

VALIDADE E FIDELIDADE DE INSTRUMENTOS DE MEDIDA EM PSICOLOGIA: TUDO O QUE SEMPRE QUIS SABER E NUNCA TEVE CORAGEM PARA PERGUNTAR

DINIZ LOPES & TIAGO RÔXO AGUIAR
Iscte-Instituto Universitário de Lisboa

Palavras-chave: Psicometria; Validade; Fidelidade.

OBJETIVO

Este capítulo visa:

- (a) Definir conceitos fundamentais da psicometria, os seus objetivos e principais correntes inspiradoras;
- (b) Definir validade e fidelidade e caracterizar os tipos de validade e fidelidade que podemos determinar relativamente a um instrumento de medida psicológica;
- (c) Salientar a importância da obtenção de características psicométricas de instrumentos de medida psicológica.

INTRODUÇÃO

Uma análise clínica (e.g., um hemograma, i.e., uma contagem de glóbulos vermelhos) ou a datação de um artefacto antigo (e.g., a datação por carbono 14) são procedimentos de medida testados e consensuais na prática médica e arqueológica, respetivamente. Tal como com estas técnicas, as medidas que aplicamos no contexto da investigação ou intervenção psicológica (seja num contexto clínico, social-organizacional ou educacional) devem ser sujeitas a um cuidadoso processo de avaliação e testagem. De facto, é imperativo confiarmos nos instrumentos de medida que utilizamos e estarmos seguros de que foram criados seguindo processos rigorosos de construção e validação. Estes processos asseguram (pelo menos tanto quanto é possível) que o instrumento que estamos a utilizar está apto a aceder ao(s) construto(s) psicológico(s) que se pretende(m) medir, e que tem em consideração as especificidades do contexto em que é aplicado.

Com este artigo, pretendemos apresentar, de forma sucinta, alguns conceitos básicos respeitantes à avaliação de qualidades psicométricas de instrumentos de medida em psicologia (popularmente chamadas de “escalas” ou “testes”), incidindo especialmente naqueles que são

mais habituais em psicometria – a validade e a fidelidade –, bem como indicar recursos que possibilitam um aprofundamento adicional sobre estes temas.

PSICOMETRIA: PRINCÍPIOS E DEFINIÇÕES A RETER

Falar de psicologia e não falar da medida de construtos psicológicos é praticamente impossível. A medida em psicologia é um dos aspetos mais fundamentais desta disciplina científica e é a área de estudo primordial da psicometria. A psicometria, cujo significado é, justamente, “a medida em psicologia”, compreende os modelos teóricos e as técnicas que permitem ao psicólogo realizar a operacionalização, desenvolvimento e avaliação da qualidade de instrumentos de medida de construtos psicológicos.

A psicometria encontra-se fortemente ancorada na história e desenvolvimento da Psicologia, bem como de outras ciências próximas. Como principais fontes inspiradoras, a psicometria conta com a teoria da medida (Spearman, 1904), a estatística e os métodos quantitativos, e a sua constante preocupação em quantificar “grandezas psicológicas”, ou seja, construtos teóricos de elevado nível de abstração, apenas mensuráveis através de manifestações cognitivo-comportamentais dos indivíduos. Este é um dos desafios fundamentais da psicometria: o de saber fazer a ponte entre um construto que se pretende medir – tipicamente teórico e abstrato e que se crê “latente” nos indivíduos – e a forma como se vai medi-lo na realidade – tipicamente através de um instrumento de medida que torna o construto psicológico “*latente*” num conteúdo “*manifesto*”. Como o hiato entre estes dois mundos é, por defeito, grande, torna-se importante garantir que estes instrumentos de medida realizam, com qualidade, esta travessia do plano latente para o plano manifesto da mensuração.

