI, R. et al (2013). Avaliação neuropsicológica e deficiências físicas: revisão de instrumentos viáveis no Brasil. *Contextos Clínicos*, v. 6, n. 1, jan-jun. Disponível em <<http://pepsic.bvsalud.org/pdf/cclin/v6n1/v6n1a05.pdf>>. Acesso em 15set2016.

WHITTEMORE, R., CHASE, S. K., & MANDLE, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, v. 11, n. 4, p. 522-537.