

Definições contemporâneas de validade de testes psicológicos

Ricardo Primi, Monalisa Muniz & Carlos Henrique Sancineto Nunes

Programa de Pós Graduação Stricto Sensu em Psicologia, Universidade São Francisco

Correspondências devem ser enviadas para:

Ricardo Primi, Universidade São Francisco, Laboratório de Avaliação Psicológica e Educacional (LabAPE), Mestrado em Psicologia, Rua Alexandre Rodrigues Barbosa, 45, CEP 13251-900, Itatiba, São Paulo, Fone (0XX11) 45348118, correio eletrônico: ricardo.primi@saofrancisco.com.br ou rprimi@uol.com.br.

As atividades de pesquisa dos autores desse capítulo são financiadas pelo CNPq, FAPESP e CAPES.

A avaliação psicológica figura entre uma das mais importantes competências profissionais do psicólogo, que envolve a aplicação de conhecimentos teóricos no entendimento do funcionamento psicológico de pessoas e grupos relacionados a uma demanda específica de entendimento e previsão do comportamento. Em tais situações, os profissionais utilizam instrumentos de avaliação que são procedimentos sistemáticos de observação de amostras de comportamento, que têm como objetivo descrever e/ou mensurar características e processos psicológicos (Anastasi e Urbina, 1997 Cronbach, 1996).

Um ponto fundamental relativo à cientificidade desses instrumentos é tratado no conceito de validade. Historicamente validade tem sido definida como o alcance em que o teste mensura o que se propõe medir ou ao conhecimento do que o teste mede e de quão bem ele faz isso (Anastasi & Urbina, 1997). Ainda nos *Standards for Educational and Psychological Testing* (chamado daqui em diante de *Standards*) de 1986, validade foi definido como “o grau em que as evidências embasam as inferências feitas a partir dos escores dos testes” (*American Educational Association, American Psychological Association, National Council on Measurement in Education, AERA, APA NCME, 1986 p. 9*).

De maneira geral, os instrumentos de avaliação propõem tarefas específicas às pessoas - resolver problemas, concordar ou não com frases auto descritivas, desenhar, contar histórias, entre outras - a partir das quais os profissionais observam e registram seus comportamentos (desenhos, histórias, acertos) e por meio deles inferem a presença de características psicológicas específicas. Na base desse processo, há a suposição de que as distintas maneiras como as pessoas respondem aos testes (comportamentos) são determinadas pelas suas características psicológicas (traços latentes) e por isso as respostas dadas aos testes servem como fonte de conhecimento dessas características. Nesse contexto, a questão da validade é o ponto basilar relativo aos estudos que desafiam e corroboram ou não as associações entre os comportamentos na testagem e os traços latentes inferidos por meio destes. O resumo dos resultados positivos dessas pesquisas embasando a relação comportamento-traço caracterizam a validade dos instrumentos. Referem-se, portanto, à legitimidade das interpretações que são feitas, portanto, ao conjunto de evidências favoráveis às interpretações propostas obtidas em pesquisas destinadas a testar os pressupostos de tais interpretações.

Se um instrumento não possui evidências de validade, não há segurança de que as interpretações sobre as características psicológicas das pessoas sugeridas pelas suas repostas na testagem são legítimas. Nessa situação, não se tem certeza sobre o instrumento psicológico efetivamente avalia e quão bem o faz e, portanto, sua utilização em práticas profissionais é ética e tecnicamente impedida.

Como argumenta Muniz (2004), de uma perspectiva ampla, o processo de validação de instrumentos não é diferente do processo científico de validação de hipóteses. Assim a validação de testes é um caso específico do processo geral de validação de teorias psicológicas por meio de formulação de hipóteses, previsões derivadas da teoria sobre manifestações comportamentais e verificação empírica da correspondência entre expectativas teóricas e os fatos observados. Assim, além de serem ferramentas profissionais, os instrumentos são objetivações de teorias psicológicas e, por isso, tem um papel fundamental não só na prática profissional mas também no avanço do conhecimento da Psicologia.

Assim, a validade era operacionalizada por meio de estudos classificados em três tipos, chamados por Muniz (2004) de “santíssima trindade da validade”: validade de conteúdo, critério e construto. O primeiro tipo, validade de conteúdo, apareceu com o objetivo de avaliar o conteúdo dos itens de determinado instrumento, verificando se são adequados para representar um domínio de comportamentos a serem mensurados. Esse tipo de validade é freqüentemente utilizado no contexto educacional, no qual é usual a verificação se os itens que compõem um teste para mensurar desempenho em uma área específica de conhecimento cobrem todos os conteúdos englobados por tal área.

O segundo tipo, validade de critério, relaciona-se à eficácia com que teste prediz variáveis externas ou variáveis critério como, por exemplo, desempenho profissional, notas escolares, acidentes de trabalho, diagnóstico psiquiátrico, dentre outros. Tais variáveis são importantes em si mesmas e se relacionam ao construto avaliado pelo instrumento de forma que a verificação de associação teste-critério atesta a utilidade desses em prever eventos objetivos relevantes em contextos aplicados. Esse tipo de validade surge em uma época de forte influência do empiricismo na Psicologia. Estudos de validade de critério são essenciais para instrumentos cujos resultados são usados como fonte de informação para inferências complexas e usualmente com alto impacto, como é o caso de testes para avaliação psicológica no contexto clínico, forense, organizacional, entre outros.

