

Discutindo procedimentos psicométricos a partir da análise de um instrumento de avaliação neuropsicológica breve¹

Josiane Pawlowski – Universidade Federal do Rio Grande do Sul²
 Clarissa Marcell Trentini – Universidade Federal do Rio Grande do Sul
 Denise Ruschel Bandeira – Universidade Federal do Rio Grande do Sul

Resumo

A desatualização e a falta de testes psicológicos adequados para uso no Brasil apontam a necessidade de pesquisas sobre validação, fidedignidade, padronização e normatização dos testes. Com relação à avaliação neuropsicológica, existem poucos instrumentos nacionais para avaliar os prejuízos das funções cognitivas causados por lesões/disfunções cerebrais. O objetivo deste artigo é discutir teoricamente o estudo psicométrico de testes neuropsicológicos e exemplificá-lo através das etapas de validação do Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN, desenvolvido para avaliar algumas funções cognitivas. Especificamente, serão enfatizadas as dificuldades encontradas no desenvolvimento de métodos psicométricos para a validação do NEUPSILIN e sugeridas soluções para essas dificuldades.

Palavras-chave: Validação; Avaliação neuropsicológica; Métodos psicométricos.

Discussing psychometric procedures through the analysis of a brief neuropsychological assessment instrument

Abstract

There are still some outdated and unstandardized psychological tests in Brazil. This evidence points to the lack of researches about validity, reliability and normative data of psychological tests to be used with Brazilian people. Besides, there are few instruments to assess damaged cognitive functions caused by brain injuries. The purpose of this paper is to discuss the psychometric study of neuropsychological tests theoretically and to give the example of NEUPSILIN, developed to assess some cognitive functions briefly. More specifically, we will point the difficulties of developing psychometric methods to NEUPSILIN validation, as well as suggest solutions to these difficulties.

Keywords: Validation; Neuropsychological assessment; Psychometric methods.

Introdução

Nos últimos anos têm crescido o interesse dos pesquisadores pelas áreas de avaliação psicológica e neuropsicológica. A primeira retomou a atenção da comunidade científica para a construção de instrumentos de avaliação principalmente após as discussões sobre a desatualização e a falta de instrumentos com qualidades psicométricas adequadas para uso no Brasil. Nesse sentido, o sistema de avaliação dos testes psicológicos, criado através da Resolução nº 2/2003 do Conselho Federal de Psicologia (CFP, 2003), apontou a necessidade de maior oferta de instrumentos válidos, fidedignos, padronizados e normatizados para serem utilizados com a população brasileira. Com relação à área da neuropsicologia, o crescimento pode ser evidenciado pela sua recente regulamentação, através da Resolução CFP nº 2/2004, como uma especialidade da Psicologia (CFP, 2004).

Entretanto, essa área mostra carência no que se refere a instrumentos para avaliação dos prejuízos das funções cognitivas em decorrência de lesões/disfunções cerebrais. Considerando a importância de ambas áreas e a carência de instrumentos para avaliação neuropsicológica, o presente artigo objetiva evidenciar algumas das dificuldades no estudo psicométrico de instrumentos de avaliação neuropsicológica, exemplificado a partir do processo de validação do Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN (Fonseca, Salles & Parente, no prelo). Também serão sugeridas algumas propostas de solução diante das dificuldades encontradas naquele processo.

Os instrumentos de avaliação neuropsicológica e sua carência no Brasil

Os primeiros estudos sobre baterias de testes

¹ Os autores agradecem ao Conselho Nacional de Pesquisa Científica – CNPq pelo apoio e financiamento à pesquisa.

² Endereço para correspondência:

Universidade Federal do Rio Grande do Sul – Instituto de Psicologia
 Rua Ramiro Barcelos, 2.600 – sala 120 – 90035-003 – Porto Alegre-RS – Telefone: (51) 3308-5352
 E-mail: josipski@yahoo.com.br

neuropsicológicos, de acordo com Peña-Casanova (1985), foram publicados na obra de Weinsenburg e McBride em 1935, ao investigarem as afasias. Nessa mesma época, foi desenvolvida a bateria de Halstead-Reitan em duas etapas, primeiro por Ward Halstead e, na década de 50, por Halph Reitan. Segundo Kane (1991), entre as principais vantagens das baterias fixas são destacadas: (1) a possibilidade de descobrir déficits não esperados em função do detalhamento de um maior número de funções cognitivas; (2) o desenvolvimento de uma ampla base de dados normativos em razão de serem aplicados os mesmos testes a todos os pacientes; e (3) a aprendizagem mais rápida da aplicação por estudantes em treinamento.

Atualmente, dentre as baterias internacionais mais utilizadas estão *Halstead-Reitan Neuropsychological Test*, *Luria Nebraska Neuropsychological Battery* e *Iowa-Benton Neuropsychological Assessment* (Tupper, 2000). A limitação dessas baterias consiste no tempo prolongado de aplicação e levantamento, o que torna seu uso inviável em algumas situações clínicas, além da necessidade de um treino especializado para sua aplicação (Ostrosky-Solís, Ardila & Rosselli, 1999). Em consequência, instrumentos breves têm sido desenvolvidos, tais como o *Consortium to Establish a Registry for Alzheimer's Disease – CERAD* (Morris & cols., 1989) e o *NEUROPSI: Evaluación Neuropsicológica Breve en Español* (Ostrosky-Solís & cols., 1999). Recentemente, novos estudos de normatização da bateria CERAD foram conduzidos por Beeri e colaboradores (2006), considerando as variáveis idade, sexo e escolaridade.

