

Validade dos Testes Psicológicos: Será Possível Reencontrar o Caminho?

Luiz Pasquali¹
Universidade de Brasília

RESUMO – Desde o trabalho de Cronbach e Meehl dos meados do século passado, o conceito de validade dos testes em Psicologia e Educação vem perdendo o seu sentido original. Embora os autores quisessem precisamente salvar esse conceito, a introdução do modelo da rede nomológica, concebida dentro da visão do positivismo lógico veio, na verdade, destruir o conceito de validade, originalmente concebido por Kelly na década de 1920 e, depois, por Cattell. O conceito de validade, finalmente, foi totalmente descaracterizado com a definição do mesmo dada pelo grande psicometrista Samuel Messick, em 1989. Parece fundamental que esse conceito seja redescoberto para salvar as bases da Psicometria. A exposição procura mostrar as confusões que o modelo da rede nomológica introduziu em Psicometria e tentar recuperar o verdadeiro significado de validade no contexto das medidas em ciências psicossociais, em particular, em Psicologia e Educação.

Palavras-chave: validade; rede nomológica; psicometria.

The Validity of the Psychological Tests: Is It Possible to Find the Way Again?

ABSTRACT – Since the work of Cronbach and Meehl of mid last century, the concept of validity of the tests in Psychology and Education has been losing its original meaning. Despite the intention of the authors to precisely save this concept, the introduction of the model of the nomological network, conceived within the framework of the logical positivism, was in fact decisive to destroy the concept of validity as originally conceived by Kelly in the 1920s and later on by Cattell. The concept of validity was finally completely mischaracterized with the definition given by the great psychometrist Samuel Messick in 1989. It seems to be fundamental that this concept is rediscovered in order to save the bases of Psychometry. The present article will try to show the confusion that the nomological network model brought to Psychometry and also to try to recuperate the real meaning of validity within the context of the measures in psychosocial sciences, especially in Psychology and Education.

Key words: validity; nomological network; psychometry.

A Confusão do Conceito Validade

*“A verdade surge mais facilmente
do erro do que da confusão”*
(Francis Bacon, 1869)

Ouve-se com mais frequência ultimamente que um dado teste psicológico que pode ser válido numa situação não o é em outra. Ou, ainda, que tal teste tem tal validade e um outro, uma outra. Inclusive, que existem níveis diferentes de magnitude de validade de um teste. Este modo de falar sobre a validade dos testes psicológicos é, pelo menos, confuso.

Para tornar essa situação mais maluca, veio o grande psicometrista Samuel Messick (1989, p. 13) dizendo o seguinte: “*Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment*” (validade é um julgamento avaliativo integrado do grau em que evidência empírica e racionalizações teóricas apóiam a adequação e propriedade de inferências e ações baseadas em escores de teste ou outros modos de avaliação). Este modo de entender e definir a validade dos testes, que

é considerado o modo moderno de entender validade, inclusive assumido pela American Psychological Association (APA, 1985), tornou o tema um verdadeiro zoológico. Na verdade, você vai encontrar na literatura tradicional e atual sobre o tema, especialmente a norte-americana, um elenco sem fim de expressões ou tipos de validades que um teste pode ter. Vejamos alguns deles:

1) Validade de construto (*construct validity*): o teste mede um atributo ou qualidade que não é “operacionalmente definido”; (Cronbach & Meehl, 1955).

2) Validade de conteúdo (*content validity*): o teste constitui uma amostra representativa de um universo de conteúdo (Cronbach & Meehl, 1955; Haynes, Richard, & Kubany, 1995), além de ser relevante (Messick, 1989).

3) Validade de critério (*criterion-oriented validity*): o teste prediz um critério externo (Cronbach & Meehl, 1955).

4) Validade preditiva (*predictive validity*): variedade da validade de critério, em que este é medido temporariamente depois de obtidos os dados do teste (Cronbach & Meehl, 1955).

5) Validade concorrente (*concurrent validity*): variedade da validade de critério, em que este é medido simultaneamente à coleta dos dados do teste (Cronbach & Meehl, 1955).

6) Validade aparente (*face validity*): consiste em se ter “peritos” revendo os conteúdos de um teste para ver se eles são apropriados “em sua cara” (Mosier, 1947, 1951).

1 Endereço: LabPAM, Instituto de Psicologia, Universidade de Brasília, Campus Universitário Darcy Ribeiro, ICC sul, Bloco A, Brasília, DF, Brasil 70910-900. E-mail: labpam@unb.br

7) Validade generalizável (*validity generalization*): a informação dos escores do teste deve ser generalizável sobre populações e tempo (Mosier, 1947, 1951; Messick, 1989).

8) Validade discriminante (*discriminant validity*): um teste tem validade discriminante se mostrar correlação nula com um teste que mede um traço independente de personalidade (Campbell & Fiske, 1959).

9) Validade convergente (*convergent validity*): um teste tem validade convergente se mostrar correlação alta com um teste que mede um traço de personalidade teoricamente relacionado ao que o teste mede (Campbell & Fiske, 1959);

10) Validade incremental (*incremental validity*): a questão de se uma medida particular aporta poder explicativo sobre e além de outra medida para prever um critério relevante (Bryant, 2000).

11) Validade fatorial (*factorial validity*): um tipo de validade de construto em que testes são submetidos à análise fatorial para verificar se possuem variância comum (caso em que se diz que estão cobrindo o mesmo construto) (Guilford, 1946).

12) Validade lógica (*logical validity*): um teste julgado válido por peritos (Cronbach, 1949).

13) Validade empírica (*empirical validity*): Cronbach (1949).

14) Validade conseqüencial (*consequential validity*): os aspectos sociais dos escores dos testes devem ser levados em conta (Messick, 1989).

15) Validade intrínseca (*Intrinsic validity*): Gulliksen (1950).

16) Validade substantiva (*Substantive validity*): validade baseada em bases racionais ou teóricas (Messick, 1989).

17) Validade estrutural (*structural validity*): as respostas devem ser internamente consistentes sobre diferentes partes do teste (Messick, 1989).