O terceiro tipo de validade, denominada validade de construto, foi definida como a medida que um teste mede determinado construto ou traço, ou seja, em que medida as evidências apóiam os significados atribuídos aos escores do teste. Essa última aparece em razão de uma crescente importância das formulações teóricas sobre o comportamento humano dentro da Psicologia (Anastasi & Urbina, 1997; Cronbach, & Meehl, 1955).

Com o tempo essa definição tripartite foi questionada e aprimorada especialmente pelas contribuições de Messick (1986). Um dos principais questionamentos levantados por esse autor dizia respeito à definição de validade de construto. Messick argumenta que tanto a validade de conteúdo quanto a de critério sempre apresentarão informações referentes ao construto, no sentido que levantarão evidências sobre o sentido das interpretações dos escores dos testes embora, evidentemente, de natureza distinta. A primeira é associada à representatividade, relevância e abrangência das unidades que compõem o teste em relação ao construto que o fundamenta; e a segunda aborda questões sobre a utilidade prática na previsão de variáveis externas relevantes. Assim Messick destaca que quase toda informação sobre o teste irá contribuir para sua validade de construto e, por esse motivo, todo estudo de validade, em certo sentido, é uma verificação da validade de construto já que apresentará peças de informação sobre como interpretar os escores de um teste.

Em razão dessa discussão, foram propostas reformulações do conceito de validade de construto reconhecendo-o como sinônimo de validade. Assim, este passou a ser entendido como um conceito abrangente que inclui sob si todas as outras formas de validade, sendo, portanto, inadequado sustentar a existência de procedimentos específicos para validade de construto. Em última análise, todos os métodos, de uma maneira ou de outra, informam algo sobre como interpretar os escores do teste, ou seja, agregam algo específico sobre quão bem o teste mensura determinado construto.

Na sequência das reformulações propostas, a expressão “tipos de validade” foi intencionalmente mudada para “fontes de evidência” de validade. A alteração visou reforçar a conotação do termo como um conceito único, que explicita que as diferentes formas de estudá-lo (conteúdo, correlações teste-critério, etc.) trazem dados para sustentar a argumentação a respeito da validade das interpretações dos escores do teste com uma finalidade particular. Neste sentido, cada evidência específica busca aspectos diferentes de um ponto conceitualmente único sobre como interpretar os escores de um

teste. Assim, na última revisão dos *Standards* não há mais um tipo específico de procedimento chamado validade de construto e os antigos tipos de validade passaram a ser definidos como fontes de evidências de validade como se verá mais adiante (Anastasi & Urbina, 1997; APA, AERA, NCME, 1999).

Messick (1989) define validade como a medida em que as evidências empíricas e os racionais teóricos, ou o sumário indutivo, embasam os significados interpretativos e os usos propostos para o teste. Essa definição associa validade a uma sindicância científica que averigua os sentidos atribuídos aos escores do teste e também as conseqüências atuais ou potenciais do uso interpretativo dos escores. Nessa proposição há quatro questões inter-relacionadas relativas à quatro facetas progressivas da validade sumarizadas na Tabela 1: (a) Quais evidências (balança de evidências) suportam a interpretação ou o sentido dos escores (validade de construto)?, (b) Quais evidências suportam a relevância dos escores para o propósito aplicado (validade de construto + relevância/utilidade)?, (c) Quais razões fundamentais tornam credíveis as implicações de valor das interpretações dos escores e qualquer uma das implicações para ação ou tomada de decisão associadas a estas interpretações (validade de construto + implicações de valor)? e (d) Quais evidências e argumentos irão indicar o valor funcional em termos de usos intencionais e conseqüências não intencionais das interpretações do teste (validade de construto + relevância/utilidade+ implicações de valor + conseqüências sociais)?

Tabela 1. Quatro Facetas Progressivas da Validade definidas por Messick (1989)

	Interpretação do teste	Uso do teste
Base evidencial	Validade de Construto (VC)	VC + Relevância e utilidade (R e U)
Base conseqüencial	VC + Implicação de Valor (IV)	VC + R + U + IV + Conseqüências Sociais

Messick (1989) enfatiza que a validade das inferências diz respeito a cada um dos significados ou interpretações dos escores e não ao teste ou o dispositivo observacional em si. Sua definição amplia a noção de significado dos escores incluindo as valorações positivas e negativas implicadas nas interpretações uma vez que esses sentidos são socialmente relevantes ao disparar ações e decisões a partir dos escores do

teste sendo parte integrante de seu sentido. Para exemplificar essa idéia pode-se pensar que um teste de *inibição vs impulsividade* seria entendido diferentemente e levaria a ações distintas se essa dimensão fosse renomeada para *auto-controle vs auto-expressão*. Essa mudança no sentido atribuído ao escore do teste não está circunscrita ao instrumento em si, mas às valorações implicadas nas nomeações do construto que, em um sentido amplo, tem a ver com a maneira como o escore é interpretado e, portanto, relaciona-se à validade, embora não esteja localizada no “objeto concreto” instrumento e sim, mas em local contextual mais abstrato onde ele é utilizado. O autor argumenta, portanto, que na interpretação dos escores, além da base evidencial dos seus significados, há também, de maneira mais ampla, uma base consequencial das interpretações assentada nas valorações implicadas. Assim, além da dimensão sentido dos escores há uma dimensão valor dos escores que é socialmente implícita e altera a dimensão sentido.