Desde cedo que a psicometria se centrou em torno de duas correntes essenciais: a **teoria clássica dos testes** (TCT) e a **teoria da resposta aos itens** (TRI). A TCT foi proposta por Spearman (1904) e desenvolvida por Gulliksen (1950), enquanto a TRI foi elaborada inicialmente por Rasch (1960) e Lord (1980) e, posteriormente, desenvolvida por Birnbaum (1968). Neste contexto, é ainda fundamental a noção de “erro” de medida, para o qual contribuíram significativamente Campbell e Stanley (1963). Estas duas correntes estabeleceram definitivamente a questão fundamental da psicometria: a análise do significado das respostas dadas pelos indivíduos numa série de tarefas, normalmente designadas por “itens”. A TCT centra a sua análise na explicação do resultado total obtido pelas respostas de indivíduos a um conjunto de itens (i.e., o “score total”; onde o “score total” = “score verdadeiro” + erro; Spearman, 1904). Já a TRI, não se interessando tanto pelo “score total”, procura nas respostas dos indivíduos a cada um dos itens que compõem um teste a matéria-prima para analisar a qualidade dos mesmos. Isto significa que a TCT está centrada na produção de testes com qualidade, enquanto a TRI se focaliza na produção de itens com qualidade. Assim, de acordo com os postulados da TCT obtemos testes válidos; com a TRI obtemos itens válidos (o que a limite permite a obtenção de testes também eles válidos).

Tanto a TCT como a TRI lançaram as bases de dois outros conceitos fundamentais da psicometria: a validade e a fidelidade. No contexto das ciências sociais, a validade é um parâmetro de qualidade das medidas que é habitualmente discutido e teorizado, não sendo tipicamente analisada nas ciências físicas. A preocupação destas últimas recai sobretudo na noção de fidelidade, ainda que esta seja, também, uma preocupação das ciências sociais. De acordo com os *Standards for Educational and Psychological Testing* (AERA et al., 1999), a validade é “o grau com que a evidência e a teoria suportam a interpretação dos resultados de um teste, decorrente do seu uso” (p.9). Dito de outra forma, a validade procura analisar a congruência entre o construto latente que deu origem ao instrumento de medida e os itens que compõem este instrumento e operacionalizam este construto (Furr, 2017). A evidência relativa à validade de qualquer instrumento de medida em Psicologia é, por este facto, necessária e fundamental. Ela ocupa um lugar central em psicometria, mais especificamente na teoria da medida.

A fidelidade é outra das características psicométricas fundamentais que um instrumento de medida deve apresentar. A fidelidade é, normalmente, associada ao erro de medida (Gregory, 2013). Assim, um instrumento de medida sem erro é aquele que medindo o mesmo construto latente, nos mesmos indivíduos, em ocasiões diferentes (ou utilizando medidas equivalentes que medem o mesmo construto latente nos mesmos indivíduos), deve produzir resultados aproximadamente iguais. Por outras palavras, ao analisarmos a associação entre estes resultados a sua correlação deve ser 1 ou próxima de 1. Contudo, e recordando a equação fundamental de Spearman (1904), sabemos que nenhuma medida está isenta de erro. Por esta razão, a determinação da fidelidade torna-se crucial de modo a testemunhar o grau de erro com que estamos a medir um determinado construto, bem como sinalizar situações em que a medida deve ser melhorada.

QUALIDADES PSICOMÉTRICAS DE INSTRUMENTOS DE MEDIDA PSICOLÓGICOS: VALIDADE E FIDELIDADE

A história do desenvolvimento e validação de instrumentos é vasta e relativamente antiga (Freeman, 1990). Ao longo de muitas décadas, foram desenvolvidas inúmeras metodologias e indicadores que podem ser usados e publicados juntamente com os respetivos instrumentos de medida para testemunhar as suas qualidades psicométricas e argumentar a favor da sua utilização. Estes indicadores e metodologias garantem que um instrumento de medida possui uma de duas (ou ambas) qualidades: validade e fidelidade (Gregory, 2013). Estas qualidades psicométricas são fundamentais em qualquer medida em psicologia, e permitem-nos dizer com confiança se um instrumento está pronto a ser utilizado para o propósito a que se propõe.

VALIDADE

De acordo com as perspetivas mais clássicas da Psicometria (Freeman, 1990), a validade de um instrumento de medida refere-se à sua capacidade para avaliar corretamente o que se propõe medir (Gregory, 2013). Isto significa, genericamente, que os resultados obtidos com este instrumento devem permitir realizar inferências acerca do construto teórico (latente) que está na sua base (e.g., inteligência geral; Spearman, 1904). Da mesma forma, os resultados obtidos com este instrumento de medida devem permitir, também, realizar inferências acerca das diversas facetas que compõem este construto teórico (e.g., a inteligência geral é concebida como unidimensional; a análise da estrutura do instrumento de medida deve refletir este facto). E finalmente, os resultados deste instrumento de medida devem mostrar-se relacionados com os resultados obtidos com outros instrumentos de medida de construtos teórica e conceptualmente próximos (Gregory, 2013; e.g., inteligência fluída vs. inteligência cristalizada; Cattell, 1963).