18) Validade externa (*external validity*): os escores do teste devem se correlacionar com outras medidas ou variáveis de fundo (Messick, 1989) ou a medida pode ser generalizada através de várias situações (Emory, 1985; Lönnqvist & Hannula, s/d).

19) Validade interna (*internal validity*): são as validades de critério, de conteúdo e de construto (Emory, 1985; Lönnqvist & Hannula, s/d).

20) Validade de hipótese (*hypothesis validity*): uma medida tem validade de hipótese se, em relação a outras variáveis, ela “se comporta” como dela se espera (Weber, 1990).

21) Validade indireta (*indirect validity*): o mesmo que validade de hipótese (Janis, 1965).

22) Validade positiva (*positive validity*): o oposto de validade preditiva (Haynes & cols., 1995).

23) Validade curricular (*curricular validity*): constitui uma extensão da validade de conteúdo e consiste em verificar o aumento da aprendizagem (se se descobre que há aumento de aprendizagem em dois testes com validade de conteúdo, então se verifica validade curricular).

24) Validade diferencial (*differential validity*): validade de uma bateria de testes avaliada pela capacidade de prever diferenças no desempenho em dois ou mais critérios.

25) Validade cruzada (*cross validity*): confirmar a validade dos resultados a partir de um novo exame com estudo empírico feito com uma segunda amostra independente.

26) Validade de grupos mistos (*mixed-group validity*): duas amostras com formatos diferentes no traço ou diferen-

tes probabilidades em expressar dado comportamento são comparadas.

27) Validade múltipla (*multiple validity*): um teste tem validade múltipla quando estiver atrelado a uma amostra vasta de critérios.

28) Validade ecológica (*ecological validity*): o quanto um instrumento psicológico mede fatores espaciais, temporais e situacionais do campo de aplicação.

29) Validade sintética (*synthetic validity*): validade de teste complexo ou de uma bateria de testes baseada no fato de que vários fatores foram representados num único escore composto.

30) Validade condicional (*conditional validity*): a validade do teste depende do uso que dele se faz.

31) Validade incondicional (*unconditional validity*): a validade do teste depende do construto sendo medido e não do uso que dele se faz.

32) ?

Você está convidado a acrescentar outros tipos de validade, se quiser utilizar sua criatividade ou sobrar espaço! Quer dizer: parece que perdemos o rumo! Isso, porque se reduziu a validade de um instrumento de medida a um julgamento sobre as condições de obtenção de uma dada medida (o escore no teste), a utilidade e os usos que se fazem ou se podem fazer da mesma. Ela já não é mais um parâmetro objetivo de instrumento. Assim, validade significa tudo o que diz respeito aos testes psicológicos e, conseqüentemente, não explica mais nada.

Tomar a expressão validade como um guarda-chuva para tudo que diz respeito aos testes psicológicos torna esse conceito, pelo menos, confuso, dizem Borsboom, van Heerden e Mellenbergh (2003). Esses autores insistem em que essa extensão do conceito validade nos desenvolvimentos teóricos recentes não ajudou na clarificação da semântica da validade. De fato, essa mistura no conceito encerra diferentes problemas, ou seja,

- Questão ontológica: O que significa para um teste dizer que é válido? Isso significa perguntar: O que faz a proposição “Escore no teste X medem o atributo Y” verdadeira?

- Questão epistemológica: Como podemos saber se um teste é válido?

- Questão metodológica: Como podemos investigar se um teste é válido?

- Questão ética: Quando e como deveríamos utilizar os escores de um teste?

Infelizmente, a primeira questão, que aparece como a mais fundamental na discussão do que seja validade dos testes psicológicos, é a que praticamente não tem recebido atenção na literatura sobre o assunto. Parece, então, que a intenção de Messick de “integrar” os diferentes aspectos de validade dos testes psicológicos, de fato introduziu confusão, misturando alhos e bugalhos.

Borsboom, Mellenbergh e van Heerden (2004, p. 1) colocam muito bem esse problema ao afirmarem:

O conceito de validade com o qual os teóricos estão interessados parece estranhamente divorciado do conceito que pesquisadores têm em mente quando colocam a questão da validade. Isso se deve a que, no século passado, a questão da validade evoluiu da questão de se a gente mede o que se pre-

tende medir (Cattell, 1946; Kelly, 1927) para a questão de se relações empíricas entre escores de um teste se emparelham com relações teóricas numa rede nomológica (Cronbach & Meehl, 1955) e, finalmente, para a questão de se interpretações e ações baseadas em escores de testes são justificadas – não somente à luz de evidência científica, mas com respeito a conseqüências sociais e éticas do seu uso (Messick, 1989).

A rede nomológica

Para melhor enquadrar a problemática dos testes psicológicos, vamos verificar o que exatamente significa esta história da rede nomológica, inventada por Cronbach e Meehl (1955), e que domina hoje o campo da validade dos testes psicológicos. Esses autores também introduziram o conceito de validade de construto no contexto dos testes psicológicos. Inclusive, eles afirmaram que esse tipo de validade exige um novo enfoque científico. De fato, dizem eles, que a compreensão do que seja tal conceito está ligada ao que chamaram de rede nomológica (*nomological network*). Para introduzir essa problemática, os autores estabeleceram seis princípios fundamentais, quais sejam:

1) Falando cientificamente, para “tornar claro o que algo seja” significa estabelecer as leis em que ele ocorre. Chamaremos a esse sistema interlaceado de leis que constitui a teoria, de rede nomológica.

2) As leis numa rede nomológica podem relacionar a) propriedades ou quantidades observáveis entre elas mesmas; ou b) construtos teóricos a observáveis; ou c) diferentes construtos teóricos entre eles mesmos. Essas “leis” podem ser estatísticas ou determinísticas.

3) Uma condição necessária para se admitir cientificamente um construto consiste em que ele deve ocorrer numa rede nomológica, onde, pelo menos, *algumas* de suas leis envolvem observáveis. Construtos admissíveis podem estar bem longe da observação, i.é, uma derivação longa pode intervir entre os nomológicos que definem implicitamente o construto, e os (derivados) nomológicos de tipo *a*. Estas últimas proposições permitem predições sobre eventos. O construto não é “reduzido” às observações, mas apenas combinado com outros construtos na rede para se fazerem predições sobre observáveis.