Ainda ao considerar a faceta referente ao uso, a base evidencial diz respeito à relevância e utilidade do teste em um contexto aplicado e a base consequencial refere-se às decorrências potenciais e/ou reais intencionais e não intencionais do uso dos testes. Essa concepção amplia a extensão dos estudos de validade para o contexto social macroscópico no qual o teste está inserido. Nesse contexto busca-se analisar as implicações antecipadas ou inesperadas da utilização do teste para ver se, de fato, o teste efetivamente agrega informações relevantes sobre o funcionamento psicológico no contexto específico em que se pretende realizar uma avaliação.

Como exemplo, pode-se pensar em instrumento clínico de avaliação da personalidade. Estudos de base evidencial sobre sua utilidade incluirão pesquisas da relação teste-critério mostrando que o instrumento prevê uma série de variáveis relevantes como diagnóstico, severidade, tipos de transtornos de personalidade e de dificuldades mais salientes, por exemplo. Estudos de base consequencial movem-se das investigações sobre o teste para contextos mais amplos sobre o processo de avaliação no qual o instrumento está inserido. Nesse exemplo, esses estudos poderiam perguntar se o uso do teste em processos psicoterápicos agregam informação e tornam a psicoterapia mais efetiva do que se não fosse utilizado. Em relação a esse aspecto Meyer e cols. (2001) argumentam que: “os pesquisadores tem historicamente focado o micro nível para avaliar os parâmetros psicométricos de validade e precisão das escalas os quais estão divorciados de um contexto individualizado. Esse foco é certamente importante.

Entretanto, os pesquisadores deveriam também focar o macro nível para avaliar o valor prático de clínicos que usam testes como ferramentas para ajudá-los a oferecer suas consultas profissionais e beneficiar os pacientes” (p. 154). Em suma, essa sugestão de Meyer refere-se aos estudos de base consequencial da utilidade dos instrumentos para a avaliação.

As idéias de Messick (1986) tiveram grande influência quando ocorreu a revisão dos *Standards* de sua edição de 1985 para a mais atual versão de 1999 (*American Educational Research Association AERA, American Psychological Association APA, National Council on Measurement in Education NCME*, 1999). Neste novo documento, em consonância com as idéias discutidas acima, validade foi definida como se referindo “ao grau em que evidência e teoria sustentam as interpretações dos escores dos testes vinculados aos usos propostos dos testes” (p. 9). O processo de validação é cumulativo, agregando um conjunto de evidências científicas que asseguram as interpretações dos escores do teste e a relevância e utilidade dos usos propostos. Assim, mais uma vez enfatiza-se que nesse processo não se valida o teste em si, mas as interpretações propostas por ele e as aplicações práticas intentadas. Neste sentido, as alterações propostas têm grande impacto às pesquisas realizadas com base nas concepções anteriores de validade. Estas apontam à fragilidade da generalização de inferências elaboradas a partir de pesquisas realizadas em um contexto específico para outros muito diversos. Questiona-se, por exemplo, se fatores de risco para o desenvolvimento de um quadro de ansiedade identificados, com uso de testes, no contexto hospitalar em pessoas com doenças crônicas podem ser considerados os mesmos em um contexto organizacional e, com base nisso, tomar decisões referentes à contratação ou realocação de profissionais em uma empresa.

Esse novo documento define cinco fontes usadas ao evidenciar a validade de um teste (AERA, APA, NCME, 1999, ver sumário na Tabela 2): (a) evidência com base no conteúdo, (b) evidência com base no processo de resposta, (c) evidência com base na estrutura interna, (d) evidência com base na relação com variáveis externas e (e) evidências baseadas nas consequências da testagem.

Tabela 2. Fontes de evidências de validade

<i>Fonte</i>	<i>Definição</i>
--------------	------------------

Evidências baseadas no conteúdo	Levanta dados sobre a representatividade dos itens do teste investigando se eles consistem em amostras abrangentes do domínio que se pretende avaliar com o teste.
Evidências baseadas no processo de resposta.	Levanta dados sobre os processos mentais envolvidos na realização das tarefas propostas pelo teste
Evidências baseadas na estrutura interna	Levanta dados sobre a estrutura das correlações entre itens avaliando o mesmo construto e também sobre as correlações entre subtestes avaliando construtos similares.
Evidências baseadas nas relações com variáveis externas	Levanta dados sobre os padrões de correlação entre os escores do teste e outras variáveis medindo o mesmo construto ou construtos relacionados (convergência) e com variáveis medindo construtos diferentes (divergência). Também traz dados sobre a capacidade preditiva do teste de outros fatos de interesse direto (critérios externos) que possuem importância por si só e associam-se ao propósito direto do uso do teste (por exemplo, sucesso no trabalho).
Evidências baseadas nas consequências da testagem.	Examina as consequências sociais intencionais e não intencionais do uso do teste para verificar se sua utilização está surtindo os efeitos desejados de acordo com o propósito para o qual foi criado.

