

Avaliação Psicológica no Brasil: Fundamentos, Situação Atual e Direções para o Futuro¹

Ricardo Primi²
Universidade São Francisco

RESUMO - Este artigo apresenta um panorama de questões importantes da área de avaliação psicológica nos últimos 25 anos no Brasil. Buscou-se discutir fundamentos epistemológicos da área, bem como suas relações com a ciência e a integração do pensamento nomotético e idiográfico da pesquisa com a prática profissional. Faz-se um apanhado histórico de eventos importantes e uma análise geral da produção de artigos. Descreve-se também a produção de instrumentos e o Sistema de Avaliação de Testes Psicológicos do Conselho Federal de Psicologia. Posteriormente, são apontadas perspectivas para o futuro em quatro áreas: avanços metodológicos e tecnológicos, integração de abordagens e avanço dos seus métodos, validade consequencial e relevância social, e incentivo à formação e à criação da especialidade em avaliação psicológica.

Palavras-chave: avaliação psicológica; métodos quantitativos; testes psicológicos; SATEPSI.

Psychological Assessment in Brazil: Foundations, Current Situation and Future Directions

ABSTRACT - This article presents an overview of important issues in the area of psychological assessment for the last 25 years in Brazil. It discusses the area's epistemological elements as well as their relations to science, and the integration of nomothetic and idiographic thinking in research with professional practice. It is presented a historical overview of major events and a general analysis of the publication in the area. The production of instruments and the Psychological Test Evaluation System of the Federal Council of Psychology are also described. Later, future perspectives are highlighted in four areas: methodological and technological advances, integration of approaches and advancement of their methods, consequential validity and social relevance, and the need to encourage training and the creation of expertise in psychological assessment.

Keywords: psychological assessment; quantitative methods; psychological tests; SATEPSI.

Este artigo foi produzido por ocasião da comemoração dos 25 anos da publicação da revista *Psicologia: Teoria e Pesquisa*, a partir de um desafio colocado pela sua editora para que se apresentasse a trajetória teórica, conceitual, empírica e tecnológica da área de avaliação psicológica no Brasil nos últimos 25 anos, as influências mútuas em relação ao panorama internacional, apontando, ainda, as perspectivas futuras. Portanto, nas páginas que se seguem, tenta-se responder a esse desafio. Inicialmente, alguns aspectos dos fundamentos epistemológicos da área são apontados; em seguida, são apresentados eventos históricos importantes e uma avaliação de seu crescimento; e, por fim, é realizado um exercício de reflexão sobre o que seria necessário para o avanço da área. Buscou-se selecionar tópicos peculiares da diversidade de pensamentos na psicologia, trazendo referências internacionais de estudos e métodos que podem ser úteis para o avanço da área.

Avaliação como Objetivação das Abordagens Teóricas

A avaliação psicológica é, talvez, uma das áreas mais antigas da psicologia. Ao nascer, teve uma de suas aplicações práticas – o desenvolvimento dos testes psicológicos e da psicometria – voltada para seleção de soldados nas grandes guerras (Anastasi & Urbina, 2000). Dessa forma, a avaliação é muitas vezes identificada com um segmento particular da psicologia dedicado à criação de instrumentos e técnicas. No entanto, a avaliação, em geral, e, em particular, o desenvolvimento de instrumentos, representa uma área central da ciência psicológica porque permite a objetivação e operacionalização de teorias psicológicas. Em outro momento ressaltou-se esse aspecto:

A avaliação psicológica é geralmente entendida como uma área aplicada, técnica, de produção de instrumentos para o psicólogo, visão certamente simplista da área. A avaliação psicológica não é simplesmente uma área técnica produtora de ferramentas profissionais, mas sim a área da psicologia responsável pela operacionalização das teorias psicológicas em eventos observáveis. Com isso, ela fomenta a observação

1 As atividades de pesquisa do autor são financiadas pelo CNPq, FAPESP e CAPES/INEP.

2 Endereço para correspondência: Rua Ferreira Penteado, 1518, Apto. 41. Bairro Cambuí. Campinas, SP. CEP 13025-357. Fone: (19) 33655164; (19) 81492244. E-mail: rprimi@mac.com.

sistemática de eventos psicológicos, abrindo os caminhos para a integração teoria e prática. Ela permite que as teorias possam ser testadas, eventualmente aprimoradas, contribuindo para a evolução do conhecimento na psicologia. Portanto, a avaliação na psicologia é uma área fundamental de integração entre a ciência e a profissão. Disso decorre que o avanço da avaliação psicológica não é um avanço simplesmente da instrumentação, mas sobretudo das teorias explicativas do funcionamento psicológico. (Primi, 2003, p. 68)

Segundo Muniz (2004), o processo de validação de instrumentos psicológicos se constitui em um caso particular de um processo mais geral, de validação de hipóteses científicas. Em ambos os casos, tenta-se validar explicações por meio de um processo hipotético-dedutivo, no qual se levantam hipóteses teóricas, planejam-se estudos empíricos, coletam-se e analisam-se dados, buscando-se testar as hipóteses explicativas, falseando-as ou corroborando-as. Esse processo interativo teoria-hipótese-falseamento encontra-se na base do desenvolvimento do conhecimento e da maturidade da psicologia como ciência. A diferença entre validar uma teoria ou um teste situa-se nos seguintes fatos: no primeiro caso, há um processo mais amplo, visto que tenta-se validar a existência de construtos e as relações causais entre eles; e no segundo caso, tenta-se validar as interpretações sobre o construto psicológico que são feitas a partir do instrumento. As semelhanças e diferenças podem ser visualizadas na Tabela 1.

O que se pode notar é que há uma relação estreita entre os instrumentos e a pesquisa científica uma vez que os estudos empíricos fazem uso dos instrumentos para observar determinados construtos no percurso de validar determinadas explicações sobre o comportamento humano. Wright (1999), um dos pioneiros no desenvolvimento do modelo de Rasch nos Estados Unidos, propôs um modelo de filosofia da ciência envolvendo cinco estágios: exposição, observação, medida, análise e teoria. A produção do conhecimento científico se inicia com a exposição ou consciência dos fenômenos. Em seguida, são organizados meios mais sistemáticos de observação, como itens e testes. A essas observações são aplicados modelos matemáticos, como a Teoria de Resposta ao Item, transformando as observações em medidas. Só então é que essas medidas são transformadas em teorias entendidas como abstrações que servem para prever eventos da realidade de maneira mais generalizada.

Assim, a avaliação psicológica, especialmente aquela parte que se dedica ao desenvolvimento de instrumentos, é uma área nuclear da psicologia e de sua edificação enquanto ciência. Em primeiro lugar porque envolve a objetivação dos conceitos teóricos em elementos observáveis. Em segundo lugar porque requer aplicação de método científico baseado no conhecimento sobre quais delineamentos (levantamento, correlacional, quasi-experimental e experimental) são mais adequados ao conhecimento que se deseja ter. Em terceiro lugar porque envolve também o uso de modelagem matemática na representação dos processos psicológicos, abordagem que vem gradativamente substituindo o modelo clássico de análise de dados baseado somente no teste de significância da hipótese nula (Rodgers, 2010). E, por último, porque seus produtos (instrumentos de medida) são peças necessárias ao desenvolvimento do conhecimento científico dentro da psicologia. Por esses motivos, ao se tratar do tema avaliação, sua história e seu desenvolvimento, não se está falando de um assunto restrito a uma determinada área, mas sim dos fundamentos mais gerais da psicologia.