Dentre os instrumentos de avaliação das funções cognitivas de forma breve utilizados no Brasil encontra-se o Mini-Exame do Estado Mental (MEEM) (Folstein, Folstein & McHugh, 1975, publicado no Brasil por Bertolucci, Brucki, Campacci & Juliano, 1994), cuja finalidade é rastrear a presença de comprometimento cognitivo, principalmente em populações com um possível quadro de demência. Os estudos nacionais com o MEEM apresentam algumas sugestões de modificação no mesmo a fim de uniformizar o uso com a população brasileira (Almeida, 1998; Brucki, Nitrini, Caramelli, Bertolucci & Okamoto, 2003). Além desse instrumento, Bertolucci e colaboradores (2001) estudaram a aplicabilidade da bateria neuropsicológica CERAD em idosos brasileiros. Ainda, visando ao diagnóstico diferencial entre demência leve e depressão, Chaves e Izquierdo (1992) organizaram uma bateria que inclui tarefas breves para avaliar prejuízos nas funções de memória, linguagem, abstração, cálculo, julgamento, bem como avaliar apraxias e agnosias. Essa bateria também foi utilizada por Ferreira (1997) em uma pesquisa sobre a avaliação de pacientes com demências, tendo sido denominada como Critérios de Positividade para Déficit Cognitivo. Contudo, é importante ressaltar que nenhum desses instrumentos destina-se especi-

ficamente à avaliação de pacientes com lesões cerebrais focais.

Evidencia-se, assim, que a elaboração de instrumentos de avaliação neuropsicológica tem crescido nos últimos anos, mas essa área ainda é pouco desenvolvida no Brasil, onde os testes neuropsicológicos apresentam limitações de uso principalmente em virtude da falta de dados normativos e estudos com grupos clínicos (Alchieri, 2004; Andrade, 2002; Bandeira, Serafini, Fonseca, Salles & Parente, 2005; Capovilla, Joly & Tonelotto, 2006; Mäder, 2002; Simões, 2002). Além disso, a área carece de testes que não demandem tempo prolongado de aplicação, que sejam sensíveis a danos cerebrais leves ou que apontem prejuízos cognitivos específicos. A fim de suprir essa lacuna, foi elaborado no Brasil, por Fonseca e colaboradores (no prelo), o Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN. O objetivo desse teste é fornecer um perfil neuropsicológico breve mediante a avaliação de áreas de competência e déficits nas funções cognitivas: orientação temporoespacial, atenção, percepção, memória, cálculias, linguagem, praxias e função executiva (resolução de problemas e fluência verbal), as quais são avaliadas por meio de 32 tarefas. A pontuação final do NEUPSILIN ainda está sendo avaliada, pois depende da coleta de dados normativos e da ponderação de alguns escores. As habilidades, as tarefas e os escores brutos foram distribuídos na construção do instrumento, como segue:

I) Orientação temporoespacial – Tempo: resposta de dia da semana, dia do mês, mês e ano, e Espaço: resposta de local, cidade, estado e país (escore máximo = 8 pontos);

II) Atenção – 1) Contagem inversa: contagem de 50 a 30 pelo examinando e contagem do tempo de realização da tarefa pelo examinador para avaliar a velocidade de processamento (escore máximo = 20 pontos) e 2) Repetição de seqüência de dígitos: repetição de uma seqüência de sete números (escore máximo = 7 pontos);

III) Percepção – 1) Verificação de igualdade e diferença de linhas: verificação de seis pares de linhas (escore máximo = 6 pontos); 2) Heminégligência visual: percepção de todo espaço de uma folha de papel repleta de traços que deverão ser riscados (escore = 1 ponto); 3) Percepção de faces: avaliação, como iguais ou diferentes, de três pares de fotografias de faces, sendo uma delas de frente e outra de perfil (escore máximo = 3 pontos); e 4) Reconhecimento de faces: memorização de dois rostos desenhados, que devem ser reconhecidos entre quatro posteriormente (escore máximo = 2 pontos);

IV) Memória – 1) Memória de trabalho: a) Ordenamento ascendente de dígitos: repetição em ordem crescente de 10 conjuntos de dois até seis dígitos (escore máximo = 10 pontos); b) Span auditivo de palavras em sentenças:

memorização e evocação das últimas palavras de frases, após a leitura de conjuntos de dois, três, quatro e cinco sentenças pelo examinador (score máximo = 14 pontos); 2) Memória verbal: a) Evocação imediata: repetição de nove palavras ditas pelo examinador (score máximo = 9 pontos), b) Evocação tardia: emissão das mesmas nove palavras em tempo posterior (score máximo = 9 pontos), e c) Reconhecimento: indicação, em uma lista de 18 palavras, da presença ou não da palavra na lista original falada pelo examinador (score máximo = 18 pontos); 3) Memória semântica de longo prazo: resposta a duas perguntas referentes a conhecimentos gerais (score máximo = 5 pontos); 4) Memória visual de curto prazo (três estímulos): memorização de uma figura sem sentido de cada vez e reconhecimento entre um conjunto de três figuras semelhantes (score máximo = 3 pontos); 5) Memória prospectiva: lembrança, ao final da testagem, da instrução de escrever o nome em uma folha de papel fornecida no início da avaliação (score máximo = 2 pontos);