4) “Aprender mais sobre” um construto teórico é uma questão de elaborar a rede nomológica na qual ele ocorre, ou de aumentar a precisão dos componentes. Pelo menos na história inicial do construto, a rede será limitada e o construto terá ainda poucas conexões.

5) Um enriquecimento da rede, tais como acrescentar um construto ou uma relação à teoria, é justificado se tal gerar nomológicos que são confirmados pela observação ou se ele reduz o número de nomológicos exigidos para prever as mesmas observações. Quando observações não se enquadram na rede como ela se apresenta no momento, o cientista tem certa liberdade em escolher onde modificar a rede. Isto é, pode haver construtos alternativos ou formas diferentes de organizar a rede que, por enquanto, são igualmente defensáveis.

6) Podemos dizer que “operações” que são qualitativamente muito diferentes “sobrepõem-se” ou “medem a mesma coisa”, se as suas posições na rede nomológica as

ligam à mesma variável construto. Nossa confiança nessa identificação depende do montante de suporte indutivo que possuímos para as regiões envolvidas da rede. Não é necessário que se faça uma comparação observacional direta das duas operações – ficamos contentes com uma prova intra-rede indicando que as duas operações produzem estimativas da mesma quantidade definida pela rede. Assim, os físicos ficam satisfeitos ao falar da “temperatura” do sol e da “temperatura” de um gás numa temperatura ambiente de quarto, mesmo que as operações de testagem sejam não-sobrepostas, isso porque tal identificação faz sentido teórico.

Esses princípios fundamentais podem, mínima e esquematicamente, ser representados na seguinte estrutura (Figura 1):

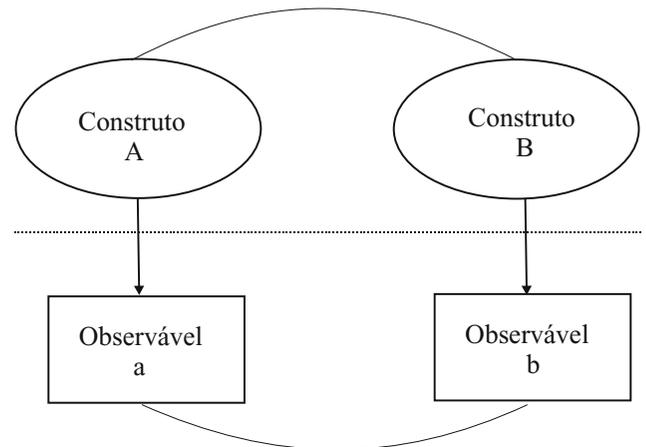


Figura 1. Rede nomológica.

A rede nomológica relaciona construtos com construtos (A e B), observáveis com observáveis (*a* e *b*), construtos com observáveis (A e *a*; B e *b*). Essas relações podem ser de covariância (estatísticas: ligações em curvas) ou de causalidade (determinísticas: ligações em setas). Ademais, um construto é admissível unicamente se ele pode ser inserido numa rede nomológica e ele se torna recomendado se for útil para prever observações. Dessa forma, um construto isolado não tem significado e, portanto, não faz sentido, porque ele adquire sentido em relação a outros construtos. Essa é a conversa (interpretada) de Cronbach e Meehl (1955).

Na ilustração da Figura 1, o que existe de real é a relação, empiricamente verificada, entre observável *a* e observável *b*. Esses dois observáveis não são idênticos, mas estão relacionados, então eles devem estar significando coisas diferentes, ainda que relacionadas. Para explicar tal ocorrência é preciso inventar dois construtos diferentes, mas relacionados, que justifiquem porque os dois observáveis estão relacionados. Assim, se os dois construtos inventados são diferentes, mas estão relacionados e o mesmo ocorre com seus observáveis, isso valida a relação entre os observáveis, porque ocorre um emparelhamento entre a rede teórica dos construtos e a rede empírica dos observáveis. É uma circularidade muito conveniente para provar a validade dos testes psicológicos via construtos. De fato, esses tais construtos são simplesmente um espelho virtual da relação entre os observáveis. Judd, Kidder e Smith (1986) ilustram isso como na Figura 2, onde há um fator A associado a três outros fatores (B, C, D) na

rede teórica (lado esquerdo da Figura), tendo como espelho as respectivas medidas empíricas (lado direito da Figura).

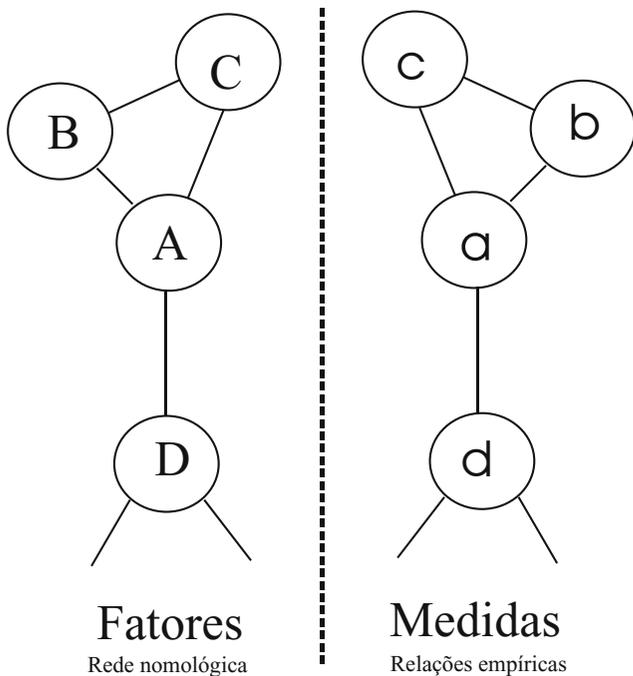


Figura 2. Correspondência entre fatores da rede nomológica e padrão empírico das relações das medidas.

Parece honesto se poder perguntar qual é a natureza ontológica desses construtos ou fatores?