A primeira fonte de informação sobre validade busca evidências com base no conteúdo e refere-se aos estudos que coletam dados sobre a abrangência e representatividade dos itens de um teste em abarcar e todas as facetas e nuances do construto que o teste se propõe a medir. As pessoas que desenvolvem testes freqüentemente desenvolvem definições operacionais e constitutivas que especificam o conteúdo a ser abarcado no instrumento (Pasquali, 1999), descrevendo-o cuidadosamente em detalhes com uma classificação das áreas e tipos de itens, por exemplo. Essa evidência pode incluir uma análise lógica ou empírica pela qual se pretende verificar se o conteúdo do teste representa adequadamente o domínio em relação à teoria que especifica as facetas do construto e se há relevância deste para as interpretações propostas pelos escores do teste. Evidência embasada no conteúdo também pode ser obtida por meio de especialistas que julgam a relação entre as partes do teste com as facetas do construto.

A segunda fonte de informações sobre a validade busca evidências com base no processo de resposta e refere-se aos estudos que visam compreender os processos mentais subjacentes a produção de respostas (ou comportamentos) impelidas pelos itens

do instrumento. Essas informações são, em essência, especificações mais detalhadas do construto que o teste propõe medir. A partir do referencial teórico do construto elaboram-se modelos explicativos do processamento mental que ocorre ao responder os itens do teste. Elaboram-se também previsões sobre parâmetros do comportamento - acerto, tempo de reação, por exemplo - em relação a diferentes propriedades dos itens. O modelo teórico inclui explicações sobre como as diferentes características dos itens afetam o processamento e, especialmente, como essas alterações processuais culminam em alterações nos parâmetros comportamentais – ocorrência de acertos, maior ou menor tempo de reação, etc. Assim, o modelo teórico pode ser empiricamente testado comparando-se as observações dos padrões de resposta ou os padrões dos parâmetros comportamentais em diferentes itens verificando-se o quanto esses padrões correspondem às previsões feitas pelo modelo. Quanto mais correspondentes mais confiança se tem no modelo teórico de interpretação sobre o que o teste mede.

Como comentam Anastasi e Urbina (1997), essa forma de validar instrumentos surge a partir da década de oitenta e tem sido principalmente aplicada aos testes de inteligência decorrentes da Psicologia Cognitiva e não é ainda freqüentemente usada. Os trabalhos de Embretson (1994) são pioneiros no desenvolvimento desse tipo de estudo definindo-o como representação de construto tendo como objetivo geral a identificação dos processos cognitivos básicos subjacentes à resolução ou resposta aos itens dos testes. No Brasil o início de aplicações desse tipo na construção de testes pode ser encontrado em Primi, (1995, 1998, 2002), Primi e Rosado (1995).

Outro método para o estudo de evidências de validade pelo processo de resposta é por meio do questionamento dos respondentes sobre suas estratégias de resolução dos itens. Inferências sobre os processos envolvidos no desempenho podem também ser desenvolvidas, analisando a relação entre as partes do teste ou o teste com outras variáveis. Esse procedimento de validação também pode incluir estudos empíricos de como observadores ou juízes registram e interpretam os dados a fim de verificarem se estão em consonância com o que pretendem medir, de acordo com a definição de construto.

Em suma, essa estratégia para verificação de validade refere-se “à análise teórica-empírica das relações entre os processos mentais ligados ao construto em causa e as respostas aos itens do instrumento, isto é, as propostas explicativas dos processos mentais subjacentes às respostas aos itens e à coerência entre as explicações e os dados

empíricos. Tais evidências podem, por exemplo, associar-se à verificação de hipóteses sobre o modo de operação dos processos mentais subjacentes ao construto em avaliação durante a realização do teste; à análise pormenorizada das verbalizações e/ou dos passos do indivíduo em resposta aos diferentes estímulos do teste...” (Conselho Federal de Psicologia, 2007, p. 133).

A terceira fonte de informações para se estudar a validade refere-se à análise da estrutura interna do instrumento. Este tipo de evidência de validade geralmente é obtido por meio da análise das estruturas de covariância entre partes do teste sejam elas itens ou subescalas compostas pela soma de itens. Geralmente esses estudos empregam a análise fatorial e/ou a análise da consistência interna para indicar o quanto a estrutura de relações entre os itens do teste e/ou seus componentes ou subtestes são coerentes com a estrutura proposta pela teoria. A fundamentação teórica do construto de um teste pode implicar em uma ou mais dimensões, também pode pressupor uma estrutura hierárquica ou de fatores isolados. Nesses casos, a busca por evidências com base na estrutura interna envolve a verificação da coesão entre a estrutura prevista com a observada. Mais recentemente, além dos métodos exploratórios tradicionais, têm sido utilizados recursos como análise fatorial confirmatória e modelagem com equações estruturais que permitem uma grande flexibilidade no teste confirmatório de modelos teóricos (Byrne, 2001).

A quarta fonte de informação chama-se evidência com base na relação com variáveis externas. Nesse tipo de estudo, em contraste com os estudos da estrutura interna, que analisam as relações entre as variáveis internas ao próprio teste, investigam-se as relações dos índices obtidos no teste com variáveis externas relevantes à validade do instrumento. De acordo com os *Standards* de 1999, as variáveis externas podem ser classificadas em quatro tipos: (a) variáveis critério (b) testes medindo os mesmos construtos, (c) testes medindo construtos relacionados (d) testes medindo construtos diferentes. A Figura 1 exemplifica essa categorização.