Avaliação Psicológica, Testes Psicológicos e as Abordagens de Pensamento Nomotético e Idiográfico

Ao se tratar do termo amplo, avaliação psicológica, deve-se, em primeiro lugar, distingui-lo dos instrumentos de avaliação. A avaliação psicológica é uma atividade mais complexa e constitui-se na busca sistemática de conhecimento a respeito do funcionamento psicológico das pessoas, de tal forma a poder orientar ações e decisões futuras. Esse conhecimento é sempre gerado em situações que envolvem questões e problemas específicos. Já os instrumentos de avaliação constituem-se em procedimentos sistemáticos de coleta de informações úteis e confiáveis que possam servir de base ao processo mais amplo e complexo da avaliação psicológica. Portanto, os instrumentos estão contidos no processo mais amplo da avaliação psicológica (Primi, Nascimento & Souza, 2004).

Em geral, os instrumentos são meios padronizados de se obter amostras/indicadores comportamentais que irão revelar diferenças individuais nos construtos, traços latentes ou processos mentais subjacentes. Presume-se, então, que os

Tabela 1. Estágios da validação de teorias científicas comparada à validação de testes.

	Validação de Teorias	Validação de Testes
Teoria	Explicações sobre fenômenos Psicológicos e suas causas	Definição do construto e das interpretações dos indicadores ou escores derivados do instrumento
Hipóteses e Objetivos	Deduções de previsões sobre eventos da realidade	Deduções de associações internas e externas
Delineamento	Planejamento do levantamento de dados procurando testar as explicações derivadas das deduções	Planejamento dos estudos de validade testando as previsões derivadas
Parte Empírica	Coleta e análise dos dados	Coleta e análise de dados
Conclusão	Falseamento ou corroboração das hipóteses explicativas e realimentação ou reformulação das teorias	Falseamento ou corroboração das interpretações pretendidas para os escores ou indicadores do teste

traços latentes são as variáveis causais dos comportamentos que se manifestam na situação de testagem. Dessa forma, o processo amplo de medida consiste em uma via indireta que, por meio da observação dos indicadores, torna possível se inferir algo sobre o construto que se deseja avaliar (Gottfredson & Saklofske, 2009). Assim, de acordo com Borsboom, Mellenbergh e Heerden (2004), os estudos de validade tentam provar a relação causal entre as variações no construto subjacente e as variações nos indicadores comportamentais avaliados pelo instrumento, justificando, dessa maneira, os sentidos atribuídos aos escores em relação ao construto. Essa conceituação deixa mais claro que, no âmbito dos estudos de validade, há uma questão de relação de causalidade entre o construto e os indicadores. Consequentemente, o processo de validação dos testes envolve todos os desafios metodológicos ao se deparar com a necessidade de estabelecer relações funcionais entre duas variáveis, nesse caso, entre uma variável latente, o construto, e outra observada, os indicadores.

Ao se tratar dos fundamentos da avaliação psicológica, é preciso entender a diversidade de estilos de pensamento que são subjacentes às práticas de diferentes grupos dentro da área. Essa diversidade de métodos e estilos revelam aspectos fundamentais da avaliação que precisam ser compreendidos e integrados em um modelo mais amplo com vários níveis que se tentará esboçar mais adiante. Cronbach (1996) define essas diferenças ao descrever as características dos estilos psicométrico (nomotético) e impressionista (idiográfico), apresentados na Tabela 2. Esses modos de raciocínio podem ser compreendidos por diferentes tradições, uma mais psicométrica, voltada à pesquisa e à descoberta de leis gerais, e outra mais aplicada, voltada à prática clínica e ao entendimento da riqueza de um indivíduo.

A primeira delas, a tradição psicométrica, pode ser exemplificada ao olharmos, por exemplo, para o início das teorias fatorias de personalidade nos trabalhos de Cattell (1957, 1973). O autor afirma que:

A mensuração é o fundamento da ciência. Mas, em personalidade, deve-se começar com a descoberta das formas naturais de padrões de comportamentos humanos. Devemos definir os traços unitários naturais, por exemplo, ansiedade, conscienciosidade, força do ego, dominância, que constituem a topografia (ou taxonomia) da personalidade. Somente depois estaremos prontos para construir escalas e baterias para medir tais traços. Chamo o primeiro passo de pesquisa da estrutura (ou taxonômica) e o segundo, desenvolvimento estrutural de escalas. (Cattell, 1973, p. 2)

Seguindo esse objetivo, Cattell passou a analisar estruturas manifestas em diferentes dados observacionais oriundos de três fontes: (a) respostas a questionários em que as pessoas são as próprias observadoras e reladoras de seu comportamento (dados Q), (b) dados de pessoas por meio de observações de terceiros que relatam o que vêem nelas em sua vida diária (dados L); e (c) medidas relativamente diretas de comportamentos em situação de testagem em laboratório (dados T). Usando a análise fatorial com o objetivo de sistematizar as correlações entre indicadores, de forma a inferir os traços unitários, Cattell (1957) encontrou mais de 23 traços básicos que supostamente seriam as forças causais dos comportamentos observados. Há uma analogia de Cattell que é muito interessante para esclarecer aspectos dessa metodologia:

O problema que por muitos anos desconcertou os psicólogos era encontrar um método que deslindasse essas influências funcionalmente unitárias na floresta caótica do comportamento humano. Mas como é que numa floresta tropical de fato decide o caçador se as manchas escuras que vê são dois ou três troncos apodrecidos ou um só jacaré? Ele fica à espera de movimento. Se eles se movem juntos - aparecem e desaparecem juntos - ele conclui por uma única estrutura. Da mesma forma, como John Stuart Mill observou em sua filosofia da ciência, o cientista deveria ter em mira a "variação concomitante" na busca de conceitos unitários. (Cattell, 1965, p. 55)

Tabela 2. Abordagens de raciocínio na avaliação psicológica segundo Cronbach (1996).

Estilo Psicométrico	Estilo Impressionista
Foco nomotético: interpretações focadas na aplicação de regras gerais aos casos individuais derivados dos estudos de validade	Foco idiográfico: interpretações focadas na combinação impressionista de dados individuais
Semelhança entre variáveis	Semelhança entre pessoas
Analítico: olha uma variável de cada vez	Holístico: tenta olhar várias variáveis ao mesmo tempo
Ênfase na padronização dos estímulos e respostas fechadas, elaboradas previamente para maximizar a objetividade	Ênfase na liberdade das respostas construídas pelo sujeito para maximizar a abrangência e riqueza individual de expressão
Inventários e testes de inteligência na área educacional	Testes projetivos na área clínica
Ênfase no instrumento	Ênfase no profissional

Observa-se que os conceitos fundamentais da abordagem psicométrica (traços) são ancorados nas dimensões interindividuais que explicam consistências comportamentais. Abstraem-se as dimensões psicológicas a partir da análise da semelhança entre variáveis interindividuais. Por exemplo, se, dentro de um grupo de pessoas, observa-se que elas diferem em suas capacidades cognitivas, mas as mais habilidosas são igualmente boas em guardar nomes, rapidez de leitura, comunicação verbal e escrita (variáveis observadas), pode-se, a partir disso, inferir uma dimensão que poderia se chamar capacidade verbal (variável latente). Portanto, a atenção volta-se às variáveis abstraídas dos indivíduos. Nesse modelo, observa-se as diferenças individuais e, pela semelhança entre as variáveis (por meio das correlações entre elas), abstrai-se um conceito. Essa abordagem constitui-se como mais analítica ao buscar definir as dimensões da personalidade, descobrindo-as uma a uma, de forma a segmentar o indivíduo em seus elementos estruturais por meio de observações das diferenças individuais dentro de grupos, de forma a construir um sistema taxonômico que, subsequentemente, será utilizado para descrever um indivíduo de maneira mais completa. Assim, retomando a analogia de Cattell (1965), essa abordagem, ao olhar para a floresta, focaliza a atenção nos jacarés.