Poder-se-ia, em princípio, dar duas respostas a essa questão, dentro da inteligência de rede nomológica. A primeira seria dentro do positivismo lógico – que, aliás, fundamenta a história da rede nomológica e que Suppe (1977) mostrou ter fracassado –, dizendo que a pergunta é impertinente, porque os construtos adquirem significado pela relação que um têm com o outro, sem com isso implicar que eles existam na realidade. Assim, se dissermos que raciocínio dedutivo se relaciona positivamente com raciocínio abstrato e negativamente com agressividade, o que é esse raciocínio dedutivo? Resposta: ele é algo que se relaciona positivamente com algo e negativamente com outro algo, sendo estes “algos” simplesmente termos de uma relação teórica. No final das contas, construtos são simplesmente rótulos embutidos numa rede teórica. Parece, então, que finalmente eles são artefatos estatísticos, necessários para explicar as covariâncias entre os observáveis.

Uma segunda resposta poderia ser dada dentro de um contexto epistemológico. Aqui, a própria palavra “construto” já deixa entrever o que se pretende significar com construto. Ela vem de “construir”. Mas, construir a partir de quê? Será ela apenas uma representação mental de uma realidade observável ou seria ela uma realidade mental independente do observável? Se for o primeiro caso, então construto constitui apenas uma representação e, com isso, se insere num contexto puramente epistemológico, de conhecimento, e assim ele é de fato construído pelo sujeito conhecedor e não pela natureza. Se, contudo, construto for concebido como a segunda alternativa, então ele é uma realidade em si e, com isso, ele se insere num contexto ontológico, uma realidade construída pela natureza. Parece claro que a posição dos defensores da

rede nomológica entenderia construto unicamente no contexto epistemológico, no máximo, quando não o entende como puro rótulo sem nenhuma consistência de realidade. Ele está ali unicamente para dar significado à relação que existe entre os observáveis. Dessa forma, se entendido como realidade epistemológica, o construto não se constitui apenas como um rótulo sem concretude própria, mas é uma realidade, somente que se expressa como uma representação e, portanto, totalmente dependente do seu observável. Mas, nesse caso, a seta que vai do construto para o observável, como na Figura 1, é injustificável, porque ela deveria ser direcionada ao contrário, porque não é o construto que causa o observável e, sim, o observável que causa o construto. Só para adiantar, esse modo de pensar se contrapõe frontalmente à visão da psicometria atual da Teoria de Resposta ao Item, por exemplo, que afirma exatamente o oposto dessa concepção assim entendida de rede nomológica (veja também Da Silva, 2005).

Então, o que significa agora dizer que o teste é válido se de fato mede aquilo para o qual ele foi feito para medir? Claro, ele mede o construto; mas o construto foi inventado a partir do observável (do score do teste) e, assim, é óbvio que será medido por este. Parece uma tautologia, que soa quase como uma safadeza. O conceito ou a percepção do gato corresponde ao gato físico. Certo! Mas, é o gato percepção que justifica, valida o gato físico ou será este que valida sua percepção? Num contexto epistemológico, é óbvio que é o gato físico que valida a percepção do gato, porque esta é simplesmente uma representação mental do gato físico e ela deve corresponder a este gato físico para ser legítima. Mas, os construtos de que fala a Psicometria e a Psicologia em geral serão também apenas representações mentais ou serão eles processos ou realidades em si, sem serem puras representações? É a eterna briga entre a visão materialista e dualista do ser humano, que sorratamente surge de novo nessa visão da rede nomológica. Infelizmente, a solução desse problema constitui uma questão filosófica e não empírica. E, dessa forma, ela vai continuar a ser um problema enquanto houver pesquisadores que pensam de forma diferentes sobre o tema.

Sintetizando: O conceito construto pode ser entendido de três formas, ou seja:

1) Construto é um rótulo, que representa uma classe de comportamentos (como, inteligência representando ou resumindo todos os comportamentos chamados de inteligentes), sem nenhuma consistência de realidade.

2) Construto é um conceito, que representa alguma realidade observável (comportamento). Assim, ele é uma realidade epistemológica, dependente de uma realidade observável.

3) Construto é uma realidade psíquica, independente e preliminar ao observável que o representa; assim, construto é uma realidade ontológica e, conseqüentemente, deve ser entendido em si mesmo e não em referência a observáveis.

No caso dos testes psicológicos, poder-se-ia aduzir a balbúrdia do conceito de validade a que levou a visão positivista, materialista, subentendida na rede nomológica, para desacreditar essa visão, que se tornaria uma tentativa frustrada. Contudo, no caso da visão dualista, em que existem os processos psíquicos independentes, fica o grave problema de saber o que eles são e quais são as leis que regem esses processos, isto é, a teoria psicológica. E esta está longe de ser uma realidade minimamente consensual em seus detalhes entre os psicólogos. Então, parece que estamos

num mato sem cachorro! Minha opinião é de que devemos desenvolver honestamente a teoria psicológica, descobrindo e não inventando as leis que governam os processos psíquicos, e deixar de lado subterfúgios ou palavrões como a tal da rede nomológica. Aliás, a história de aceitar um construto que explica fenômenos observáveis tem sido corriqueira em qualquer empreitada científica, inclusive na Física e na Biologia. Em Biologia, o gene era aceito como construto mesmo antes de sua descoberta recente. Em Física, a gravidade (G) continua sendo elusiva, mas ninguém duvida da sua existência (Da Silva, 2005).

Incongruências da rede nomológica

Cronbach e Meehl (1955, p. 290) dizem:

Validação de construto ocorre quando um investigador crê que seu instrumento reflete um construto particular, ao qual estão ligados certos significados. A interpretação proposta gera hipóteses testáveis específicas, que constituem o meio pelo qual se confirma ou desconfirma a afirmativa.