V) Habilidades aritméticas – resolução de quatro cálculos, um de cada operação aritmética básica (score máximo = 8 pontos);

VI) Linguagem – 1) Linguagem oral: a) Nomeação de dois objetos e duas figuras (score máximo = 4 pontos); b) Repetição de oito palavras reais e duas pseudopalavras (score máximo = 10 pontos); c) Linguagem automática: contagem de um a dez e verbalização de todos os meses do ano, em ordem (2 pontos); d) Compreensão oral (três conjuntos de figuras): indicação da figura correspondente ao enunciado verbal do examinador (score máximo = 3 pontos); e) Processamento de inferências: explicação do significado de um provérbio e de duas metáforas (score máximo = 3 pontos); 2) Linguagem escrita: a) Leitura em voz alta de dez palavras reais e de duas pseudopalavras (score máximo = 12 pontos); b) Compreensão escrita (três estímulos): leitura em silêncio de palavras e frases e indicação das figuras correspondentes (score máximo = 3 pontos); c) Escrita espontânea de uma frase (score = 2 pontos); d) Escrita copiada de uma frase (score = 2 pontos) e e) Escrita ditada de dez palavras reais e de duas pseudopalavras (score máximo = 12 pontos);

VII) Praxias – a) Ideomotora: realização de três gestos, conforme instrução verbal do examinador (score máximo = 3 pontos); b) Construtiva: cópia de três figuras (quadrado, flor e cubo) e desenho de um relógio (score máximo = 16 pontos); c) Reflexiva: repetição de uma seqüência de três gestos (score máximo = 3 pontos);

VIII) Resolução de problemas – resposta a duas perguntas envolvendo raciocínio abstrato (score máximo = 2 pontos);

IX) Função executiva (fluência verbal) – verbalização, durante um minuto, de palavras que iniciem com a letra F. Score: contagem do número de itens verbalizados.

Como pode ser observado por essa descrição, as tarefas do NEUPSILIN são curtas e de fácil resolução por pessoas que possuem preservadas as funções cognitivas avaliadas pelo teste. O NEUPSILIN pode ser classificado como um instrumento de avaliação neuropsicológica breve, porque possui um tempo reduzido de aplicação (entre 30 e 40 minutos) e inclui tarefas para acessar várias funções cognitivas.

As etapas de construção de um instrumento de avaliação psicológica

A construção de um instrumento de qualidade requer etapas bem definidas e procedimentos rigorosos. De acordo com Pasquali (1999), existem alguns passos necessários, que incluem tarefas e métodos específicos, os quais devem ser realizados em uma seqüência temporal determinada. Esses passos compõem três eixos distintos, denominados por Pasquali procedimentos teóricos, empíricos ou experimentais, e analíticos ou estatísticos.

Os procedimentos teóricos incluem a escolha do sistema a ser estudado, a definição de suas propriedades (atributos), a concepção da dimensionalidade desses atributos, a definição constitutiva e operacional dos mesmos, a construção de itens e a validação de conteúdo. Essa etapa tem como primeiro objetivo a especificação das categorias comportamentais que representam o objeto psicológico a ser medido, tornando-se necessária a explicitação dos fundamentos teóricos que orientam a criação dos itens para que seja garantida a legitimidade da medida (Pasquali, 1999). Dessa forma, a investigação teórica deve preceder a construção dos itens, o que também fornece uma direção às etapas seguintes de construção. Como o sistema psicológico escolhido não pode ser medido diretamente, são delimitados os atributos ou propriedades que o definem, cuja escolha depende de fatores tais como o interesse do pesquisador, a saturação de estudos prévios e a relevância do atributo para o instrumento. Desses atributos são derivadas dimensões dependentes tanto de argumentos teóricos quanto de análises fatoriais realizadas em estudos exploratórios posteriores, a fim de investigar se a teoria que fundamenta a construção do teste é ou não pertinente. Após definidas as dimensões, os construtos são descritos com base em conceitos da teoria e devem ser transformados em itens mensuráveis, ou seja, devem ser definidos operacionalmente. Construídos os itens, estes devem ser submetidos à análise de juízes (peritos na área do construto) e à análise semântica que objetiva verificar se os mesmos podem ser

compreendidos e se apresentam validade aparente (credibilidade), etapa que compreende o estudo de validade de conteúdo do instrumento.

Após a verificação da validade de conteúdo do teste, segue-se a realização dos procedimentos empíricos e analíticos. Os primeiros incluem a definição de amostras e de instruções do teste, a administração do instrumento piloto e a coleta válida para proceder à verificação de suas qualidades psicométricas. Já os procedimentos estatísticos (analíticos) são a última etapa da construção de um instrumento, abrangendo análises estatísticas a serem efetuadas para validação, precisão e normatização do teste. Nessa etapa, de acordo com Pasquali (2003), são realizadas as análises da validade de construto e de critério do instrumento.