Em outro extremo, encontra-se a segunda tradição, o estilo impressionista, oriundo do modelo clínico que tem como foco o indivíduo e estudos de casos. Nessa tradição, busca-se compreender mais profundamente o indivíduo, considerando todas as variáveis disponíveis sobre a pessoa, bem como sua interação na configuração de um padrão individual único. Muitas vezes, entende-se que essa configuração é tão única que dificilmente se repetirá em outra pessoa, derivando-se daí a noção do idiossincrático. Um ponto importante a ser destacado refere-se ao fato de que os conceitos fundamentais dessa abordagem mais clínica (táxons) estão ancorados em configurações intraindividuais que são usadas para explicar e entender o sujeito. Retomando a analogia, essa abordagem focaliza a atenção na floresta, incluindo tudo o que a compõe, não somente o jacaré mas também todas as outras formas lá existentes que interagem com ele.

Assim, em contextos mais clínicos da saúde, assim como na abordagem psicométrica, foram derivados sistemas taxonômicos a partir do estudo das semelhanças entre perfis, isto é, configurações de características internas de cada indivíduo. Um exemplo é o sistema de classificação constante no eixo II do DSM-IV, que define os transtornos de personalidade (Millon, Grossman, Millon, Meagher & Ramnath, 2004). Nesse, ao se observar em um indivíduo a combinação de pelo menos três das seguintes características: o fracasso em se conformar com as normas sociais, propensão a enganar, impulsividade, agressividade, desrespeito com a segurança, irresponsabilidade e ausência de remorso, de maneira estável e inflexível associada ao prejuízo no funcionamento social ou ocupacional, pode-se inferir uma configuração do transtorno anti-social, uma das categorias de classificação do Eixo II. Embora tais sistemas de classificação diagnóstica tenham sido originados de uma tradição mais clínica, os adeptos dessa abordagem, com um entendimento mais radicalmente idiossincrático, são bastante

resistentes ao uso dessas classificações devido ao seu caráter nomotético e simplificador.

Entretanto, um aspecto importante situa-se na diferença entre conceitos basilares dessas duas abordagens, uma iniciando sua conceituação pelas variáveis interindividuais abstraídas dos indivíduos, deixando-os em segundo plano, e a outra iniciando-se pelas configurações intraindividuais, de forma a colocar o indivíduo em primeiro plano e deixando as variáveis em segundo plano. A Figura 1 exemplifica esses focos diferenciados. Nela, há 10 indivíduos exibindo níveis diferentes de ansiedade e depressão. Há uma correlação significativa entre essas duas variáveis de 0,45 ($p < 0,05$), indicando uma tendência a se observar que pessoas ansiosas tendem a ser deprimidas. Essa informação é abstraída da relação entre as variáveis e descreve o perfil de uma parcela importante dos indivíduos (alta depressão e ansiedade), mas não o de alguns deles que são só deprimidos ou só ansiosos. Então, no nível individual, pode-se notar uma quantidade considerável de contradições diante do sentido mais geral, indicado pela correlação entre as variáveis. Assim, é possível entender como, às vezes, esses dois focos diferenciados chegam a dados aparentemente contraditórios.

O estilo de pensamento mais clínico e impressionista encontra-se na origem das técnicas projetivas, por exemplo, ao enfatizar interpretações mais holísticas e flexíveis e ao considerar de maneira mais livre o conjunto de variáveis expressas, de maneira a buscar a formulação de entendimentos mais amplos sobre a pessoa. Os exemplos contraditórios das predições nomotéticas da abordagem psicométrica, tal como o ilustrado na Figura 1, são enfatizados, sugerindo que os métodos mais psicométricos empobrecem a avaliação e não dão conta de entender toda a complexidade individual. Por outro lado, o estilo de pensamento mais psicométrico, que deu origem às escalas de autorrelato e aos testes de inteligência, afirma-se por meio do embasamento empírico e pelos procedimentos mais sistemáticos que culminam em um sistema taxonômico descritivo mais objetivo e sustentado, criticando outras abordagens justamente pela ausência desses elementos.

É claro que as interpretações validadas nomoteticamente a partir da análise das diferenças individuais não serão válidas para uma parcela importante de casos. Para essa parcela, uma abordagem idiográfica tentará explicar as incongruências encontradas, considerando-se outras informações disponíveis por meio da análise de configurações intraindividuais. Mas a abordagem puramente clínica-individual só seria justificada se a realidade fosse radicalmente idiográfica, isto é, se as combinações psicológicas características ocorressem somente uma vez e nunca mais se repetissem. Entretanto, o que se nota é a existência de tendências gerais que superam o acaso, de forma que adotar uma abordagem puramente idiográfica implicaria não se beneficiar dos inúmeros estudos que a psicologia tem produzido, cujos resultados têm indicado a existência de certas “regularidades” no comportamento humano. As predições, derivadas de estudos nomotéticos de validade, baseiam-se em fatos mais frequentes e, por isso, devem ser utilizadas como guias para as hipóteses iniciais em um processo de avaliação. Mas, eventualmente, essas interpretações precisarão ser adaptadas aos casos individuais

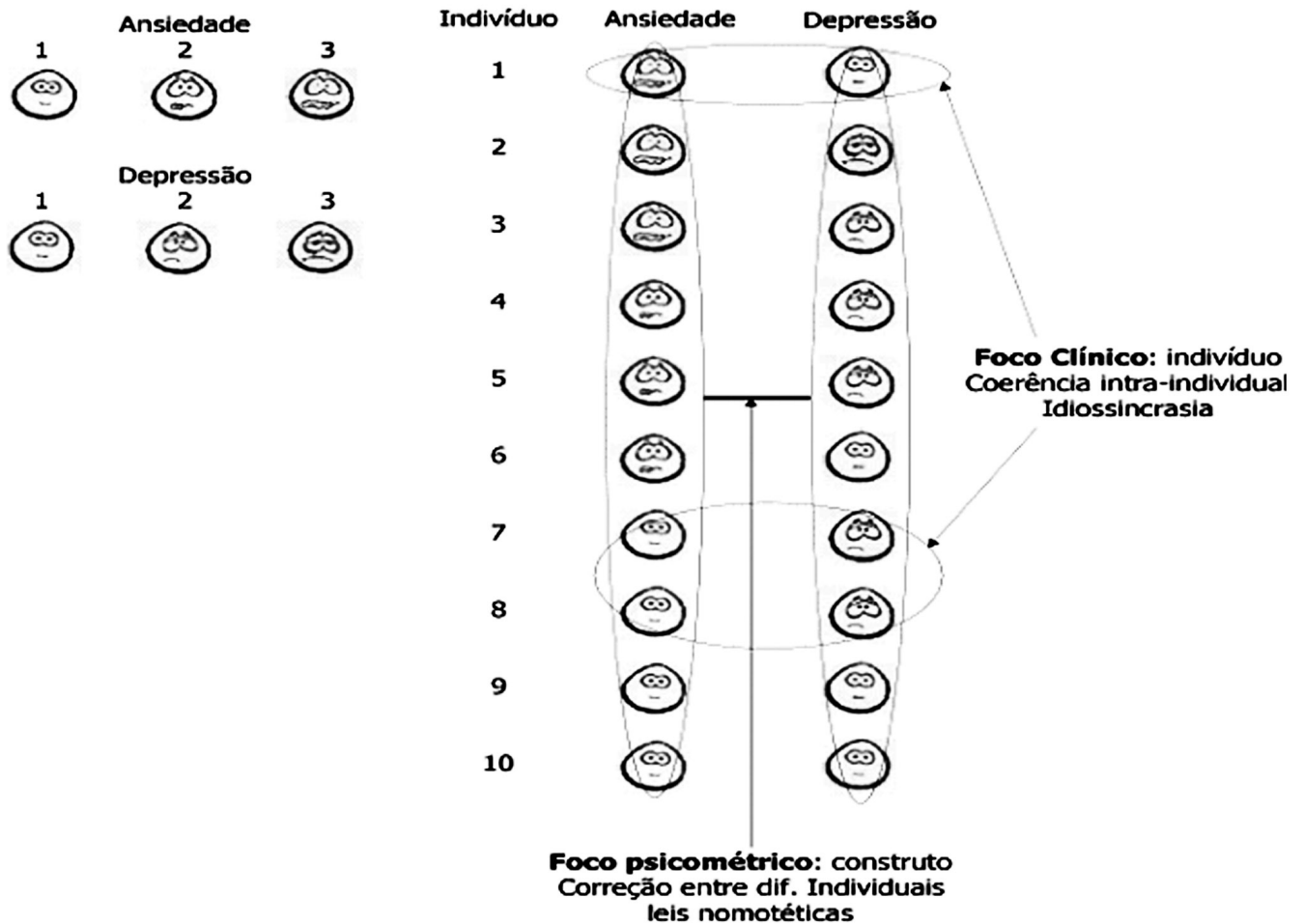


Figura 1. Exemplificação dos focos psicométrico (análise da semelhança das diferenças interindividuais) e clínico (análise dos perfis intra-individuais).