Carmines e Zeller (1979, p. 23) explicam: “Fundamentalmente, a validade de construto diz respeito ao tanto que uma dada medida se relaciona com outras medidas consistentes com hipóteses teoricamente derivadas com respeito aos conceitos (ou construtos) que estão sendo medidos”. Isso parece significar que a validade de construto diz respeito ao quanto à relação existente (vista, observada: a correlação) entre duas medidas (de observáveis) é consistente com hipóteses teoricamente deduzidas sobre os construtos sendo medidos por elas. Ou seja, se tenho duas medidas (*a* e *b*; veja Figura 1) que se apresentam relacionadas (na verdade, é só isso que tenho como dado empírico), então elas são consistentes se puder supor que os dois construtos (A e B) que elas medem foram definidos como correlacionados.

Vamos procurar entender melhor essa história, utilizando o dilema da validade de construto de Brown e Burrows (1992), adaptado de McGraw (1982) e analisado na Figura 3.

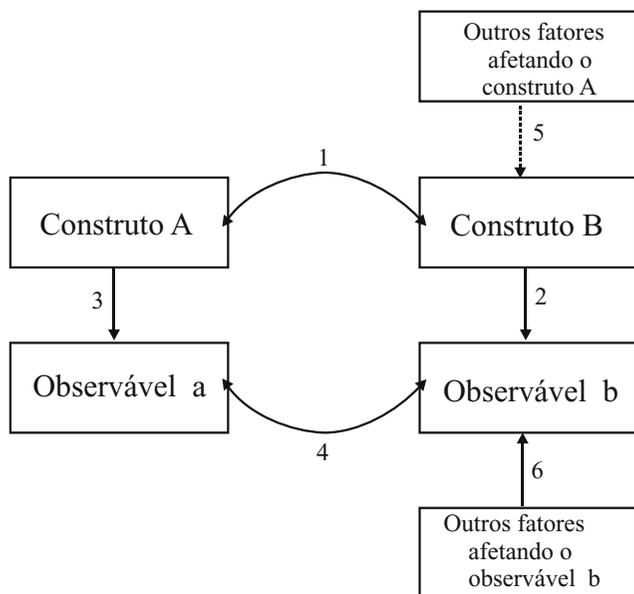


Figura 3. Ilustração da validação de construto (adaptada de McGraw, 1982 e Brown & Burrows, 1992).

Quero validar o instrumento que mede o observável *a*, que constitui uma medida do construto A (relação 3), utilizando a relação que existe (tida como já validada) entre a medida do observável *b* como medida do construto B (relação 2). Então digo o seguinte:

O construto A se relaciona ao construto B (premissa teórica: relação 1).

A medida *a* se correlaciona à medida *b* (dado empírico: relação 4).

A medida *b* mede o construto B (hipótese confirmada).

Conseqüentemente, a medida *a* mede A (hipótese: relação 3) porque se relaciona à medida *b* que mede B (hipótese confirmada: relação 2), dado que este construto está correlacionado ao construto A. Isto é, *a* mede A assim como *b* mede B.

Parece difícil ver a lógica desse silogismo. A base da argumentação se fundamenta na correlação que existe entre *a* e *b*, supondo que A e B também estejam correlacionados, e que a relação 2 seja empiricamente verificada verdadeira.

Vamos analisar essa história em dois momentos: a) a validade da relação 2 e b) a validade da conclusão de que se relação 2, então relação 3.

a) A validade da relação 2: *b* mede B

Como se pode estabelecer empiricamente que a relação 2 (*b* para B) seja verdadeira, isto é, válida? Se a resposta for: do mesmo modo que se estabeleceu a validade da relação 3 (*a* para A), isto é, trabalhando uma relação entre medida *c* e construto C. Obviamente, esse argumento leva a uma seqüência ou redução *ad infinitum* e é, por conseguinte, inviável. Então, como finalmente estabelecer empiricamente validada uma relação de medida com seu construto, que possa servir de base para o silogismo acima apresentado? Aqui entra aquele argumento indutivista que já Hume (1739/1888) e Popper (1972) mostraram que não funciona. O argumento é o seguinte: A evidência com respeito à relação 3 aumenta com o aumento de estudos. Zeller e Carmines (1980, p. 82) explicam essa história da seguinte forma:

A validade de construto não se estabelece com a confirmação de um único estudo em diferentes ocasiões ou pela confirmação de muitas predições num único estudo. Ao contrário, a validade de construto idealmente requer um padrão de descobertas consistentes envolvendo diferentes pesquisadores por um período significativo de tempo e com respeito a uma variedade de variáveis diversas e teoricamente relevantes. Somente se e quando essas condições foram atendidas, pode-se falar com confiança da validade de construto de uma medida particular.

Assim, agindo como esses autores sugerem, a gente vai acumulando evidências sobre a validade de construto de uma dada medida até que, no fim do mundo, a gente tem a certeza absoluta dessa validade! E se nesse processo ocorrer evidência negativa, qual a conseqüência sobre a validade da medida? Carmines e Zeller (1979) dizem que isso pode ocorrer por várias razões, quais sejam,

1) A medida não tem validade de construto. Nesse caso, a medida é simplesmente falsa e deve ser abandonada como medida do construto de interesse.

2) A rede nomológica que serviu para gerar as predições empíricas está incorreta. Nesse caso, a rede teórica está errada e é preciso revê-la.

3) O procedimento para testar as hipóteses teoricamente derivadas está incorreto. Nesse caso, ou a medida (o teste) foi mal construída ou as análises estatísticas foram mal feitas. Assim, tanto a medida quanto as análises têm que ser revistas.

4) Falta de validade de construto ou de precisão de alguma(s) outra(s) variável(eis) na análise. Isto é, um dado construto vem sempre associado a outros construtos. Nesse caso, foi feita uma relação entre construtos que, de fato, não se relacionam. Há, portanto, falha na parte teórica da rede nomológica.