Assim, e dentro destas perspetivas clássicas, é comum distinguir três tipos de validade e obter evidência a seu respeito, de forma a testemunhar a qualidade psicométrica de um instrumento de medida:

| A **validade de conteúdo** – uma das fontes mais importantes de validade e uma das primeiras a ser analisada na “vida útil” de um instrumento de medida (Messick, 2005). A validade de conteúdo visa obter evidência de que o conjunto de itens que compõem este instrumento são a “amostra mais representativa” de uma “população de itens” passível de operacionalizar o construto teórico que está na sua base (Gregory, 2013). Tipicamente, esta evidência é obtida através daquilo a que se convencionou chamar “análise de peritos” (Lawshe, 1975). É constituído um painel de peritos na área do construto a ser medido e solicitam-se avaliações da adequação do conteúdo do instrumento de medida com base na(s) teoria(s) que sustenta(m) o construto; um caso especial da validade de conteúdo toma o nome de validade facial (ver Holden, 2010). Este tipo de validade procura analisar a usabilidade do instrumento de medida junto da população alvo da medida, ou seja, os examinandos. É pedido que estes se pronunciem acerca da compreensibilidade, adequação das escalas de resposta e outros aspetos formais dos conteúdos do instrumento de medida (Lopes & Pinto, 2016). Desta forma, garantimos que se recolhem dados com uma elevada adequação para a medida do construto em análise, minimizando possíveis fontes de erro relacionadas com a sua administração.

| A **validade de construto** – uma vez que a maior parte dos construtos alvo de medida em psicologia se caracterizam pelo seu elevado nível de abstração (e.g., capacidades, crenças, valores, atitudes), é muito importante construir e utilizar instrumentos de medida que operacionalizem, da forma o mais correta possível, os construtos teóricos que lhes estão subjacentes. Mas é muito importante que o instrumento de medida, uma vez construído, permita obter dados a partir dos quais possamos realizar inferências corretas do construto subjacente. Este último aspeto é crucial à validade de construto, e constitui a sua definição fundamental (Lopes & Pinto, 2016). Para determinarmos a qualidade das inferências realizadas a partir dos

resultados de um instrumento de medida de volta para o construto que lhe está subjacente, diversos níveis de prova são normalmente exigidos: (i) confirmação de que a estrutura interna do instrumento respeita a estrutura teórica proposta para o construto (i.e., unidimensionalidade vs. multidimensionalidade), tipicamente determinado com recurso a técnicas estatísticas do tipo análise fatorial (Lopes & Pinto, 2016); (ii) determinação da sensibilidade da escala, recorrendo-se a diversos tipos de indicadores (e.g., variáveis de caracterização sociodemográfica) de acordo com o estabelecido na definição teórica do construto subjacente (e.g., medida de inteligência sensível à idade do examinando; Ryan et al., 2000); (iii) determinação da associação dos resultados do instrumento de medida com outros instrumentos medindo um construto semelhante (validade convergente; Campbell & Fiske, 1959) ou construto diferente (validade discriminante; Campbell & Fiske, 1959) e finalmente (iv) as consequências individuais ou sociais da utilização do instrumento, podendo afetar, de forma decisiva, a sua qualidade e validade (cf., utilização do teste de Inteligência Geral de Raven na avaliação de populações não-norte-americanas ou não-europeias; Gonthier, 2022; Lopes & Pinto, 2016).

| **A validade relativa a um critério** – diz respeito à capacidade de um instrumento de medida poder ser utilizado para prever, de forma eficaz, resultados de uma outra medida (dita critério) não diretamente mensurados pelo instrumento (Cronbach & Meehl, 1955; Lopes & Pinto, 2016; e.g., uma prova de cultura geral aplicada a candidatos a um determinado curso superior permite prever a sua média final de curso). Este tipo de validade pode assumir dois aspetos distintos (Cronbach & Meehl, 1955): **validade preditiva**, quando a obtenção da medida e do critério ocorre em tempos distintos (como é o caso da prova de cultural geral referida anteriormente); **validade concorrente**, quando medida e critério são recolhidos ao mesmo tempo, ou com um intervalo de tempo muito diminuto (e.g., um questionário de triagem substitui um exame médico para admissão de doentes no serviço de urgência de um hospital).