para efetivamente se chegar a uma compreensão válida de um indivíduo. Nesse sentido, Tavares (2003) define essas adaptações com o termo “validade clínica”, argumentando que só assim poderemos chegar a compreensões efetivamente válidas sobre os indivíduos.

Dessa forma, podemos ver que as duas abordagens não são contrárias, mas sim complementares, e representam estágios do raciocínio sistemático da psicologia, na tentativa de medir e avaliar, durante a busca pela compreensão do ser humano. Millon e cols. (2004) propõem o conceito de níveis de interpretação, um modelo multinível integrativo de diferentes abordagens, que pode ser aplicado aqui. Nesse modelo, os instrumentos de avaliação contêm itens ou indicadores que são combinados para produzir escalas. Tem-se aí, atrelados ao instrumento, dois níveis: o primeiro, dos indicadores, e o segundo, das escalas (medidas de traços) que agregam itens/indicadores. Nesses níveis, os métodos psicométricos têm um papel importante para validar os construtos (traços) e seus indicadores e também para produzir medidas nas quais é possível aplicar, por exemplo, o modelo de Rasch, com a finalidade de se criar medidas de cada dimensão, que consiste atualmente no estado da arte em construção de medidas quantitativas (Wilson, 2005; Wright & Stone, 2004). Em seguida, em um nível hierarquicamente superior, estão os perfis (configurações de notas nas escalas ou de presença ou

ausência de indicadores qualitativos) que irão caracterizar de maneira mais global cada indivíduo. Nesse nível, como ponto de partida, tornam-se relevantes os sistemas taxonômicos clínicos baseados em estudos das semelhanças entre perfis de indivíduos. Entretanto, convém salientar que:

a abordagem idiográfica nos lembra que os construtos diagnósticos são somente pontos de referência que facilitam o entendimento. Se, por exemplo, o indivíduo é caracterizado como narcisista, a próxima questão seria: quão diferente essa pessoa é do tipo narcisista puro? Tais questões desviam a atenção de categorias diagnósticas simples em direção ao entendimento do indivíduo. Como o objetivo é o entendimento idiográfico da pessoa, a avaliação é, em realidade, um esforço para mostrar as limitações das variáveis e categorias diagnósticas ao descrever aquela pessoa avaliada ... o estudo da personalidade começa como ciência e termina com arte. (Millon & cols., 2004, p. 120)

Um último nível, hierarquicamente superior, refere-se a todas as outras variáveis referentes ao contexto, tais como outras características da pessoa e o contexto histórico imediato e de longo prazo que podem interferir ou interagir no sentido das variáveis e do perfil a ser interpretado. Como pode ser notado, essa concepção prevê diversos níveis, desde a testagem (itens e escalas) até a avaliação (perfis e contexto),

dentro dos quais os diferentes estilos de pensamento precisam ser integrados de maneira a trazer as vantagens de sua abordagem, servindo ao entendimento mais amplo das pessoas.

Nesse momento, é oportuno mencionar a existência de algumas correntes de pensamento muito influentes na psicologia, que possuem uma abordagem muito crítica em relação à avaliação psicológica. As críticas se baseiam na análise dos condicionantes históricos das concepções em psicologia, considerando-os como forma de legitimar algumas ideologias, resultando certos conceitos psicológicos em mecanismos sociais de perpetuação da dominação e das injustiças cometidas. Essa visão tem feito críticas à mensuração e a certos construtos na psicologia como, por exemplo, a personalidade antissocial, déficit de atenção e hiperatividade, questionando sua existência e mesmo a possibilidade de medi-los. Se, por um lado, há sim maus exemplos de uso preconceituoso e irrefletido de instrumentos de avaliação, por outro, a generalização acrítica dessa atitude pode criar um grande prejuízo para a psicologia, pois acaba por desmerecer conceitos e instrumentos com sólido valor para a prática e teoria. Os instrumentos podem, até mesmo, ser usados para a implementação de estudos e de intervenções sociais, como querem esses críticos. Os problemas do mau uso reforçam a necessidade constante de uma reflexão crítica sobre o contexto e a construção histórica de certos entendimentos em psicologia, mas a crítica deve ter um caráter mais construtivo, melhorando a prática, e não destrutivo. Tal atitude negativa dissemina o preconceito em relação aos instrumentos e aos psicólogos que os utilizam, como se todos fossem acrílicos e socialmente descompromissados.

Novamente, não há porque haver contradição entre o raciocínio crítico mais amplo e uma prática mais técnica ligada à mensuração e avaliação, visto que são processos complementares. Como se salientou acima, o raciocínio tem vários níveis, e as reflexões críticas sobre o processo amplo (fundamento da avaliação, história das demandas) são elementos importantes em um nível mais amplo da prática em avaliação. Um conceito importante a esse respeito tem a ver com a validade consequencial (Messick, 1980), que será discutida mais adiante quando forem tratadas as perspectivas de avanço para a área.

Eventos Históricos e o Desenvolvimento da Área no Brasil

Tendo discutido aspectos relacionados aos alicerces da avaliação psicológica e à diversidade de pensamentos que se observa atualmente nas discussões relacionadas a esse tema, nesta seção do artigo pretende-se fazer uma descrição de eventos históricos importantes da área que ocorreram nos últimos 25 anos. Não se pretende fazer um levantamento histórico compreensivo, mas sim tentar elencar os eventos que, na opinião do autor, foram relevantes para a área.

Ao longo desse período, especialmente a partir da década de 1990, pode-se observar a movimentação da área por meio da organização de eventos dedicados à avaliação psicológica, com mais de uma edição anual, promovidos em São Paulo pela USP (Encontro de Técnicas do Exame Psicológico), em Porto Alegre pela UFRGS e PUC-RS (Encontro Nacional

sobre Testes Psicológicos), em Minas Gerais pela UFMG, PUC-MG, Centro Universitário Newton Paiva e FUMEC (Encontro Mineiro de Avaliação Psicológica) e em Ribeirão Preto pela USP-RP (Encontro da Sociedade Brasileira de Rorschach e outros métodos projetivos, que em 2004 passou a se chamar Associação Brasileira de Rorschach e Métodos Projetivos - ASBRo).