Isso é muito bonito. Mas, duas coisas: primeiro, os estudos que vão dando resultados positivos, isto é, apoiando a hipótese de validade de construto, vão aumentando a probabilidade de certeza dessa validade. Agora, acontece que, após ter aumentado essa probabilidade, um resultado negativo ocorre. O que acontece com essa probabilidade? Ela diminuiu ou acabou? Se o resultado negativo resulta de uma pesquisa bem conduzida, esse resultado constitui a morte da hipótese, porque é uma regra básica, de qualquer pesquisa científica, que um único resultado negativo torna a hipótese inválida, por mais resultados positivos que tenha tido anteriormente. Trata-se da regra da confirmação empírica (Pasquali, 2005a). Em segundo lugar, mesmo que esse proposto processo funcionasse, tratar-se-ia de validação de um teste e não de validade do teste. Validade constitui uma propriedade de algo, enquanto validação constitui uma atividade do cientista para verificar tal propriedade. Esta se insere num processo epistemológico, de conhecimento, enquanto aquela se insere num campo ontológico, da realidade da propriedade e do algo em questão. As duas coisas são importantes; mas, uma não é a outra, porque uma fala “do que” se trata, enquanto a outra fala do “como fazer”. A validade parece ser um conceito bastante simples, se entendida como a propriedade de um teste estar de fato se referindo a (representando) um construto, enquanto que validação já se apresenta bastante mais complexa, porque se trata de provar cientificamente a validade do instrumento. Neste último contexto, uma teoria como a da rede nomológica é útil, sobretudo quando operacionalizada como fizeram Campbell e Fiske (1959) com a sua validade convergente-discriminante. Apenas que os construtos em jogo não são conceituados pela rede nomológica; eles são preliminares e definidos em termos da teoria substantiva psicológica e, então, entrosados numa rede nomológica.

b) A validade da relação se 2 e 1 e 4, então 3

O modelo de validação de construto afirma que se as relações 1, 2 e 4 forem verdadeiras, então segue que a relação 3 também o é. Entretanto, todas essas relações postuladas no modelo da rede nomológica são expressas como correlações, isto é, em termos de covariância. Agora, acontece que

a covariância nunca é perfeita, ou seja, a correlação nunca é 1. Suponha, por exemplo, que as correlações propostas entre as variáveis da rede nomológica sejam todas de 0,70, que representam correlações muito expressivas. Obviamente, trata-se das correlações que expressam as covariâncias com respeito às relações 1, 2 e 4, porque a relação 3 é aquela que você quer demonstrar empiricamente. Em outras palavras, você está postulando que

- $r_{AB} = 0,70$
- $r_{Bb} = 0,70$
- $r_{ab} = 0,70$.

Pergunta-se: $r_{Aa} = ?$

Veja isso ilustrado na Figura 4.

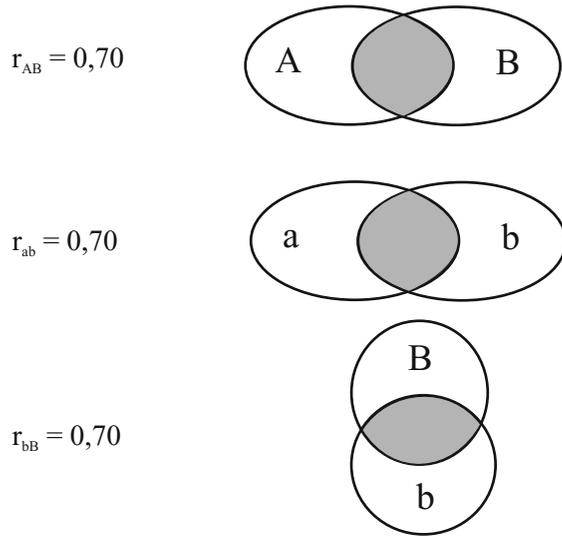


Figura 4. Correlações entre construtos e observáveis.

Dessa situação, poderiam surgir várias possibilidades para a correlação entre a medida *a* e seu suposto construto *A*, duas das quais estão ilustradas na Figura 5.

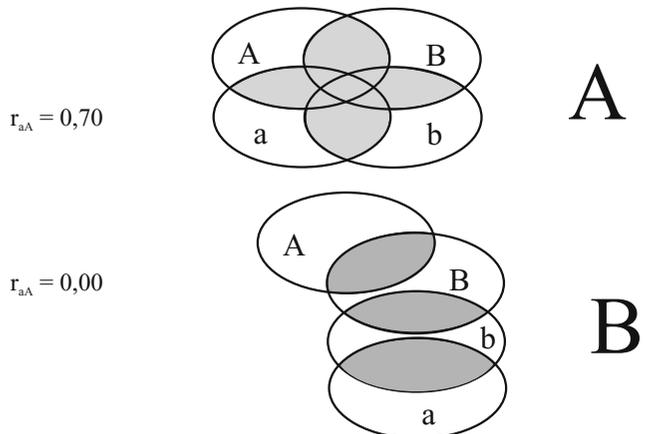


Figura 5. Correlação entre a medida *a* e o construto *A*.

Mesmo salvando todas as correlações postuladas, a correlação entre *a* e o construto *A* tanto pode ser alta (parte A da Figura) quanto nula (parte B da Figura). Dessa forma, o silogismo que argumenta em prol da validade de construto

da medida *a* constitui, no máximo, uma prova circunstancial de tal validade. Isto é, não prova nada.

Como, então, entender a validade de construto de um teste psicológico? Vamos discorrer um pouco sobre esse problema.

Reencontrando o conceito de validade

Entende-se por teste psicológico um conjunto constituído de comportamentos que o sujeito deve exibir. Ele é um teste se todos os comportamentos envolvidos no conjunto se referem à “mesma coisa”, é a questão da unidimensionalidade. Ademais, alguns sujeitos podem ser capazes de executar todos os comportamentos envolvidos, outros sujeitos, apenas alguns dos comportamentos e outros, ainda, nenhum dos comportamentos. Por que isso ocorre? Porque o sujeito que executa todos os comportamentos tem mais daquela “mesma coisa” que o sujeito que executa apenas alguns comportamentos ou nenhum deles. Se isso for verdade, então aquela “mesma coisa” deve ser a) uma realidade, porque os comportamentos envolvidos no teste são reais, b) aquela “mesma coisa” é a causadora das respostas dos sujeitos e c) aquela “mesma coisa” deve variar entre diferentes sujeitos, uma vez que, sendo a causa dos comportamentos, ela produz diferenças nos comportamentos dos sujeitos.