A validade é, portanto, um conceito abrangente e complexo, e a literatura atual mostra a existência de diversas perspetivas relativamente a este conceito. Ainda que não sejam consensuais (para uma revisão deste dissenso ver Borsboom et al., 2004; Markus & Borsboom, 2013), existe uma série de princípios que compõem aquilo a que atualmente se chama a “conceção contemporânea da validade” (Chan, 2014; Cronbach, 1988; Hubley & Zumbo, 2011; Kane, 2013; Messick, 1989; Zumbo, 2009):

1. A validade diz respeito não ao instrumento em si, mas às inferências e decisões que tomamos com base nos resultados obtidos com o instrumento de medida psicológica;
2. O grande foco da validade é a validade de construto. A restante evidência acerca da validade de um instrumento de medida psicológica serve de suporte à validade de construto e, em consequência, à interpretação e utilização dos resultados obtidos;
3. A validade é um processo evolutivo de acumulação de evidência, não se estabelecendo apenas nas primeiras utilizações de um instrumento de medida psicológica;
4. Para além das fontes tradicionais de obtenção de evidência de validade (conteúdo da medida, relações com outros construtos – validade convergente, discriminante, concorrente e preditiva), bem como estruturação interna – dimensionalidade, a validade

compreende ainda evidência baseada nos processos de respostas dos indivíduos (e.g., processos cognitivos despoletados durante a resposta aos itens) e consequências da utilização das medidas (utilização pretendida e mau uso).

FIDELIDADE

A fidelidade refere-se essencialmente à consistência de um instrumento de medida na avaliação do construto subjacente ao mesmo, ou seja, o grau com que os resultados obtidos com o instrumento são precisos e livres de erro de medida (Gulliksen, 1950). Assim, se na ausência de crescimento, aprendizagem ou maturação de um indivíduo, lhe for aplicado o mesmo instrumento de medida em momentos diferentes, devemos obter, sensivelmente, os mesmos resultados (Gregory, 2013). Pensemos num instrumento de medida banal: uma balança. Se estiver bem calibrada, e nos pesarmos a cada 5 minutos no mesmo dia, a balança vai apresentar o mesmo peso de forma consistente, com pequenas oscilações entre medições (ou pequenos erros de medida). Um instrumento de medida em psicologia deverá, idealmente, comportar-se da mesma forma, ainda que na prática isso seja extremamente raro (dada a natureza dos construtos sob medida – ver secção da validade de um instrumento de medida).

De acordo com a teoria clássica dos testes (Novick, 1966), o resultado obtido com a aplicação de um instrumento de medida equivale ao “score verdadeiro” presente (e latente) num examinando (o “true score”; Spearman, 1904) adicionado a erro associado à medida. Por exemplo, se fizermos sequencialmente vários testes de matemática, será esperado que o nosso resultado seja diferente em todos eles, fruto de diversas condicionantes (cansaço, duração da prova, etc). A estas pequenas diferenças chamamos “erros de medida”. No entanto, o nosso “resultado verdadeiro de conhecimento matemático” não terá sofrido grandes alterações entre medições, ainda que os resultados sejam diferentes. Este “score verdadeiro” não é acessível diretamente, e obriga-nos a utilizar instrumentos de medida (i.e., os testes de matemática). Idealmente, a pontuação obtida num instrumento de medida seria igual à pontuação verdadeira, caso o instrumento fosse desprovido de qualquer tipo de erro de medida. No entanto, subtrair totalmente o erro de medida de um instrumento é essencialmente impossível (Schmidt & Hunter, 1996). Assim sendo, e uma vez que a fidelidade está intimamente relacionada com a ideia de erro de medida, podemos entendê-la como expressando a distância entre o resultado obtido com o instrumento de medida e o “score verdadeiro” do indivíduo. Assim, quanto mais alta a fidelidade menor esta distância e maior a consistência do instrumento para medir a pontuação verdadeira do construto latente que lhe está na base.