Tais eventos fomentaram a reunião, intercâmbio e organização de pesquisadores e profissionais da área, o que culminou na criação e consolidação das duas sociedades científicas mais representativas da área atualmente: a já mencionada ASBRo³, fundada em 1993, e o Instituto Brasileiro de Avaliação Psicológica (IBAP)⁴, fundado em 1997. Ambas integram o Fórum Nacional das Entidades Brasileiras de Psicologia. Posteriormente à criação do IBAP, os Encontros Nacionais sobre Testes Psicológicos passaram a ser realizados por esse instituto, sendo que a sua última edição, em 2009, foi resultado de uma organização conjunta do IBAP e da ASBRo. O processo de consolidação das entidades pode ser notado na criação da primeira revista específica da área, intitulada “Avaliação Psicológica”, cujo início deu-se em 2002⁵. Também, em 2000, foi criada uma lista de discussão de profissionais e estudantes interessados na área (avalpsi@yahoogrupos.com.br), que gradativamente se tornou um importante espaço de discussão e troca de informações. Um dos temas discutidos na lista, por exemplo, resultou na anulação de duas questões mal elaboradas do Exame Nacional de Cursos de Psicologia de 2001 e 2002. Atualmente, a lista contém 569 membros.

Em nível de pós graduação *stricto sensu* também é possível observar a ampliação da área ocorrida no mesmo período. Tradicionalmente, a avaliação psicológica se organiza em laboratórios e linhas de pesquisa. Dos 65 programas existentes no Brasil, nove (UFMG, UFU, UFRGS, PUC/RS, UFSC, USP/RP, USP/SP, PUCCAMP, UnB) contam com linhas de pesquisa na área. Em 2001, a Universidade São Francisco deu início ao primeiro programa com área de concentração em Avaliação Psicológica, até o momento o único no país. O que se pode notar é que, se por um lado houve um aumento importante da área nos programas de pós-graduação, por outro, em número absoluto, essa representação é ainda pequena. Também é possível notar o crescimento de grupos ligados à área nas reuniões da Associação Nacional de Pesquisa e Pós-Graduação em Psicologia (ANPEPP) que, na sua última reunião, realizada em junho de 2010, contou com quatro grupos de trabalho (GT-7: Avaliação de crianças e adolescentes, GT-8: Avaliação de programas, GT-28: Métodos projetivos nos contextos de avaliação psicológica e GT-33; Pesquisa em avaliação psicológica).

Todos esses eventos, em geral, indicam que a área da avaliação psicológica cresceu muito nos últimos 20 anos. Uma constatação desse crescimento pode também ser notada na quantidade de publicações. Uma busca simples, para este artigo, realizada no INDEX-Psi⁶, que é uma base de dados de artigos de periódicos brasileiros em psicologia, utilizando como palavra chave os descritores “avaliação” ou “testes”,

3 <http://www.asbro.org.br>.

4 <http://www.ibapnet.org.br>.

5 <http://www.ibapnet.org.br/avalpsi.html>.

6 <http://www.psi.bvs.br>.

resultou nos dados resumidos na Figura 2. Essa figura mostra um claro padrão de crescimento no número de publicações nos últimos anos. Há um padrão relativamente constante entre 1985 até 1993, período em que se encontra, em média, 32 publicações por ano. A partir de 1994, nota-se o início de um padrão de crescimento, sendo que, nos últimos cinco anos, excluindo-se o ano de 2010 por ainda estar em curso, observam-se, em média, 95 trabalhos por ano. Portanto, o que se pode visualizar é que, na última década, a produção quase triplicou em relação à década de 1980.

Em termos de impacto na prática profissional, com cer-

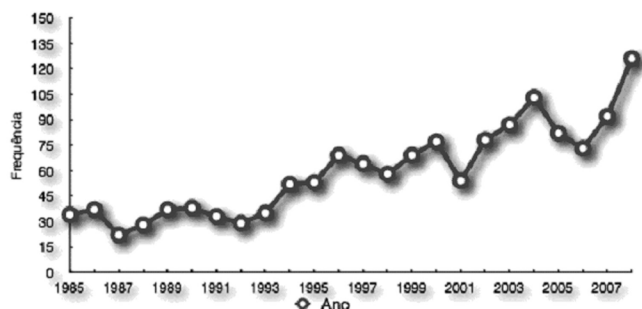


Figura 2. Número de artigos em periódicos, por ano, cujas palavras-chave eram 'avaliação' ou 'testes'.

teza um marco histórico muito importante desse período foi a criação do Sistema de Avaliação dos Testes Psicológicos (SATEPSI)⁷, do Conselho Federal de Psicologia (CFP), ao final de 2001. Esse sistema é resultado de uma série de ações anteriores do CFP, na tentativa de responder a uma grande demanda de processos éticos envolvendo a avaliação psicológica. O SATEPSI consiste em uma norma de certificação de instrumentos de avaliação psicológica que avalia e qualifica os instrumentos em apto ou inapto para uso profissional, a partir da verificação objetiva de um conjunto de requisitos técnicos mínimos (fundamentação teórica, precisão, validade e normatização), definidos pela área (*American Educational Research Association, American Psychological Association, Nacional Council on Measurement in Education, 1999; CFP, 2004; Nunes & Primi, 2010; Primi, Muniz & Nunes, 2009; Primi, Nascimento & Souza, 2004; Primi & Nunes, 2010*). Esse sistema é gerido por uma comissão consultiva em avaliação psicológica mantida pelo CFP e por um grupo de pareceristas composto por pesquisadores e profissionais da área. A comissão consultiva também auxilia a plenária do CFP a responder questionamentos da sociedade, por meio do oferecimento de consultoria técnica sobre a área.

O SATEPSI tem como meta a elevação da qualidade dos instrumentos de avaliação psicológica, uma vez que já havia sido notado que inúmeros testes, utilizados na prática profissional, não eram baseados em nenhum estudo que comprovasse seus fundamentos científicos (Noronha, Primi & Alchirei, 2004; Pereira, Primi & Cobêro, 2003). Isso configurava uma situação de um mal velado à população, difícil de ser descoberto, frente ao qual ela não possuía mecanismos para se defender (Primi & Nunes, 2010).

No cadastro do SATPSI, em 2004, havia 106 testes avaliados, sendo 51 desfavoráveis (48,1%). Em 2010, há 214 testes, sendo 77 desfavoráveis (35,9%), 114 favoráveis (53,2%) e 23 em processo de análise (10,7%). Portanto, o que se pode ver é que, nos últimos cinco anos, o número de testes praticamente dobrou. Essa medida provocou uma série de efeitos indiretos na área, tais como o aumento do investimento no desenvolvimento de instrumentos, o aquecimento do mercado de testes, uma divulgação de princípios técnicos da qualidade dos testes entre os profissionais e estudantes, de forma que o SATEPSI tornou-se uma referência para se lidar com outros setores da sociedade, consumidores da avaliação psicológica, tais como a procuradoria da justiça, polícia, dentre outros. Toda essa movimentação tem atraído a atenção internacional de organizações congêneres como a *American Psychological Association (APA) – Division 5* e a *International Test Commission (ITC)*, por exemplo (Hutz & Primi, 2006; Nunes, Hutz & Nunes, 2010).

É importante destacar que embora existam sistemas de revisão de testes como o *Buros Institute of Mental Measurements – BIMM*⁸, que já funciona há mais de 70 anos nos Estados Unidos, o Brasil é pioneiro na implementação de um sistema de certificação baseado em critérios internacionais de qualidade de testes, de forma a abarcar todos os instrumentos usados profissionalmente em um país. Medidas similares estão sendo feitas, por exemplo, pela *European Federation of Psychologists' Associations – EFPA*, que preparou uma ficha de avaliação dos testes psicológicos em 2008⁹, a qual vem sendo utilizada por uma editora espanhola na avaliação de seus testes. É interessante notar que os sistemas estrangeiros, como o *BIMM*, sistematizam informações sobre revisão de instrumentos e as disponibilizam à comunidade de profissionais e pesquisadores para que esses possam tomar decisões mais eficazes ao escolherem instrumentos. Embora o SATEPSI, atualmente, informe somente um parecer com a decisão final (apto ou inapto para uso), ele está sendo organizado para fornecer informações mais detalhadas sobre os testes em um banco de dados informatizado, disponibilizado aos profissionais pelo CFP nos moldes dos sistemas internacionais (Primi & Nunes, 2010).