Vamos especificar um pouco essas três afirmações, substituindo aquela “mesma coisa” pelo seu sinônimo, ou seja, “construto”.

Na tentativa de reencontrar o verdadeiro conceito de validade de construto, os autores holandeses Borsboom, Mellenbergh e van Heerden (Borsboom & cols., 2003; Borsboom & cols., 2004), vêm fazendo um trabalho fundamental. A presente exposição se inspira nesses pesquisadores, bem como de trabalhos anteriores do presente autor (Pasquali, 1999, 2004). Veja, também, José Aparecido da Silva (2005), embora este autor tenha predileção por interpretação biológica dos construtos psicológicos.

A base da argumentação consiste em se entender que validade constitui uma propriedade do instrumento de medida e que não tem nada a ver com a idéia de que ela consistiria num julgamento sobre o significado dos escores de um teste, retomando, assim, o sentido original que Kelly (1927) quis dar a esse conceito, ao afirmar que um teste é válido quando mede aquilo que supostamente deve medir.

Antes de continuar, gostaria de lhe propor uma questão: Você construiu um metro, feito de ferro, para medir o comprimento das coisas. Você mediu um pedaço de pau e conseguiu uma medida do mesmo. Em seguida, você esquentou o metro e foi medir novamente o pedaço de pau e conseguiu outra medida dele, agora, porém, ela era bem diferente da primeira. Pergunta: O metro, na segunda vez que você o utilizou, já não está mais medindo comprimento?

Sendo sensato, talvez a sua resposta seja: sim, ele está ainda medindo comprimento, no entanto está medindo errado! Muito bem, então o metro continua medindo aquilo para o qual ele foi feito para fazer, isto é, medir comprimento. Isto é validade. Mas, ele está medindo errado? De acordo, mas continua medindo comprimento, no entanto, o está fazendo erradamente porque o metro, com o calor, ficou descalibrado. Por que ficou descalibrado? Porque houve a presença de

fatores perturbadores no uso do instrumento (o calor). Mas a presença de fatores perturbadores é corriqueira em qualquer empreitada científica (as chamadas variáveis estranhas ou *confounding variables*). De qualquer forma, o instrumento ser calibrado ou não, é questão de precisão, não de validade, porque ele continua medindo aquilo para o qual ele foi construído. Assim, continua sendo verdadeiro que o metro mede comprimento (validade), embora na prática ele o faça erradamente (precisão). Dessa forma, validade responde a se algo é verdadeiro ou falso, enquanto que precisão responde a se algo está correto ou errado. A primeira questão diz respeito a um problema ontológico e a segunda a um problema psicométrico (de mensuração) ou metodológico. No caso dos testes psicológicos, os dois problemas são relevantes e importantes, mas um não é o outro. Assim, conclui-se que os parâmetros de validade e precisão são características do instrumento de medida, do teste, e não da medida feita de um objeto. Esta última é confiável e legítima se o instrumento que a produziu for válido (pertinente, relevante) e preciso (calibrado). Quem garante a qualidade da medida é a qualidade do instrumento. E a validade do instrumento diz respeito exclusivamente à pertinência do instrumento com respeito ao objeto que se quer medir; é a questão da referência. Agora, acontece que referência supõe que dois objetos sejam reais, porque, se um deles não o for, então a relação é fictícia, dado que ela é feita entre uma realidade e uma racionalização (fantasia, invenção). É o que fazem, ao que parece, os positivistas quando falam de rede nomológica.

O posicionamento aqui assumido tem implícita a aceitação de uma visão filosófica e, por isso, empiricamente não demonstrável, sobre o que seja Psicologia e, com isso, o que seja um ser humano. Trata-se da visão dualista, que considera o ser humano um ente ontológico heterogêneo, especificamente composto de elementos do que Popper (1972) chamou de primeiro e segundo mundos; ou seja, de corpo e mente, sendo estes muito distintos entre si, mas perfazendo uma composição e não um agregado (veja Pasquali, 2005a). Essa posição dualista do ser humano é tomada, dado que a visão monista (materialista) não explica a contento o ser humano e, especificamente, no caso dos testes psicológicos, leva às incongruências discutidas anteriormente. Ademais, ela se coaduna perfeitamente com a teoria psicométrica atual da Teoria de Resposta ao Item (TRI, veja Pasquali, 2005b).

Voltando aos testes psicológicos. Um teste, então, é um conjunto de estímulos comportamentais, ou seja, de comportamentos, finalmente, porque os estímulos pretendem produzir comportamentos. Assim, um teste é um conjunto de comportamentos observáveis. Agora, o psicólogo não está especificamente interessado em verificar componentes ou aspectos anatômicos, fisiológicos ou genéticos desses comportamentos, nem fatores ecológicos e nem mesmo fatores culturais que entram no aparecimento de tais comportamentos. Todos esses fatores são relevantes no comportamento humano, mas eles são do interesse do biólogo, do ecólogo ou de sociólogo. O que, então, o psicólogo procura entender desses comportamentos? O psicólogo procura entender os comportamentos com referência a processos psíquicos, os famosos construtos, estes entendidos como realidades ontológicas e não como racionalizações inseridas numa rede nomológica. Dessa forma, para entender o comportamento, o

psicólogo deve entender o seu referente, o processo psíquico, que se constitui como causa do comportamento, para satisfazer o ditado dos romanos de “*scire est per causas scire*” (conhecer se faz conhecendo as causas). O fato de que esses processos psíquicos sejam minimamente conhecidos (isto é, praticamente desconhecidos) não invalidam sua presença, muito menos sua utilidade científica.

Dessa forma, se um teste *a* mede um construto A, então ele deve se referir a este construto especificamente e não porque o teste se correlaciona com outro teste *b* que mede o construto B que, por sua vez, se correlaciona com o construto A. Assim, a validade de um teste deve ser entendida como ilustrado na Figura 6.



Figura 6. Ilustração da concepção dualista do ser humano do teste.