Tal como acontece na validade, existem várias formas de avaliar a fidelidade de um instrumento de medida. Normalmente, estas são divididas em medidas de consistência temporal (e.g., método teste-reteste; Guttman, 1945; Bland & Altman, 1986) e medidas de consistência interna (e.g., alfa de Cronbach; Cronbach, 1951).

Enquanto **consistência temporal**, a fidelidade remete-nos para a capacidade do instrumento de medida se mostrar relativamente invulnerável aos efeitos da passagem do tempo nos examinados (e.g., aprendizagem, maturação) em diferentes aplicações do mesmo, o que pode introduzir erro na correta medida do construto (Lopes & Pinto, 2016). A avaliação da consistência temporal é, normalmente, realizada de duas formas: pelo método de teste-reteste ou pelo método das formas paralelas. No método de teste-reteste, verificamos se um mesmo instrumento de medida apresenta consistência de resultados em duas ou mais aplicações do mesmo separadas temporalmente. Isto é, se correlacionarmos as pontuações obtidas numa aplicação e noutra devemos verificar uma associação razoável ($r > .75$; Portney & Watkins, 2015). Quanto ao método das “formas paralelas” (Parameswaran et al., 1979), utiliza-se, normalmente, um instrumento que possua formas “quase idênticas” (e.g., o teste de inteligência de Raven – SPM – possui diversas “formas paralelas” do “teste mãe”; avaliam exatamente o mesmo construto, mas possuem uma “aparência diferente” entre si) e em que cada uma delas é aplicada em momentos separados no tempo. Este método é útil para obviar possíveis ameaças que se colocam ao método teste-reteste, tais com a aprendizagem, familiarização com o instrumento de medida, entre outros.

No que toca à fidelidade como **consistência interna**, as diferentes metodologias existentes procuram analisar a consistência dos instrumentos de medida numa única aplicação do mesmo. A consistência interna reflete, portanto, a coerência (ou redundância) dos componentes de um instrumento de medida (McCrae et al., 2011). Para avaliar este tipo de fidelidade existem vários métodos, dos quais destacamos quatro. O método de *split-half* (ou “partição ao meio”; Rulon, 1939) diz respeito à divisão dos itens de um instrumento de medida em duas partes iguais, posteriormente correlacionando os itens de cada uma das partes para verificar a existência de correlações elevadas (ou seja, a homogeneidade interna de cada uma das partes). A aplicação da fórmula de Spearman-Brown (Spearman, 1910; Brown, 1910) permite-nos derivar um único coeficiente de fidelidade *split-half* que resume as diferentes correlações encontradas nas duas metades do instrumento de medida. Um outro método de cálculo da consistência interna de um instrumento de medida é a determinação do alfa de Cronbach (Cronbach, 1951), um dos indicadores de fidelidade mais populares na Psicologia e em diferentes ciências sociais. Este indicador de fidelidade permite a obtenção de um único coeficiente sem necessitarmos de realizar divisões ou outro tipo de operações sobre o instrumento de medida. Podemos entendê-lo como uma média de todos os coeficientes *split-half* que possam ser obtidos a partir da partição de um teste em duas partes (Cronbach, 1951). Uma variante do coeficiente alfa de Cronbach é o coeficiente de Kuder-Richardson (Kuder & Richardson, 1937), ou KR-20, utilizado para determinar a fidelidade de instrumentos de medida cujos itens são medidos através de escalas de tipo nominal dicotómico (e.g., “sim” / “não”; “verdadeiro” / “falso”).