No que se refere à construção do NEUPSILIN, esse instrumento já foi submetido à análise de validade de conteúdo, incluindo as análises de juízes e semântica (Fonseca, Salles & Parente, 2006). Além disso, já foi realizado um estudo piloto com a última versão do instrumento (Pawlowski & colaboradores, 2006). Tais análises indicaram a adequação do teste quanto à validade de conteúdo e à compreensão dos itens pelos participantes. Entendendo-se a necessidade de prosseguir as análises das qualidades psicométricas do NEUPSILIN, o instrumento ainda será submetido a estudos de outras evidências de validade e investigações de fidedignidade, padronização e normatização.

Dificuldades e possibilidades de análises de validade do NEUPSILIN

Após o estudo de validade de conteúdo do NEUPSILIN, as etapas seguintes de análise de suas qualidades psicométricas devem incluir os estudos de validade de construto e de critério. A validade de construto verifica se o teste constitui uma representação adequada do construto teórico ou traço e a validade de critério avalia a eficácia do teste em prever o desempenho de um grupo específico de indivíduos, sendo então selecionados grupos-critério referentes aos construtos avaliados pelo instrumento (Anastasi & Urbina, 2000; Pasquali, 2001). Enquanto a análise de validade de conteúdo do NEUPSILIN seguiu as propostas de execução apresentadas na sessão anterior, investigações de validade de construto desse instrumento são mais restritas. Isso ocorre em virtude das particularidades referentes à forma do teste, que inclui tarefas breves para avaliar várias funções cognitivas, e da maneira como é realizado o seu levantamento, ou seja, como o teste é pontuado. Dessa forma, mesmo existindo métodos distintos para analisar a validade de construto de um instrumento (Pasquali, 1999, 2001, 2003), algumas restrições são verificadas no que se

refere ao estudo psicométrico do NEUPSILIN, por exemplo, nos seguintes procedimentos:

1) *Diferenciação por idade*: é um procedimento utilizado quando o teste avalia características de comportamento que aumentam ou diminuem com a idade, por exemplo, a inteligência e o desenvolvimento psicomotor. Em um estudo piloto realizado por Santos (2005) com o NEUPSILIN, poucos subtestes apresentaram diferença significativa na comparação dos escores entre adultos jovens e uma população de 59 a 75 anos. Assim, o desempenho no NEUPSILIN tende a permanecer estável entre adultos jovens, podendo ser encontrada tendência de declínio no desempenho em indivíduos saudáveis em consequência do envelhecimento, o que também foi evidenciado no estudo de Ostrosky-Solís & colaboradores (1999) com a bateria espanhola NEUROPSI. É possível que a partir da idade de 75 anos seja evidenciado um declínio mais representativo nas tarefas do NEUPSILIN, devendo, para essa análise, ser controlados fatores como a escolaridade e hábitos de leitura e escrita, que podem influenciar na manutenção do desempenho cognitivo (Fonseca, 2006). Portanto, a diferenciação por idade poderia contribuir no estudo de validade do NEUPSILIN mediante a comparação de suas faixas extremas;

2) *Correlações com testes que medem o mesmo construto*: consiste na aplicação de outro instrumento que meça o mesmo construto daquele que se pretende validar. Na validação do NEUPSILIN, esse procedimento é dificultado em razão da escassez de instrumentos para avaliação de funções cognitivas que meçam conjuntamente os mesmos construtos avaliados por esse teste. De um modo geral, os autores incluem em uma bateria subtestes que avaliam diferentes funções cognitivas, de acordo com seu interesse clínico ou de pesquisa. Uma possibilidade para resolver essa questão seria administrar testes ou tarefas de instrumentos já existentes que meçam o mesmo construto de cada razão incluída no NEUPSILIN. Para não ser necessário um tempo de aplicação muito longo, o que poderia gerar efeito de redução do desempenho em razão do cansaço, poderiam ser aplicados apenas alguns subtestes do NEUPSILIN e seus testes correlatos, sendo formados grupos para a análise de todos os subtestes. Também é possível a aplicação de testes em dias diferentes e ordens diversas para controlar o efeito de aprendizagem em tarefas semelhantes, o que poderia enviesar o processo de validação;

3) *Análise convergente-discriminante*: nessa análise é necessário que o instrumento novo se correlacione significativamente a variáveis com as quais teoricamente deve estar

relacionado (validade convergente) e que não se correlacione a variáveis com as quais teoricamente deve divergir (validade discriminante) (Anastasi & Urbina, 2000; Pasquali, 2003). Considerando os diversos construtos avaliados pelo NEUPSILIN, poderia ser realizada uma análise convergente entre esse instrumento e a inteligência ou raciocínio. Já para a análise discriminante, é difícil eger algum teste que não se correlacione a qualquer das funções avaliadas pelo NEUPSILIN. É importante ressaltar, ainda, que esse tipo de procedimento isolado dos demais fornece pouca evidência de validade do novo teste;

4) *Intervenções experimentais*: objetivam verificar se o teste discrimina claramente grupos-critério produzidos experimentalmente em relação ao traço a ser medido, por exemplo, diferenciar ansiosos de não-ansiosos através de uma escala, devendo ter sido criada uma situação potencialmente geradora de ansiedade à qual um dos grupos seria submetido (Pasquali, 2001). Quanto às análises com o NEUPSILIN, esse é um procedimento de difícil delineamento metodológico em razão das questões éticas envolvidas em uma possível intervenção experimental a fim de manipular o funcionamento cognitivo de um paciente. Um procedimento nessa área poderia ser considerado demasiadamente invasivo, além de ser necessária a realização de uma pesquisa interdisciplinar pela necessidade da participação de especialistas para a realização dos procedimentos clínicos com os pacientes;