De maneira geral, pode-se argumentar que esses indicadores atestam uma avanço importante na área nos últimos 10 anos. O aumento nas publicações e nos instrumentos aprovados no SATEPSI indica um aumento de produtos qualificados, visto que ambos os sistemas envolvem avaliação por pares. Pode-se inferir indiretamente que há um avanço no domínio de metodologias psicométricas, uma vez que essa constitui-se como condição necessária para o desenvolvimento de instrumentos e realização de estudos na área. No contexto acadêmico, nota-se um avanço na organização dos profissionais e pesquisadores e de disponibilização de informações qualificadas em bases de dados de revistas científicas, o que se tornam recursos valiosos para a prática profissional. Contudo, é difícil saber o quanto esses avanços estão refletindo na prática profissional, ou seja, o quanto o psicólogo está consumindo criticamente esse conhecimento

7 <http://www2.pol.org.br/satepsi/sistema/admin.cfm>.

8 <http://www.unl.edu/buros/bimm/index.html>.

9 <http://www.efpa.eu/professional-development/tests-and-testing>.

e transformando sua prática. Mas um avanço já é notável: o aumento da produção de recursos e informações qualificadas.

Direções Futuras

Diante do crescimento que se observa nos últimos anos, podemos nos questionar quais caminhos seriam importantes trilharmos para desenvolver ainda mais a área no Brasil. Tal reflexão foi feita em três áreas amplas: (a) avanços metodológicos e tecnológicos, (b) integração de abordagens e avanços dos seus métodos, (c) validade consequencial e relevância social e (d) incentivo à formação e criação da especialidade em avaliação psicológica.

Em relação à dimensão metodológica, como foi observado, nos últimos anos pode ser notado um avanço importante. Já existe um número considerável de artigos fazendo uso de métodos mais avançados como, por exemplo, a Teoria de Resposta ao Item (TRI) e Análise Fatorial Confirmatória. Em termos de avanços tecnológicos também encontram-se exemplos de testagem informatizada e confecção de relatórios via *web* já disponíveis para testes comercializados. Entretanto, esses avanços representam ainda o início de uma nova fase que se deve buscar atingir. Com os recursos tecnológicos disponíveis, não se torna mais necessário realizar tarefas mecânicas de maneira manual (cotação, correção, conversão de escores etc.). Uma série de procedimentos envolvidos na avaliação pode ser auxiliada pelo computador, nas mais variadas maneiras, tais como a aplicação em si, a correção e auxílio na confecção de relatórios. O avanço tecnológico abre uma gama de possibilidades para a avaliação psicológica como a aplicação de testes via *web*, a utilização de diferentes estímulos multimídia, uma riqueza maior na interação do sujeito com os instrumentos e coleta de novas informações não disponíveis em testes tradicionais (ver, por exemplo, Miguel & Primi, 2010; Muniz, Seabra, Primi & Miguel, 2010; Primi, 1998, 2002; Primi, Cruz, Muniz & Petrini, 2006; Santos & Primi, 2005).

A junção da tecnologia com as metodologias da TRI permite a criação de testes adaptativos computadorizados, os quais utilizam bancos de itens que permitem a otimização do teste por meio da seleção de itens mais adequados àquela avaliação, de forma a adaptar o teste ao sujeito, em razão da dificuldade das questões, por exemplo, permitindo dessa forma uma avaliação mais curta e mais precisa (Nunes & Primi, 2009). Evidentemente que todo esse avanço depende do desenvolvimento de *softwares* e ferramentas tecnológicas aliado à implementação de modelos psicométricos da TRI. Considerando-se que a comunidade científica brasileira já demonstra possuir domínio dessas ferramentas, espera-se, portanto, ver o desenvolvimento da área caminhar para o uso mais disseminado da avaliação informatizada baseada no emprego da TRI.

Um outro avanço metodológico importante embasa-se na utilização da TRI como meio para o aprimoramento da interpretação de medidas (Primi, 2004). A interpretação dos escores de um teste é tradicionalmente conferida por meio de referência à norma, que compara os escores obtidos por um indivíduo com os escores obtidos por um grupo de referência (grupo normativo), indicando a posição relativa desse escore

frente ao grupo (Cronbach, 1996). Entretanto, esse tipo de informação descritiva não permite afirmar de maneira mais substancial quais características descrevem as pessoas nos diferentes níveis da escala ou entender o que determinado nível na escala significa em termos dos processos mentais do construto que o instrumento tenta avaliar. Esse problema tem sido referido como métrica arbitrária (Blanton & Jaccard, 2006). Recentemente, Embretson (2006) propôs integrar a psicologia cognitiva com a TRI, em um procedimento chamado Interpretação Referenciada no Item. Como a TRI modela a relação entre escala numérica (os diversos níveis do construto) e a resposta a cada item em particular, ela possibilita definir, para cada nível da escala, quais seriam os padrões de resposta esperados aos itens do teste. Assim, conhecendo melhor os itens, por exemplo, por meio de análises mais aprofundadas do conteúdo e do processo de resposta, pode-se elaborar interpretações fazendo-se referência a esses elementos, conferindo um sentido mais aprofundado à métrica da escala. Esse procedimento tem sido aplicado, por exemplo, em testes psicológicos que avaliam o raciocínio indutivo (Primi, 1998, 2000), na análise do Exame Nacional do Desempenho do Estudante (ENADE) de Psicologia (Primi, Hutz & Silva, submetido) e na interpretação da Bateria Fatorial da Personalidade (Primi, Carvalho, Miguel & Muniz, 2010). Esse tipo de aprimoramento é uma contribuição importante da psicologia para a construção de medidas em ciências humanas, trazendo seu conhecimento sobre validação de construtos e operacionalização de medidas (Wilson, 2005).

Pela perspectiva dos estilos de pensamento em avaliação, que culminam na ênfase nomotética/psicométrica *versus* idiográfica/clínica, nota-se que os pontos levantados anteriormente derivam de procedimentos mais psicométricos de refinamento de medidas das dimensões psicológicas. Mas há igual necessidade de aplicação de métodos de análise centrados no nível dos indivíduos ou de seus perfis. Para isso, é necessário focar-se um nível acima (dos indivíduos/perfis) do que aquele habitualmente estudado (dos itens/variáveis latentes). Ao se analisar a literatura sobre avaliação, nota-se uma ênfase mais psicométrica, objetivada na análise fatorial de variáveis, por exemplo, buscando-se refinar as dimensões que, a despeito de sua grande importância, não são suficientes para esclarecer como o conjunto de dimensões avaliadas em um teste, quando combinado em um perfil, irá conseguir descrever os diferentes subtipos existentes. Depois de se validar as dimensões, provando sua estrutura interna e consistência, seguem-se outras questões (e.g., Quais perfis ou tipos de pessoas essas dimensões descrevem? Como devem ser interpretados?), as quais são respondidas por meio de métodos que buscam analisar perfis e agrupar pessoas (ver exemplos em Ackerman & Beier, 2003; Almeida & Primi 2004; Westen & Shedler, 1999a, 1999b). Contudo, mesmo diante da importância desse tipo de estudo, o que se pode notar é uma escassez de pesquisas que façam uso desse enfoque. Portanto, nota-se a necessidade de desenvolvimento de investigações que utilizem a análise de conglomerados ou a metodologia Q, que consiste em uma análise fatorial que busca agrupar pessoas (linhas da matriz), ao invés da habitual análise de variáveis (colunas da matriz). Tais métodos buscam descobrir agrupamento de pessoas com perfis semelhantes, de forma a se poder analisar como esses grupos são descritos

a partir de sua configuração nos construtos avaliados (Block, 2008; ver também <http://qmethod.org/> para mais detalhes da metodologia Q).