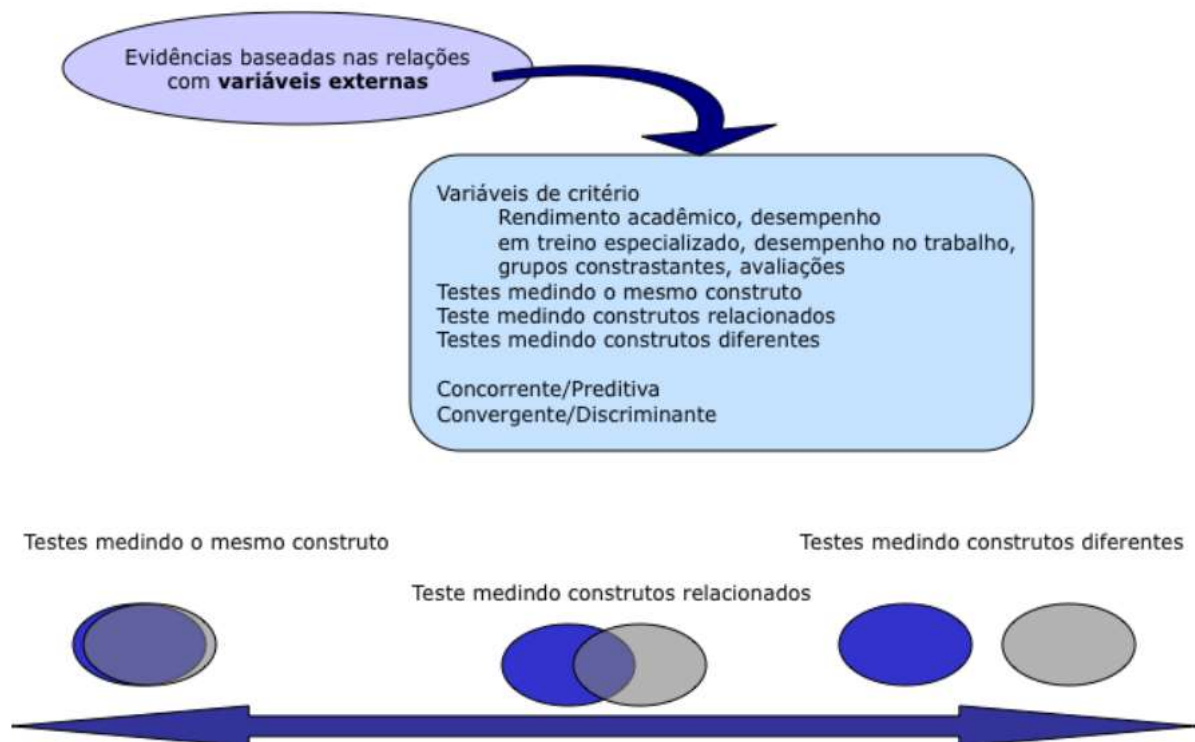


Figura 1. Classificação das variáveis externas no contexto de validação de testes

As variáveis critério, sejam elas contínuas ou dicotômicas indicativas de pertença a grupos, consistem em observações de eventos comportamentais relevantes por si só como, por exemplo, desempenho no trabalho, desempenho escolar, acidentes, adoecimento mental, escolha profissional, etc. Assim consistem em eventos importantes de serem previstos e que têm como um possível fator associado, o construto medido pelo teste. Assim embora não correspondam necessariamente a um processo mental e sejam resultantes de múltiplos fatores têm estreita relação com os construtos psicológicos medidos pelos testes, isto é, com os processos mentais que esses medem. Dessa forma, a observação de associações significativas entre teste-critério pode adjudicar validade ao instrumento, especialmente sobre a relevância e utilidade em prever certas situações importantes em ambientes aplicados específicos. Uma parte importante desses estudos refere-se à fundamentação teórico-empírica, baseada na revisão da literatura, justificando a associação antecipada entre construto-critério. Sem esta fundamentação, esses estudos envolveriam o teste de duas questões simultâneas, isto é, verificariam a associação construto-critério ao mesmo tempo em que estudam associação teste-critério. Num caso como esse, se não for encontrada relação significativa teste-critério, não é possível concluir se tal resultado decorre do fato do

teste não medir o construto como se pensava ou se o construto é que não se relaciona ao critério (Wolfe & Smith, 2007). Nesse último caso o teste poderia ser uma medida válida do construto e coerentemente não estar correlacionado com o critério. Portanto, a escolha do critério, bem como o procedimento de mensuração utilizado para obter as medidas do critério é um ponto central neste tipo de validade, pois o valor do estudo depende da relevância, precisão e validade dessas medidas de critério.

Nos estudos da relação teste-critério normalmente é feita uma distinção entre dois tipos de delineamento culminando em informações distintas. Quando a avaliação da variável critério é feita simultaneamente ao teste, a validade é chamada diagnóstica ou concorrente, pois, se as associações forem altas, atingiu-se conhecimento da situação presente. Quando a variável critério é avaliada após a aplicação do teste, ou seja, refere-se a um evento que se realizou após a sua aplicação, a validade é chamada preditiva já que uma alta associação entre o teste e o critério indica que o teste conseguiu prever uma situação futura.