5) *Análise fatorial*: consiste na análise da dimensionalidade do instrumento, que tem a finalidade de verificar o número de traços latentes (definidos teoricamente) que o teste avalia. Nesse sentido, um instrumento tem validade de construto quando há demonstração empírica da hipótese teórica que o fundamenta (Pasquali, 2001). A análise fatorial pode ser aplicada ao NEUPSILIN desde que se tenha uma amostra grande, com variabilidade de desempenho. Além disso, um mesmo item pode revelar-se com peso em mais de um fator, ou itens de diferentes subtestes podem carregar no mesmo fator. Essas hipóteses de configuração dos itens são geradas em razão da alta correlação entre alguns dos subtestes e funções avaliadas pelo instrumento, os quais podem ser interdependentes, tais como a memória e a atenção;

6) *Análise da consistência interna*: consiste em calcular a correlação entre cada item do teste e o restante dos itens ou o escore total dos itens. O problema dessa análise em um teste neuropsicológico breve tal como o NEUPSILIN consiste na composição do instrumento com itens que avaliam diferentes construtos, o que torna inadequado o uso da análise de consistência interna do instrumento como um todo. Da mesma forma, essa análise não se aplica aos subtestes, visto que muitos

escores apresentam pouca variabilidade, vários subtestes são compostos por um número reduzido de itens e alguns possuem itens cujo grau de dificuldade aumenta, o que não permite a realização da análise pelo α de Cronbach. Uma alternativa seria realizar análises de correlação simples entre os itens dos subtestes e o total de cada função cognitiva avaliada. Com relação a esse procedimento, é importante ressaltar ainda a consideração de Pasquali (2001) de que a “análise de consistência interna não constitui prova cabal de validade de construto do teste” (p. 172), já que correlações altas entre itens não indicam necessariamente que estejam medindo o mesmo construto.

Seguindo os procedimentos necessários para a validação de um instrumento, a análise da *validade de critério* verifica a eficácia do teste em prever o desempenho de um grupo específico de indivíduos, devendo ser selecionados grupos-critério referentes aos construtos avaliados pelo instrumento (Anastasi & Urbina, 2000; Pasquali, 2001). Os critérios são avaliados por outro tipo de técnica e se referem a aspectos externos ao que o instrumento mede, por exemplo, desempenhos acadêmico, profissional ou em um treinamento especializado, diagnóstico psiquiátrico, avaliações de professores, orientadores ou colegas e outros testes disponíveis (Pasquali, 2003). Os procedimentos para avaliar a validade de critério são as análises concorrente e preditiva. Na análise concorrente é realizada a avaliação simultânea da medida e do critério, por exemplo, a avaliação diagnóstica por um instrumento neuropsicológico e por avaliação neurológica. Já na análise preditiva, a avaliação do critério é feita em um período posterior à avaliação da medida em questão (predição de resultados futuros) e pode consistir na avaliação de grupos contrastantes, correlação com outros instrumentos já validados e avaliações externas (Pasquali, 2001). Esse é um tipo de validade pertinente à análise psicométrica do NEUPSILIN, visto que uma das suas finalidades é diferenciar alterações produzidas por lesões no hemisfério direito daquelas produzidas por lesões no hemisfério esquerdo.

Contudo, é necessário considerar a possibilidade de as lesões produzirem alterações comportamentais diferentes nos indivíduos graças à complexidade das associações neuronais do cérebro (Gil, 2002; Luria, 1973). Isso dificulta a definição de um critério, por exemplo, a formação de um grupo clínico mais homogêneo, ou seja, com alguma lesão e com manifestações comportamentais semelhantes. Ainda, a sua realização requer um tempo mais prolongado, visto que é evidente a dificuldade para delimitar uma amostra de sujeitos com um tipo específico de lesão em número

suficiente para a realização de análises estatísticas. Apesar desses entraves, esse é um dos procedimentos que deve ser realizado para a validação do NEUPSILIN, por ser o que evidenciaria a validade do teste em medir diferenças de lesão de hemisfério direito e esquerdo.

Evidências de validade como alternativa ao processo de legitimação de um instrumento

Os procedimentos citados anteriormente compõem um conjunto de passos que são a referência principal entre os pesquisadores que constroem instrumentos para avaliação psicológica. Entretanto, o enfoque na análise de diferentes tipos de validade (conteúdo, critério e construto) para obter um instrumento de medida apropriado tem sido questionado entre alguns pesquisadores (Urbina, 2007), sendo sugeridas novas formas de estudar a validade de um teste. Nesse sentido, a *American Educational Research Association* (AERA), a *American Psychological Association* (APA) e o *National Council on Measurement in Education* (NCME) (1999) elaboraram padrões para testagem educacional e psicológica. Essas associações não consideram necessárias as divisões de estudos de validade segundo seus tipos distintos (conteúdo, critério e construto), pois entendem que diferentes linhas de evidências de validade podem fornecer informações relevantes para uma determinada interpretação específica dos escores de um teste. Dessa forma, para AERA e colaboradores (1999), a validade é um conceito único entendido como “o grau em que todas as evidências acumuladas suportam a interpretação pretendida dos escores do teste para o propósito a que se destina” (p. 11). Assim, as evidências de validade podem ser baseadas no conteúdo do teste, no processo de resposta, na estrutura interna do teste, nas relações com outras variáveis e nas conseqüências da testagem (AERA e colaboradores, 1999).