Ainda com relação aos avanços metodológicos, uma área que seguramente pode se desenvolver bastante com o emprego de métodos mais avançados é a dos métodos projetivos. Há uma tradição antiga de trabalhos nessa área no Brasil, sendo que, durante muito tempo, predominou, em relação a esse método, a disseminação de uma ideia errônea de que não seria possível aplicar métodos psicométricos clássicos na validação dessas técnicas. Por ocasião da criação e implementação do SATEPSI, essa discussão foi aprofundada e felizmente se construiu um consenso de que os princípios de validação psicométricos são adequados para todos os tipos de técnicas, incluindo as projetivas. Mas o ponto central situa-se na constatação de que essas técnicas têm peculiaridades que requerem métodos psicométricos mais avançados, diferentes dos habitualmente empregados. Como exemplo, podemos citar as discussões que são feitas na análise de validade convergente entre Rorschach e MMPI (Meyer, 1999) ou na análise fatorial das variáveis no Zulliger (Villemor-Amaral & Primi, 2009). A metodologia de análise de perfis anteriormente discutida pode também ser útil para o desenvolvimento de pesquisas com as técnicas projetivas. Como esses métodos são centrados na análise multivariada de semelhanças sistemáticas entre pessoas, eles são mais próximos do raciocínio clínico também subjacente a essas técnicas. Contudo, o ponto principal a ser destacado é o fato que essa área necessita de um avanço no conhecimento sobre metodologia e métodos quantitativos para se desenvolver.

É muito importante para a área que isso ocorra, visto que a literatura tem mostrado a riqueza das informações que podem ser obtidas quando se utilizam métodos de avaliação diversificados. Meyer e cols. (2001), por exemplo, reuniram estudos de meta-análise sobre a validade de testes psicológicos comparados aos testes médicos e concluíram que ambos apresentam validade comparável, e que o emprego de métodos diversificados pode trazer informações únicas. Sugerem, ainda, que as avaliações que usam múltiplos métodos são mais eficazes e que clínicos que fazem avaliações exclusivamente com entrevistas e observação são propensos a entendimentos incompletos. Em outro estudo muito interessante, Shedler, Mayman e Manis (1993) demonstraram que um grupo de pessoas que se relatavam saudáveis em escalas de autorrelato, na verdade, não podiam ser assim classificadas quando foram avaliadas em medidas psicofisiológicas de estresse. Entretanto, clínicos usando uma técnica projetiva conseguiram identificar essas pessoas pelo uso de mecanismos de negação de estresse. Esse estudo é um exemplo importante da utilidade dessas técnicas na avaliação.

Em um nível mais amplo, saindo das questões relativas aos itens-escalas e perfis de pessoas, e direcionando-se ao contexto em que os instrumentos são usados, veremos que várias consequências sociais se desdobram em função do uso que é feito desses instrumentos, suscitando questões importantes da avaliação que requerem investigações específicas. Essas questões apresentam uma grande relevância social e se relacionam diretamente com as críticas mais amplas à avaliação, embora não tenham sido sistematicamente estudadas. Essas questões foram colocadas por Messick (1980) ao

conceituar a validade como um julgamento avaliativo amplo da adequação e apropriabilidade das inferências derivadas dos escores dos testes em quatro facetas progressivas: (a) base evidencial da interpretação, que indica um sumário indutivo de evidências empíricas que apoiam o sentido do construto que está sendo atribuído aos escores do teste; (b) base evidencial do uso, que indica o racional e evidências da relevância do construto e utilidade dos escores para aplicações específicas; (c) base consequencial da interpretação, que indica uma apreciação das implicações de valor das interpretações dos escores do teste e sua coerência com o sentido do construto que está sendo atribuído; e (d) base consequencial do uso, indicando a apreciação das consequências sociais potenciais do uso pretendidos para o teste, comparada às consequências observadas de fato quando usado.

As informações sobre base evidencial são aquelas decorrentes dos estudos clássicos de validade. Já os estudos de base consequencial investigam questões mais amplas como, por exemplo, se a avaliação cognitiva no diagnóstico de crianças com dificuldades de aprendizagem de fato contribui para intervenções mais eficazes e promoção do desenvolvimento dessas crianças; se a avaliação psicológica em processos seletivos contribui para produzir uma organização mais eficaz e competente; se determinadas interpretações dos testes têm uma conotação negativa inadequada, gerando prejuízo não intencional às pessoas avaliadas. Em suma, tais estudos versam sobre as consequências do uso da avaliação psicológica e coesão desse uso com os propósitos mais gerais de desenvolvimento das pessoas. Em essência, essas questões equacionam o compromisso social da psicologia no âmbito da avaliação, indagando se suas técnicas estão sendo úteis e eticamente apropriadas aos setores sociais que consomem seus serviços. Como foi salientado, embora existam muitas críticas dentro da psicologia direcionadas à avaliação psicológica em relação a esses aspectos, poucas ou nenhuma dessas críticas têm se transformado em pesquisas que buscam verificar elementos da validade consequencial. Então, a transformação dessas críticas em estudos bem delineados de validade consequencial se torna um passo importante na direção do aprimoramento da área.

Por último, há a questão da formação profissional, que é um problema geral, não só restrito à psicologia ou à avaliação. No entanto, como o avanço da área depende de competências mais complexas, tais como raciocínio matemático, conhecimentos de estatística e de métodos de pesquisa, a área acaba sendo bastante afetada pela má formação. Em um estudo dos dados do ENADE, Primi e cols. (submetido) mostram que poucos estudantes que se formam apresentam competências mais complexas ligadas à compreensão de dados estatísticos e do raciocínio crítico ligado à investigação científica. Portanto, o avanço da área requer uma melhora substancial nessa formação específica e em medidas que de fato promovam a integração entre pesquisa e prática profissional. O uso dos recursos tecnológicos, por um lado, pode ser um aliado importante para a formação na área. Por outro, é necessária uma maior qualificação da crítica em relação à avaliação psicológica de forma a dissolver falsos dilemas e evitar a perpetuação de preconceitos simplistas de que a avaliação é uma atividade técnica alienada e descompromissada socialmente,

fato que acaba desmotivando jovens estudantes a seguir se aprofundando na área.

Outro ponto importante relacionado à formação é a criação da especialidade em avaliação psicológica. Tal pedido já foi solicitado ao CFP, mas até o momento não foi aceito, com o argumento que a avaliação psicológica não corresponde a uma atividade profissional específica, mas sim geral, que perpassa todas as atividades do psicólogo nas mais diversas áreas. Ainda que esse ponto seja verdadeiro, a criação da especialidade não tem o intuito de proibir essa prática aos não especialistas, já que nenhum psicólogo é impedido de exercer a profissão pelo simples fato de não possuir um título de especialista; mas sim reconhecer que existe um campo próprio de conhecimentos específicos, como se argumenta ao longo de todo esse texto, e reconhecer o mérito dos profissionais que eventualmente venham aprofundar seus estudos nas diversas áreas desse campo. Esse fato poderá estimular a criação de cursos específicos e outras medidas que podem melhorar a formação na área.

Talvez a resistência em relação à especialidade em avaliação psicológica seja oriunda de um setor da psicologia que faz uso de uma concepção diferente de avaliação, que não é tão sistematizada e não usa instrumentos, confiando mais em procedimentos livres como entrevista e observação. Assim, a aprovação da especialidade se torna implicitamente um reconhecimento do valor do uso de instrumentos, que não é uma prática valorizada nesse grupo. Talvez por esse motivo haja tanta resistência. No entanto, como já se afirmou, a literatura tem apontado que as práticas mais competentes e eticamente comprometidas são aquelas que se utilizam de todos os recursos disponíveis (abordagem multimétodo com instrumentos diversos e outros procedimentos) e em todos os níveis importantes (testes, pessoas e contexto). Portanto, se for esse o caso, torna-se necessário um avanço em direção a esse entendimento para que a especialidade seja aceita. E quando isso ocorrer será um sinal que a psicologia deu mais um passo importante em direção ao desenvolvimento da área.