O construto (traço latente, *teta*) se posiciona como o objeto que o teste quer medir, isto é, ele é o aquilo que o teste pretende medir. Então, ele é o referente, em função do qual a qualidade do teste deve ser avaliada. Conseqüentemente, as respostas ao teste (o escore no teste, o observável, o *tau*) não criam o construto, antes, pelo contrário, é o escore no teste que depende do construto. “Quando se diz que um teste é válido, estamos assumindo uma posição ontológica de que o atributo sendo medido existe e afeta o resultado do procedimento de medida” (Borsboom & cols., 2004). Assim, o construto é o referencial para os resultados de um teste. Se estes são reais, então o referencial também deve ser real. Do contrário fica estranho se dizer, por exemplo: o teste X mede inteligência, mas inteligência não existe. As duas afirmações não podem ser simultaneamente verdadeiras. Uma delas é falsa. Como os resultados do teste X são reais, então o construto medido existe.

Comentário Final

Validade diz respeito ao instrumento e não ao uso que se faz dos seus escores. Assim, não faz sentido dizer que um teste com validade de construto é válido numa situação, e não o é em outra. Esse modo de falar faz sentido somente no contexto da teoria clássica dos testes, em que imperava a validação dos testes via critério, em que a validade do teste realmente dependia da estabilidade do critério. Assim, se o critério mudasse de *status*, o teste que o predizia perdia a validade. No caso da validade de construto, não existe esta história de validade condicionada do teste, porque o critério aqui é o *teta*. Assim, o teste não prediz um *teta*, ele representa ou modela comportamentalmente o traço latente. Como se supõe que o traço latente não mude de uma situação para outra como muda um critério, então o teste que afere validamente o tal *teta*, o fará assim em qualquer situação. O problema da validade condicionada com respeito a testes com validade

de construto ainda poderia significar alguma coisa, em duas situações, a saber:

1) O *teta* que o teste mede está mal definido. Para resolver este problema, temos que verificar duas coisas: a) rever a teoria que define o *teta* em questão e b) verificar a função de informação do teste fornecida pela TRI.

2) O traço latente que um dado teste mede validamente numa situação, não se aplica numa outra situação de seleção, por exemplo. Mas este não é um problema psicométrico de validade dos testes e, sim, um problema de psicologia organizacional. Esta deve verificar se um dado traço latente é ou não pertinente no contexto de uma dada seleção. Para isso existe a profiisografia do cargo.

Referências

- American Psychological Association (1985). *Standards for educational and psychological testing*. Washington: American Psychological Association, Inc.
- Bacon, F. (1869). *Novum organum*. Em J. Spedding, R. L. Ellis & D. D. Heath (Orgs.), *The works of Francis Bacon. Vol. VIII* (pp. 179-203). Boston: Taggard and Thompson.
- Borsboom, D., van Heerden, J. & Mellenbergh, G. J. (2003). *Validity and truth*. Internal Report. Department of Psychology, University of Amsterdam.
- Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Brown, K. & Burrows, C. (1992). *What is validity? A prologue to an evaluation of selected health status instruments*. National Centre for Health Program Evaluation, Report 1.
- Bryant, F. B. (2000). Assessing the validity of measurement. Em L. G. Grimm & P. R. Yarnold (Orgs.), *Reading and understanding more multivariate statistics* (pp.99-146). Washington: American Psychological Association.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines, .G. & Zeller, R. A. (1979). *Reliability and validity assessment*. New Park: Sage Publications.
- Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.
- Cronbach, L. J. & Meehl, P. (1955). Construct validity in psychological tests, *Psychological Bulletin*, 52, 4, 281-302.
- Da Silva, J. A. (2005). *Inteligência: Resultado da genética, do ambiente ou de ambos?* São Paulo: Editora Lovise.
- Emory, C. (1985). *Business research methods The Irwin Series in Information and Decision Sciences* (3ª ed.). Illinois: Homewood.
- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.
- Gulliksen, H. (1950). Intrinsic validity. *American Psychologist*, 5, 511-517.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247.
- Hume, D. (1888). *Treatise of human nature. Vol. I*. Oxford: Claredon Press. (Trabalho original publicado em 1739)

- Janis, I. L. (1965). The problem of validating content analysis. Em H. D. Lasswell, N. Leites & cols. (Orgs.), *Language of politics*. Cambridge: MIT Press.
- Judd, C., Kidder, L. & Smith, E. (1986). *Research methods in social relations* (6ª ed.) New York: Rinehart and Winston.
- Kelly, E. L. (1927). *Interpretation of educational measurements*. New York: Macmillan.
- Lönnqvist, A. & Hannula, M. (s/d.). *Validating performance measures with the nomological network*. Retirado em 22/09/2006 de <http://www.tut.fi/units/tuta/teta/mittaritumi/julkaisut/bryssel.pdf>
- McGraw, J. E. (1982). Dilemmatics. The study of research choices and dilemmas. Em J. E. McGraw; J. Martin & R. A. Kukla (Orgs.), *Judgment Calls in Research* (pp.69-102). Beverly Hills: Sage.
- Messick, S. (1989). Validity. Em R. Linn (Org.), *Educational measurement*. (3ª ed., 13-103). New York: American Council on Education and Macmillan Publishing Company.
- Mosier, C.I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement*, 7, 191-205.
- Mosier, C.I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement*, 11, 5-12.
- Pasquali, L. (1999). *Instrumentos psicológicos: Manual prático de elaboração*. Brasília: LabPAM / IBAP.
- Pasquali, L. (2004). *Psicometria. Teoria dos testes na psicologia e na educação* (2ª ed.). Petrópolis: Editora Vozes.
- Pasquali, L. (2005a). *Delineamento de pesquisa em ciência. Vol. I: A lógica da pesquisa científica*. (mimeo).
- Pasquali, L. (2005b). *Teoria de resposta ao item – TRI. Manual para iniciantes*. Brasília: INEP.
- Popper, K. R. (1972). *A lógica da pesquisa científica*. São Paulo: Editora Cultrix.
- Suppe, F. (1977). *The structure of scientific theories*. Urbana: University of Illinois Press.
- Weber, R. P. (1990). *Basic content analysis* (2ª ed.). Newbury Park: Sage Publications.
- Zeller, R. A. & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge: Cambridge University Press.