No que se refere às evidências baseadas no conteúdo do teste e no processo de resposta, essas se assemelham aos procedimentos para análise da validade de conteúdo sugerida por Pasquali (2003). Da mesma forma, as evidências baseadas na estrutura interna do instrumento correspondem à validade de construto por meio da análise fatorial. Para exemplificar essa semelhança, a evidência de validade baseada no processo de resposta inclui análises teóricas e empíricas dos processos das respostas dos testes, as quais podem fornecer evidências da adequação entre o construto e as respostas ou desempenhos dos sujeitos avaliados. Essas evidências são provenientes, geralmente, de análises das respostas individuais e das estratégias utilizadas para responder ao teste ou podem ser realizadas a partir de outros aspectos envolvidos em um desempenho, por exemplo, o tempo de resposta. As inferências sobre os processos envolvidos nas respostas do

sujeito podem ainda ser obtidas pela análise da relação entre partes do teste ou também entre o teste e outras variáveis. Esse tipo de evidência contribui para a interpretação dos escores do teste, a fim de ser verificada a presença de capacidades irrelevantes ao construto que podem estar influenciando sua resposta. Ainda, os estudos sobre processos de resposta podem ser obtidos pela análise da consistência entre a avaliação de juízes do desempenho e os escores de interpretação pretendidos. Evidencia-se, assim, a semelhança dessa proposta aos procedimentos de validade de conteúdo.

As fontes de evidência baseadas nas conseqüências da testagem objetivam avaliar se o teste está medindo características relacionadas ao construto ou se mede outros aspectos adicionais. Portanto, é responsável por examinar a super ou sub-representação de um determinado construto, mediante a avaliação da conseqüência da aplicação do teste em uma população específica. Já a evidência baseada nas relações com outras variáveis compreende análises da relação dos escores do teste com variáveis externas ao mesmo, por exemplo, medidas de algum critério que se espera que o teste possa prever. Esse tipo de evidência inclui procedimentos que correspondem às análises convergente e discriminante, estudos de evidências experimentais (tal como, examinar se um teste de atividade acadêmica diferencia grupos instruídos de não-instruídos), além de evidências através da relação teste-critério (delineamentos de evidências preditivas e concorrentes). Portanto, são abarcados nessa forma de avaliação de evidências de validade tanto procedimentos que compõem estudos de validade de construto quanto de critério, também sugeridos por Pasquali (2003). Um ponto a destacar nessa forma de evidenciar a validade de um instrumento é que ele também inclui a possibilidade de analisar a relação entre os escores de um teste e variáveis categóricas, incluindo as que caracterizam os membros de um grupo. Essa análise é relevante para a validade de um teste quando a teoria que embasa o seu uso sugere que diferenças entre os grupos, segundo uma determinada variável categórica, deveriam estar presentes ou ausentes para a interpretação do teste ser sustentada (AERA e colaboradores, 1999).

De acordo com a AERA e colaboradores (1999), a validade de um instrumento é garantida pela integração de várias evidências de validade, as quais, juntamente com a teoria, possam sustentar a interpretação pretendida aos escores de um teste para uso específico. Portanto, as diretrizes da AERA e colaboradores ampliam as possibilidades do pesquisador que estuda as propriedades psicométricas de um instrumento de avaliação neuropsicológica. Por exemplo, em um estudo é possível apresentar evidências da validade de um teste que mede o desempenho em diversas funções cognitivas, tal como o

NEUPSILIN, pela relação dos escores do teste com variáveis que são apresentadas pelos sujeitos que irão responder aos instrumentos, ou seja, variáveis categóricas, tal como o nível de escolaridade.

Procedimentos finais para a análise das qualidades psicométricas de instrumentos de avaliação neuropsicológica

Para finalizar o processo de análise psicométrica de um instrumento é necessária a realização do estudo da fidedignidade ou da precisão do teste, que se refere “ao quanto o escore obtido no teste se aproxima do escore verdadeiro do sujeito num traço qualquer” (Pasquali, 1999, p. 66). A fidedignidade pode ser obtida a partir das análises: (a) Teste-reteste do instrumento: a fim de avaliar a sua estabilidade temporal; (b) Formas paralelas: em que são aplicadas duas formas de um mesmo teste com o objetivo de analisar a equivalência das mesmas; e (c) Consistência interna: que pode ser analisada por meio da correlação simples, do alfa (α) de Cronbach ou do método das duas metades, em que um teste é dividido em duas partes equivalentes, sendo calculada a correlação entre os escores obtidos nas duas metades. No estudo psicométrico do NEUPSILIN, podem ser realizados os procedimentos de análise de fidedignidade teste-reteste, correlação simples entre os itens dos subtestes e o resultado total da função avaliada, e concordância entre avaliadores na pontuação de alguns escores. De acordo com Urbina (2007), avaliadores ou juízes diferentes podem designar escores distintos ao desempenho em um mesmo teste ainda que existam instruções de pontuação detalhadas em um manual, pois o julgamento subjetivo dos avaliadores pode influenciar na determinação de um escore do teste.