Uma segunda categoria de variáveis externas utilizadas na validação de testes diz respeito a outros testes, isto é, instrumentos de avaliação de processos mentais similares, relacionados ou diferentes. Uma prática comum é trabalhar com outros instrumentos do gênero destinados a avaliar o mesmo construto tal qual o teste que está sendo validado. Nesse caso antecipam-se correlações altas uma vez que o testes medem um mesmo construto, chamando-se tais resultados, quando seguem o padrão esperado, evidência de validade convergente. O método que adota uma lógica oposta a esta é realizada com o uso de testes que mensuram construtos distintos ao do teste que está sendo validado. Nesse caso antecipam-se ausência de relação, isto é, correlações próximas de zero. Quando isso ocorre diz-se que se evidenciou a validade divergente do instrumento.

Na prática, são realizados estudos que buscam simultaneamente evidências de validade convergente e discriminante. Estudos desse tipo utilizam instrumentos de medida de construtos similares e muito pouco relacionados e são esperadas correlações altas com os primeiros e virtualmente nulas com os últimos. A obtenção de resultados compatíveis com os esperados indica que o teste, além de convergir com medidas similares, possui certo grau de especificidade.

Nessa categoria ampla de variáveis externas há testes que mensuram construtos relacionados, mas que não são os mesmos do teste em processo de validação. Neste caso, são esperadas correlações moderadas entre esses testes. Evidentemente tais relações devem ser fundamentadas por um racional teórico sustentado na literatura. Há uma certa sobreposição conceitual entre esse tipo de variável e a variável critério uma vez que essa última também pode ser entendida como uma variável relacionada. Entretanto há uma diferença importante entre as duas, pois variáveis critério referem-se a eventos observáveis importantes em si mesmos que se supõe serem variáveis dependentes de inúmeros fatores incluindo a variável psicológica medida pelo teste. Variáveis-critério geralmente são resultantes de processos complexos e raramente podem ser consideradas propriamente uma medida de um construto psicológico. Em contrapartida, a variável externa composta por um teste medindo construtos relacionados é uma medida semelhante ao teste no sentido de medir processos psicológicos de dimensões relacionadas. Assim esse último tipo de variável externa não deve ser confundida com variáveis critérios, usualmente mais difíceis e trabalhosas de serem obtidas.

Com relativa frequência, estudos envolvendo a correlação de um teste com outro que mede o mesmo construto, já validado, chamado de padrão ouro, são referidos como sendo de validade de critério concorrente. Entretanto essa classificação não corresponde a uma aplicação estrita das definições propostas pelo *Standards* de 1999, sendo mais adequadamente denominados estudos de validade convergente. Assim, variáveis critério embora possam ser entendidas como variáveis relacionadas, têm um *status* diferente de outros testes medindo construtos relacionados (S. Urbina, comunicação pessoal, julho de 2005).

A quinta fonte de informação chama-se evidência com base nas consequências da testagem e inclui os pontos discutidos anteriormente sobre a conceituação ampliada de Messick (1989) sobre a validade. Essa fonte de evidência examina as consequências sociais intencionais e não intencionais do uso do teste para verificar se sua utilização está surtindo os efeitos desejados de acordo com o propósito para o qual foi criado. É evidente que essa propriedade não é do teste em si, mas sim do contexto mais amplo no qual ele é usado. Portanto, as consequências da testagem assumem grande importância no que se refere à validade no sentido amplo. Os procedimentos para esse tipo e

validade buscam mostrar se os benefícios antecipados pelo emprego do instrumento estão de fato ocorrendo.

Esse tipo de evidência não informa propriamente sobre a validade técnica do instrumento mas sobre as interpretações dos escores intencionadas ou não quando este passa a ser usado em contextos aplicados. Espera-se sempre que os resultados de um teste possam trazer algum benefício, seja, por exemplo, para definir uma estratégia mais adequada de terapia, para que as organizações contratem indivíduos a partir de critérios defensáveis e justos, para que o governo e os diferentes profissionais envolvidos com educação possam delinear projetos que visem a melhoria da qualidade de ensino, para citar somente alguns exemplos. No entanto, os resultados benéficos não decorrem simplesmente da existência de testes validados. Existem inúmeras variáveis associadas à utilização que podem interferir no propósito e sentido último e amplo do instrumento. Dentre elas, pode-se citar os aspectos mais básicos do uso de testes, como o discernimento para a sua escolha em função da situação e contexto, bem a sua aplicação e interpretação de forma coerente com o que as pesquisas indicam. A má utilização de um teste pode comprometer a validade das interpretações, fazendo ruir a legitimidade dos dados obtidos. Assim, mesmo tendo instrumentos válidos, se a sua utilização for incorreta, a validade das interpretações é comprometida.

Esse tipo de validade é um conceito novo, incluído nos *Standards* de 1999, e são ainda recentes os estudos que abordam como parte integrante da validade o que o teste, ou melhor, os resultados deste, trazem de benefícios e malefícios ao indivíduo ou à sociedade. No meio científico, essa conceituação tem gerado discordância entre os pesquisadores. Muitos contestam e não concordam que essa questão possa ser incluída no escopo da validade de um teste, pois implica situações que fogem do controle da pesquisa e da construção de um teste. Aparentemente esse tipo de evidência parece não ter relação com a validade se restringimos esse conceito ao seu sentido circunscrito ao micro nível do instrumento. Nesse nível o levantamento de informações técnicas justificando que conjunto de itens consegue medir determinado construto, pode ser feito sem se referir às questões mais amplas tratadas nos estudos de validade consequencial.