Referências

- Ackerman, P. L., & Beier, M. F. (2003). Intelligence, personality and interests in the career choice process. *Journal of Career Assessment, 11*, 205-218.
- Almeida, L. S., & Primi, R. (2004). Perfis de capacidades cognitivas na Bateria de Provas de Raciocínio (BPR-5). *Psicologia Escolar e Educacional, 8*, 135-144.
- 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.
- Anastasi, A., & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artes Médicas.
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27-41.
- Block, J. (2008). *The Q-sort in character appraisal: Encoding subjective impressions of persons quantitatively*. Washington, DC: American Psychological Association.
- Borsboom, D., Mellenbergh, G. J., & Heerden, J. V. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book.
- Cattell, R. B. (1965). *Análise científica da personalidade*. São Paulo: Ibrasa.
- Cattell, R. B. (1973). Personality structure: Principles common to Q-, L- and T-Data. Em R. B. Cattell (1973), *Personality and mood by questionnaire: A handbook of interpretative theory, psychometrics and practical procedures* (pp. 1-63). San Francisco: Jossey-Bass.
- Conselho Federal de Psicologia – CFP (2001). *Resolução Nº 002/2003*. Brasília: CFP.
- Conselho Federal de Psicologia – CFP (2004). *Avaliação dos testes psicológicos: relatório*. Brasília: CFP.
- Cronbach, L. J. (1996). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist, 61*, 50-55.
- Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology, 50*, 183-195.
- Hutz, C. S., & Primi, R. (2006). Psychological assessment in Brazil: Current trends and challenges. *The Score (Newsletter APA), 37*, 17-18.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, 1012-1027.
- Meyer, G. J. (1999). The convergent validity of MMPI and Rorschach scales: An extension using profile scores to define response/character styles on both methods and a re-examination of simple Rorschach response frequency. *Journal of Personality Assessment, 72*, 1-35.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, Robert R., Eisman, E. J., Kubiszyn, T. W., & Read, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128-165.
- Miguel, F. K., & Primi, R. (2010). Teste Informatizado de Percepção de Emoções Primárias. Em M. C. R. A. Joly & C. Reppold (Orgs.), *Estudos de testes informatizados para avaliação psicológica*. São Paulo: Casa do Psicólogo.
- Millon, T., Grossman, S., Millon, C., Meagher, S., & Ramnath, R. (2004). *Personality disorders in modern life*. New Jersey: John Wiley & Sons.
- Muniz, J. (2004). La validación de los tests. *Metodología de las Ciencias del Comportamiento, 5*, 121-141.
- Muniz, M., Seabra, A. G., Primi, R., & Miguel, F. K. (2010). Teste Dinâmico Informatizado para avaliar o raciocínio indutivo em crianças-TEDRI. Em M. C. R. A. Joly & C. Reppold (Orgs.), *Estudos de testes informatizados para avaliação psicológica*. São Paulo: Casa do Psicólogo.
- Noronha, A. P. P., Primi, R., & Alchirei, J. C. (2004). Parâmetros psicométricos: uma análise de testes psicológicos comercializados no Brasil. *Psicologia Ciência e Profissão, 24*, 88-89.
- Nunes, C. H. S. S., Hutz, C. S., & Nunes, M. F. O. (2010). Brazilian Psychological Assessment Association: Objectives, actions, and projects. *Testing International (ITC), 23*, 4.
- Nunes, C. H. S. S., & Primi, R. (2009). Teoria de resposta a item: conceitos e aplicações na Psicologia e na Educação. Em C. S. Hutz (Org.), *Avanços e polêmicas em avaliação psicológica* (pp. 25-69). São Paulo: Casa do Psicólogo.

Nunes, C. H. S. S., & Primi, R. (2010). Aspectos técnicos e conceituais da ficha de avaliação dos testes psicológicos. Em Conselho Federal de Psicologia - CFP (Org.), *Avaliação psicológica: diretrizes na regulamentação da profissão* (pp. 101-128). Brasília: CFP.

Pereira, F. M., Primi, R., & Cobêro, C. (2003). Validade de testes utilizados em seleção de pessoal segundo recrutadores. *Psicologia: Teoria e Prática*, 5, 83-98.

Primi, R. (1998). *Desenvolvimento de um instrumento informatizado para avaliação do raciocínio analítico*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo.

Primi, R. (2000). Testes informatizados de inteligência: dos produtos aos processos cognitivos. Em M. J. Gonçalves, E. C. Macedo, A. L. Sannyey & F. C. Capovilla (Orgs.), *Tecnologia em (re)habilitação cognitiva: a dinâmica clínica-teoria-pesquisa* (pp. 30-37). São Paulo: Centro Universitário São Camilo.

Primi, R. (2002). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, 30, 41-70.

Primi, R. (2003). Inteligência: avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação Psicológica*, 2, 67-77.

Primi, R. (2004). Avanços na interpretação de escalas com a aplicação da Teoria de Resposta ao Item. *Avaliação Psicológica*, 3, 53-58.

Primi, R., Carvalho, L. F., Miguel, F. K., & Muniz, M. (2010). Resultado dos fatores da BFP por meio da Teoria de Resposta ao Item: interpretação referenciada no item. Em C. H. S. S. Nunes, C. Hutz & M. F. O. Nunes (Org.), *Bateria Fatorial de Personalidade (BFP): manual técnico*. (pp. 153-170). São Paulo: Casa do Psicólogo.

Primi, R., Cruz, M. B. Z., Muniz, M., & Petrini, M. C. (2006). Validade de construto de um instrumento informatizado de avaliação dinâmica da inteligência fluida. *Psico (PUCRS)*, 37, 109-122.

Primi, R., Hutz, C. S., & Silva, M. C. R. (submetido). A prova do ENADE de Psicologia 2006: concepção, construção e análise da prova.

Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições contemporâneas de validade de testes psicológicos. Em C. S. Hutz (Org.), *Avanços e polêmicas em avaliação psicológica* (pp. 243-265). São Paulo: Casa do Psicólogo.

Primi, R., Nascimento, R. S. G. F., & Souza, A. S. (2004). Critérios para avaliação de testes psicológicos. Em Conselho Federal de Psicologia - CFP (Org.), *Avaliação dos testes psicológicos: relatório* (pp. 31-55). Brasília: CFP.

Primi, R., & Nunes, C. H. S. S. (2010). O SATEPSI: propostas de aprimoramento. Em Conselho Federal de Psicologia - CFP (Org.), *Avaliação psicológica: diretrizes na regulamentação da profissão* (pp. 129-148). Brasília: CFP.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, 65, 1-12.

Santos, M. A., & Primi, R. (2005). Desenvolvimento de um teste informatizado para a avaliação do raciocínio, da memória e da velocidade de processamento. *Estudos de Psicologia (Campinas)*, 22, 241-254.

Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist*, 48, 1117-1131.

Tavares, M. (2003). Validade clínica. *Psico-USF*, 8, 125-136.

Villemor-Amaral, A. E., & Primi, R. (2009). *O Teste de Zulliger no Sistema Compreensivo ZSC Forma Individual*. São Paulo: Casa do Psicólogo.

Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry*, 156, 258-272.

Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part II: Toward an empirically based clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273-285.

Wilson, M. (2005). *Constructing measures: A item response modeling approach*. New Jersey: Lawrence Erlbaum.

Wright, B. D. (1999). Fundamental measurement for Psychology. Em S. E. Embretson & S. L. Hershberger (Orgs.), *The new rules of measurement: What every psychologist and educator should know* (pp. 65-103). New Jersey: Lawrence Erlbaum.

Wright, B. D., & Stone, M. H. (2004). *Making measures*. Chicago: The Phaneron Press.