Por fim, Pasquali (1999) afirma que a construção de um instrumento com qualidades psicométricas satisfatórias e, portanto, com condições de medir um determinado fenômeno de forma eficaz, precisa incluir ainda pesquisas de normatização e padronização. Por meio da normatização é estabelecido um padrão de referência para os resultados do teste, e a padronização aponta os meios adequados de aplicação e levantamento do instrumento. A realização de tais procedimentos concluiria as análises psicométricas do Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN.

Conclusão e considerações finais

Este artigo objetivou discutir o processo de validação psicométrica de instrumentos de avaliação neuropsicológica breve, indicando possibilidades de análise para uma bateria com subtestes que medem diferentes construtos. Atualmente, evidencia-se na área da neuropsicologia no Brasil a escassez de instrumentos adequados quanto às condições psicométricas para avaliação das funções cognitivas. Muitos dos profissionais da área

Psico-USF, v. 12, n. 2, p. 211-219, jul./dez. 2007

utilizam baterias elaboradas por eles mesmos e que incluem diferentes testes ou tarefas, os quais não foram submetidos a estudos psicométricos de validade. Dessa forma, a comparação do sujeito avaliado com o seu grupo de iguais fica comprometida e restrita ao conhecimento teórico e prático do profissional.

Nesse contexto, a discussão dos métodos de validação do NEUPSILIN contribui para o delineamento de futuras pesquisas de análises psicométricas do instrumento. Ainda mais relevante, a preocupação em desenvolver o NEUPSILIN seguindo pressupostos psicométricos de validação pretende equipar os profissionais com um material de qualidade para uso em avaliações neuropsicológicas. Diante da carência na neuropsicologia de testes nessas condições, o objetivo do NEUPSILIN é fornecer um perfil neuropsicológico breve através da avaliação de habilidades cognitivas preservadas e deficitárias. O instrumento possibilita delimitar, em apenas uma sessão, as funções com algum prejuízo, embasando a seleção de testes ou baterias neuropsicológicas específicas a serem utilizados em uma investigação mais aprofundada. Além disso, pode fornecer dados qualitativos por meio de indícios de uso de estratégias cognitivas (análise dos tipos de erros). Apresenta também a vantagem de permitir a diferenciação de seqüelas neuropsicológicas de lesões nos hemisférios cerebrais direito e esquerdo, principalmente aquelas relacionadas aos processamentos perceptivo, mnemônico e comunicativo. O NEUPSILIN pode ser classificado como um instrumento de avaliação neuropsicológica breve, porque possui um tempo reduzido de aplicação (entre 30 e 40 minutos) e inclui tarefas para acessar várias funções cognitivas. Por essa razão, classifica-se entre os instrumentos de rastreamento (*screenings*) e as baterias completas. Em função da carência desse tipo de instrumento, evidencia-se a relevância de sua construção e a importância de questionar os meios mais adequados de demonstrar a sua validade.

Essa discussão pretende instigar e instrumentalizar os pesquisadores a construir e adaptarem instrumentos neuropsicológicos válidos e confiáveis para a realidade brasileira, considerando a especificidade da análise psicométrica para cada bateria. Demonstra também a preocupação com a área da avaliação neuropsicológica, a qual vem crescendo amplamente, mas ainda carece de instrumental de avaliação adequado.

Referências

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association Publications.