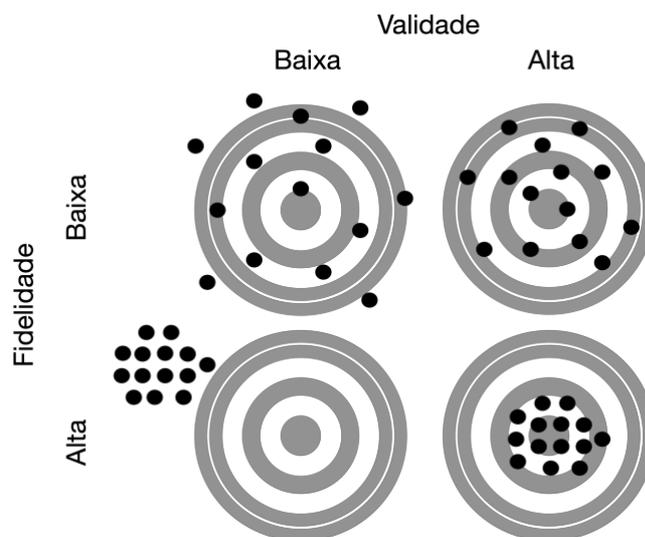
Quando pretendemos determinar a fidelidade de instrumentos de medida que geram resultados qualitativos (e.g., testes projetivos), é comum utilizar-se um coeficiente de fidelidade denominado **acordo inter-juízes** (Cohen, 1960). Este coeficiente permite determinar a consistência da avaliação de dois juízes, utilizando uma grelha de cotação, de produções escritas

ou verbais de examinandos (como é o caso de determinadas respostas em instrumentos de medida psicológica, por exemplo as baterias de inteligência de Wechsler – WAIS ou WISC; Wechsler, 1939). Caso esta tarefa seja realizada por mais de dois juízes, terá de se utilizar um coeficiente de acordo diferente, nomeadamente o coeficiente K de Cohen (Cohen, 1960). Note-se, contudo, que se trata de um cálculo de fidelidade que incide particularmente na avaliação da qualidade da grelha de cotação. De facto, o que se está a analisar é a consistência de avaliações de juízes baseadas nestas grelhas. Naturalmente que erros de cotação ou pontuação contribuem para o erro da medida no seu todo e, a limite, causam impacto na sua fidelidade.

AS RELAÇÕES ENTRE VALIDADE E FIDELIDADE

Como descrevemos anteriormente, a validade e a fidelidade são características psicométricas fundamentais na determinação da qualidade de um instrumento de medida, apontando para diferentes características desejáveis que estes devem revestir. Contudo, e apesar de distintas, estas características interagem e impactam no instrumento de medida, na maioria das situações de forma indesejável. Na figura 1, apresenta-se uma representação gráfica destas interações. Os “alvos” representam o construto subjacente à medida e os pontos negros os “itens” da medida que a operacionalizam.

FIGURA 1 | RELAÇÕES ENTRE VALIDADE E FIDELIDADE



Nota. Adaptado de Trochim (2006).

No canto superior esquerdo da Figura 1, podemos observar a situação mais indesejável para um instrumento de medida – baixa validade e baixa fidelidade. Neste caso, verifica-se uma grande dispersão nos itens quanto à sua capacidade para operacionalizar corretamente o construto de

base (não está garantido que meça o que se pretende medir), bem como a sua consistência não está garantida. No canto inferior esquerdo, novamente uma situação indesejável para um instrumento de medida – ainda que os itens meçam consistentemente “qualquer coisa”, não garantem que meçam o construto que se pretende medir. No canto superior direito encontramos um instrumento de medida em que se garante alguma validade na medida do construto de base, mas não estando garantida a consistência com que o fazemos. Finalmente no canto inferior direito, encontramos uma medida psicometricamente válida – elevada consistência na medida do construto e garantia de operacionalização correta do mesmo.

RECURSOS

De seguida listamos alguns recursos que podem ser úteis para acompanhar a aprofundar o seu conhecimento nos tópicos da validade e da fidelidade. Deixamos ainda algumas revistas que, ao longo dos anos, têm vindo a publicar artigos fundamentais nestas áreas.

| Publicações

As publicações seguintes foram fundamentais no estabelecimento de definições e metodologias de avaliação da validade e fidelidade dos instrumentos de medida psicológicos. Algumas destas publicações (assinaladas com asterisco) são fundadoras nesta área e criaram o “fiel da balança” a partir do qual derivaram a maior parte das restantes definições de validade e fidelidade:

. Standards para a testagem em Educação e Psicologia (*Standards for Educational and Psychological Testing*; AERA et al. 1999)*;

. Comissão Internacional de Testes (*International Test Commission [ITC], 2000*);

. Princípios de Validação e Utilização de Procedimentos para Seleção de Pessoal (*Principles for the Validation and Use of Personnel Selection Procedures*; Society for Industrial and Organizational Psychology, 2003)*;

. Revisão de testes para o Mental Measurement Yearbook (*Test Reviewing for the Mental Measurement Yearbook at the Buros Center for Testing*; Carlson and Geisinger, 2012);

. Modelo de revisão de testes da European Federation of Psychologists’ Association’s (EFPA review model; Evers et al. 2013)*;

. Standards de Consenso para a Seleção de Instrumentos de Medida na área da Saúde (Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN); Mokkink et al. 2010a).