Apesar de tais argumentos, considera-se relevante um amadurecimento das discussões acerca das evidências da base consequencial, já que as informações levantadas nesse nível são essenciais aos instrumentos pois lidam com o cerne de sua existência e da sua utilidade para algum bem maior dentro da Psicologia. Assim apesar

das contestações e objeções de muitos pesquisadores sobre a existência da evidência de baseada nas consequências da testagem como uma fonte de validade, não se pode ignorar sua importância fundamental em evidenciar o desígnio último dos instrumentos em serem úteis em contextos aplicados por meio da obtenção de informações sobre o funcionamento psicológico das pessoas. É uma questão ética vincular o teste as consequências que ele pode propiciar tanto ao indivíduo quanto a sociedade. Os argumentos dos críticos a esse tipo de validade, de que as evidências de validade consequencial fogem ao controle dos pesquisadores responsáveis pelo teste parecem, na verdade, explicitar o papel dos diferentes “agentes” envolvidos em uma avaliação psicológica. Neste cenário, claramente o psicólogo que está utilizando um dado teste tem papel central para a validade consequencial do mesmo no momento da sua escolha, administração, levantamento, interpretação e comunicação de seus resultados.

Outros profissionais que utilizam esses resultados para a tomada de decisões (governos, profissionais na área da saúde, forense, etc.) também têm um papel importante para a validade consequencial, uma vez que, se estes não procurarem entender o significado do construto avaliado pelo teste e os limites inferenciais de seus resultados, poderão utilizar as informações oriundas do mesmo de uma forma enviesada, injusta e prejudicial a indivíduos ou, de forma mais ampla, à sociedade. De certa forma, as pessoas avaliadas a partir de um teste psicológico também apresentam um papel para a validade consequencial do mesmo, uma vez que podem desempenhar um papel de fiscalização. Nesse sentido, é possível a identificação de iniciativas do Conselho Federal de Psicologia com a elaboração de cartilhas e comunicados à comunidade para alertar sobre as situações recomendadas para a realização adequada de um processo de avaliação psicológica (Conselho Federal de Psicologia, 2007), para detalhar o papel do SATEPSI e outros pontos relevante nesse contexto.

A comparação das definições antigas de validade de conteúdo, validade de critério e validade de construto com a visão mais contemporânea, calcada nas fontes de evidência proposta na nova edição das *Standards* (AERA, APA, NCME, 1999), permite identificar que houve uma reorganização e renovação em termos conceituais de parte desses métodos. A Figura 2 compara os pontos principais das duas conceituações.

Observando a Figura 2 pode-se perceber que as duas maiores diferenças entre uma proposta e outra, é, primeiro, que nos *Standards*, como já foi comentado, foi abolida a nomenclatura específica para validade de construto fazendo com que os

diferentes métodos discutidos com meios de se evidenciar a validade de construto apareçam como métodos específicos dentro de um das cinco fontes de validade. A segunda diferença, é que os *Standards* propõe uma nova fonte de validade, a evidência baseada nas conseqüências da testagem.