- Alchieri, J. C. (2004). Aspectos instrumentais e metodológicos da avaliação psicológica. Em V. M. Andrade, F. H. dos Santos & O. F. A. Bueno (Orgs.). *Neuropsicologia hoje* (pp. 13-36). São Paulo: Artes Médicas.
- Almeida, O. P. (1998). Miniexame do Estado Mental e o diagnóstico de demência no Brasil. *Arquivos de Neuropsiquiatria*, 56(3B), 605-612.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artmed.
- Andrade, V. M. (2002). Das bases históricas da neuropsicologia à avaliação neuropsicológica. Em R. M. Cruz, J. C. Alchieri & J. J. Sardá Jr. (Orgs.). *Avaliação e medidas psicológicas: produção do conhecimento e da intervenção profissional* (pp. 27-44). São Paulo: Casa do Psicólogo.
- Bandeira, D. R., Serafini, A. J., Fonseca, R. P., Salles, J. F. & Parente, M. A. M. P. (2005). *Contribuições da psicométrica para a área de avaliação neuropsicológica*. Trabalho apresentado no II Congresso Brasileiro de Avaliação Psicológica, Gramado, RS.
- Beerl, M. S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H. & Silverman, J. M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, 67(2), 1006-1010.
- Bertolucci, P. H. F., Brucki, S. M. D., Campacci, S. R. & Juliano, Y. (1994). O Miniexame do Estado Mental em uma população geral: impacto da escolaridade. *Arquivos de Neuropsiquiatria*, 52(1), 1-7.
- Bertolucci, P. H. F., Okamoto, I. H., Brucki, S. M. D., Siviero, M. O., Neto, J. T. & Ramos, L. R. (2001). Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arquivos de Neuropsiquiatria*, 59(3A), 532-536.
- Brucki, S. M. D., Nitrini, R., Caramelli, P., Bertolucci, P. H. F. & Okamoto, I. H. (2003). Sugestões para o uso do Miniexame do Estado Mental no Brasil. *Arquivos de Neuropsiquiatria*, 61(3B), 777-781.
- Capovilla, A. G. S., Joly, M. C. R. A. & Tonelotto, J. M. F. (2006). Avaliação neuropsicológica e aprendizagem. Em A. P. P. Noronha, A. A. A. dos Santos & F. F. Sisto (Orgs.). *Facetas do fazer em avaliação psicológica* (pp. 141-162). São Paulo: Vetor.
- Chaves, M. L. F. & Izquierdo, I. (1992). Differential diagnosis between dementia and depression: A study of efficiency increment. *Acta Neurologica Scandinavica*, 85, 378-382.
- Conselho Federal de Psicologia (2004). *Resolução CFP nº 2/2004*. Brasília: CFP.
- Conselho Federal de Psicologia (2003). *Resolução CFP nº 2/2003*. Brasília, CFP.
- Ferreira, E. D. (1997). *Avaliação de pacientes com demência do tipo Alzheimer e multinfarto: um estudo de coorte*. Dissertação de Mestrado. Porto Alegre: Universidade Federal do Rio Grande do Sul – Curso de Pós-Graduação em Clínica Médica.
- Fonseca, R. P. (2006). *Bateria Montreal de Avaliação da Comunicação: estudos teóricos, sócio-demográfico, psicométrico e neuropsicológico*. Tese de Doutorado. Porto Alegre: Universidade Federal do Rio Grande do Sul – Instituto de Psicologia.
- Fonseca, R. P., Salles, J. F. & Parente, M. A. M. P. (2006). Construção do Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN: Análise semântica dos itens. *Anais do 2º Congresso Brasileiro de Cérebro, Comportamento e Emoções*, Bento Gonçalves, Brasil.
- Fonseca, R. P., Salles, J. F. & Parente, M. A. M. P. (no prelo). *Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN*. São Paulo: Vetor.
- Gil, R. (2002). *Neuropsicologia*. São Paulo: Santos.
- Kane, R. L. (1991). Standardized and flexible batteries in neuropsychology: An assessment update. *Neuropsychology Review*, 2(4), 281-339.
- Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.
- Mäder, M. J. (2002). Avaliação neuropsicológica: da pesquisa à prática clínica com adultos. Em R. M. Cruz, J. C. Alchieri & J. J. Sardá Jr. (Orgs.). *Avaliação e medidas psicológicas: produção do conhecimento e da intervenção profissional* (pp. 47-68). São Paulo: Casa do Psicólogo.
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D. & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39, 1159-1165.
- Ostrosky-Solís, F., Ardila, A. & Rosselli, M. (1999). NEUROPSI: A brief neuropsychological test battery in Spanish with norms by age and educational level. *Journal of the International Neuropsychological Society*, 5, 413-433.
- Pasquali, L. (1999). *Instrumentos psicológicos: manual prático de elaboração*. Brasília: LabPAM/ IBAPP.
- Pasquali, L. (2001). *Técnicas de exame psicológico – TEP: manual*. São Paulo: Casa do Psicólogo.
- Pasquali, L. (2003). *Psicométrica: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes.
- Pawlowski, J., Bandeira D. R., Parente M. A. M. P., Fonseca, R. P., Salles, J. F. & Gonchoroski, M. V. C. (2006). Estudo piloto do Instrumento de Avaliação

- Neuropsicológica Breve NEUPSILIN. *Anais do 2º Congresso Brasileiro de Cérebro, Comportamento e Emoções*, Bento Gonçalves, Brasil.
- Peña-Casanova, J. (1985). La exploracion neuropsicológica. *VII Congreso Nacional de Neurología*. Barcelona: Sociedad Española de Neurología.
- Santos, D. O. (2005). *Influência do envelhecimento no desempenho de adultos no Instrumento de Avaliação Neuropsicológica Breve NEUPSILIN*. Monografia de Especialização. Porto Alegre: Universidade Federal do Rio Grande do Sul – Instituto de Psicologia.
- Simões, M. R. (2002). Avaliação psicológica em crianças e adolescentes. Em R. Primi (Org). *Temas em avaliação psicológica* (pp. 26-34). São Paulo: Impressão Digital do Brasil.
- Tupper, D. E. (2000). Introduction: Neuropsychological assessment après Luria. *Neuropsychological Review*, 9(2), 57-61.
- Urbina, S. (2007). *Fundamentos da testagem psicológica* (C. Dornelles, Trad.). Porto Alegre: Artes Médicas. (Original publicado em 2004).

*Recebido em junho de 2006
Reformulado em agosto de 2007
Aprovado em setembro de 2007*

Sobre as autoras:

Josiane Pawłowski é psicóloga, doutoranda em Psicologia pelo Programa de Pós-Graduação em Psicologia da Universidade Federal do Rio Grande do Sul (UFRGS), mestre em Psicologia (UFRGS) e especialista em Psicologia Clínica – Ênfase em Avaliação Psicológica (UFRGS).

Clarissa Marceli Trentini é psicóloga, doutora em Ciências Médicas: Psiquiatria pela Universidade Federal do Rio Grande do Sul (UFRGS) e professora do Instituto de Psicologia e do Programa de Pós-Graduação em Psicologia da UFRGS.

Denise Ruschel Bandeira é psicóloga, doutora em Psicologia pela Universidade Federal do Rio Grande do Sul (UFRGS) e professora do Instituto de Psicologia e do Programa de Pós-Graduação em Psicologia da UFRGS.