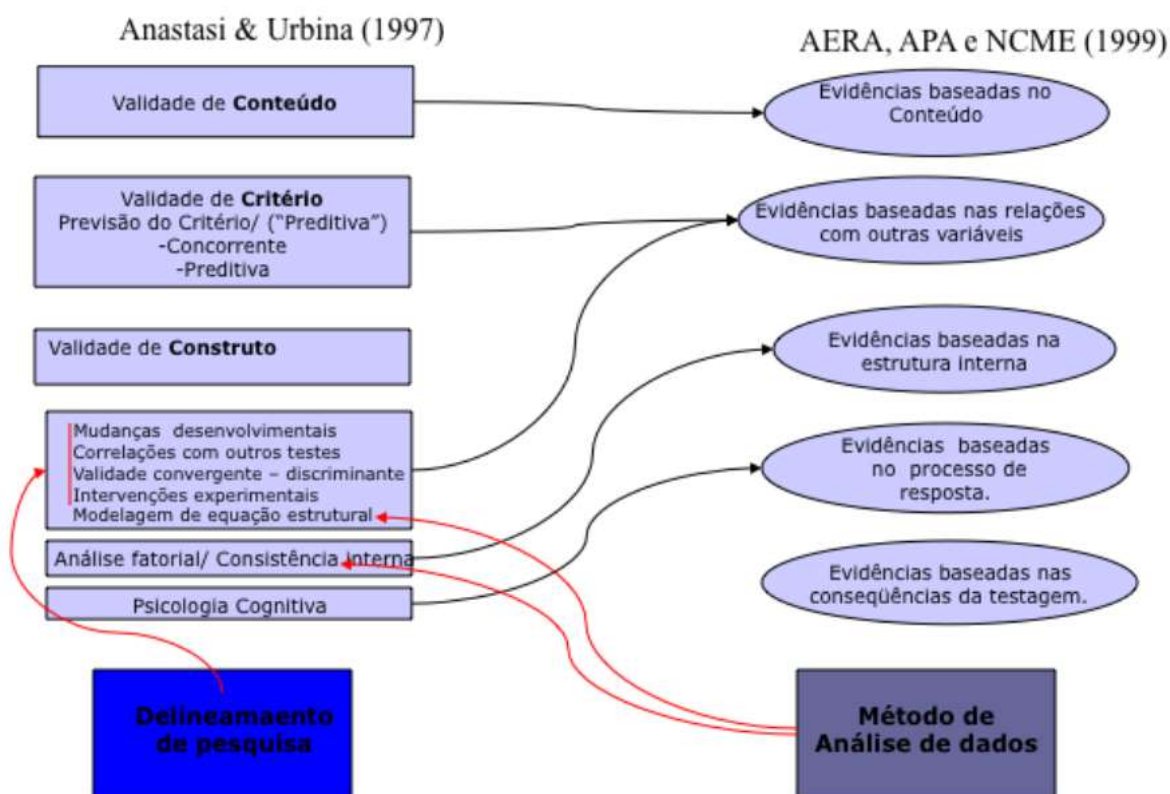


Figura 2. Comparação do antigo modelo conceitual “Santíssima Trindade da Validade” com a visão contemporânea dos *Standards* de 1999.

Como pode ser observado nas relações indicadas na Figura 2, os diferentes estudos que Cronbach e Meehl (1955) propunham como meios de como investigar a validade de construto se distribuem mais logicamente nos *Standards* de 1999. É possível verificar-se também que a nomenclatura se referia a métodos de análise de dados (análise fatorial, por exemplo) e outra aos delineamentos de pesquisa empregados (como é o caso de convergente-discriminante).

O objetivo deste capítulo foi apresentar as definições contemporâneas de validade organizadas nos *Standards* de 1999 e compará-las com a visão clássica

tripartite de validade, referida por parte dos livros publicados antes dessa edição. Ressalta-se que as terminologias validade de conteúdo, validade de critério preditiva ou concorrente e validade de construto, ainda são muito utilizadas no meio científico, bem como talvez sejam as mais conhecidas no Brasil.

A conceituação proposta pelos *Standards* gradativamente tem sido utilizada em artigos, livros, teses e dissertações, mas ainda com menor frequência. Assim, cumpre-se aqui o objetivo de divulgar esse sistema, que é adotado no Sistema de Avaliação dos Testes Psicológicos (SATEPSI), mantido pelo Conselho Federal de Psicologia (ver em http://www.psicologia-online.org.br/servicos/serv_satepsi.cfm#). Tal fato tem acelerado o uso do novo sistema em manuais técnicos de testes psicológicos que, por esse motivo, precisa ser divulgado e discutido para permitir que os usuários de tais testes se apropriem das informações técnicas sobre os mesmos. Pretendeu-se com este capítulo apresentar, de forma sumarizada, as principais diferenças conceituais de validade propostas na última versão dos *Standards* contribuindo com isso ao aumento do nível de cientificidade de sua prática profissional da Psicologia.

Referências

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1986). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- 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. (1997) *Psychological Testing*. Upper Saddle River, New Jersey: Prentice Hall.
- Byrne, B. (2001). *Structural Equation Modeling With AMOS: Basic Concepts, Applications and Programming*. New Jersey: Lawrence Erlbaum Associates.
- Conselho Federal de Psicologia – CFP (2007). *Resolução n.º 002/2003* Em: Conselho Federal de Psicologia. *Profissão Psicólogo: Legislação e resoluções para a prática profissional*. Brasília: CFP/CRP. [On-line]. Disponível: <http://www.pol.org.br>.
- Conselho Federal de Psicologia – CFP (2007). *Cartilha sobre a Avaliação Psicológica*. Brasília: CFP.

- Cronbach, L. J. (1996) *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Embretson, S. (1994). Applications of cognitive design systems to test development. In: Reynolds, C. R., (Ed.). *Cognitive assessment: a multidisciplinary perspective*. New York: Plenum Press.
- Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, 18, 2, 5-11.
- Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Muniz, J. (2004). La validación de los tests. *Metodología de las Ciencias del Comportamiento*, 5 (2), 121-141.
- Pasquali, L. (1999). Testes Referentes a Construto: Teoria e Modelo de Construção. Em L. Pasquali (Ed.), *Instrumentos Psicológicos: Manual Prático de Elaboração*. (pp. 37-71). Brasília, DF: Laboratório De Pesquisa em Avaliação e Medida – LabPAM.
- Primi, R. (1995) *Inteligência, processamento de informação e teoria da gestalt: um estudo experimental*. Campinas, 156p. Dissertação (Mestrado). Instituto de Pós Graduação em Psicologia, PUCCAMP.
- 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. (2002) Complexity of Geometric Inductive Reasoning Tasks: Contribution to the Understanding of the Fluid Intelligence. *Intelligence*, 30 (1), 41-70.
- Primi, R. & Rosado, E. M. S. (1995) Os princípios de organização perceptual e a atividade inteligente: um estudo sobre testes de inteligência. *Estudos de Psicologia*, 11(2), 3-12.
- Wolfe, E. W. & Smith Jr. E. V. (2007). Instrument Development Tools and Activities for Measure Validation Using the Rasch Models: Part II – Validation Activities. *Journal of Applied Measurement*, 8(2), 204-234.