. Revelle, W., & Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395–1411. <https://doi.org/10.1037/pas0000754>

. Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients.

Organizational Research Methods, 19(4), 651-682. <https://doi.org/10.1177/1094428116656239>

| Revistas científicas

De seguida, apresentam-se algumas revistas que têm vindo a publicar, ao longo da sua história, artigos fundamentais no estudo e compreensão da validade e fidelidade:

- . *Applied Psychological Measurement*, Sage Publications
- . *Assessment*, Sage Publications
- . *Educational and Psychological Measurement*, Sage Publications
- . *Psychological Assessment*, APA
- . *Psychological Methods*, APA
- . *Psychometrika*, Springer

CONSIDERAÇÕES FINAIS

Como referimos neste artigo, a validade e a fidelidade são duas características psicométricas fundamentais que devem revestir qualquer instrumento de medida psicológica. Determinar a validade e a fidelidade possibilita, não só, a obtenção de instrumentos válidos e fiáveis para medir o construto que subjaz a qualquer instrumento de medida, como, muitas vezes, permite que futuros/as investigadores/as tenham confiança nestas medidas e não se vejam obrigados a escolher entre investir o tempo necessário para validarem os instrumentos que desejam utilizar ou utilizarem instrumentos que não estejam validados.

Validar e determinar a fidelidade de um instrumento de medida não é uma tarefa simples, principalmente para quem o está a fazer pela primeira vez. São processos complexos, envolvem vários tipos de metodologias, e não têm um final tão delineado como noutros tipos de estudos. Ainda assim, este tipo de trabalho é fundamental para a ciência psicológica em geral, e para os instrumentos de medida em particular. São estudos que não tendem a contribuir diretamente para o que sabemos sobre um determinado tema, mas que possibilitam uma progressão mais célere na investigação sobre o mesmo.

SOBRE OS AUTORES

[Diniz Lopes](#) é doutorado em Psicologia Social e das Organizações (2007) pelo Iscte-Instituto Universitário de Lisboa. É Professor Associado com Agregação no Iscte, Membro do grupo de investigação Behavior, Emotion and Cognition (BEC). Tem ampla experiência no desenvolvimento e validação de instrumentos e coordena a UC de Psicometria.

[Tiago Roxo Aguiar](#) é estudante de doutoramento em Psicologia no Iscte-Instituto Universitário de Lisboa e assistente de investigação no Grupo de Comportamento Organizacional e Recursos Humanos da BRU-Iscte - Business Research Unit (IBS). É membro do grupo de investigação Behavior, Emotion and Cognition (BEC) e tem lecionado a UC de Psicometria.

REFERÊNCIAS

- AERA, APA, & NCME (1999). *The standards for educational and psychological testing*. AERA Publications Sales
- Birnbaum A. (1968). Some latent trait models and their use in inferring and examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp.17-20). Addison Wesley
- Bland, J.M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327(8476), 307-310
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061-1071
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296-322. <https://doi.org/10.1111/j.2044-8295.1910.tb00207.x>
- Campbell D. T. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22. <https://doi.org/10.1037/h0046743>
- Chan, E. K. H. (2014). Standards and guidelines for validation practices: Development and evaluation of measurement instruments. In B. Zumbo & E. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 9-24). Springer. https://doi.org/10.1007/978-3-319-07794-9_2
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. <https://doi.org/10.1007/BF02310555>
- Cronbach, L. J. (1988). Five perspectives on validation argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Erlbaum
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Freeman, F. S. (1990). *Teoria e prática dos testes psicológicos*. Fundação Calouste Gulbenkian
- Furr, R. M. (2017). *Psychometrics: An introduction* (3rd ed.). Sage
- Gonthier C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications*, 7(1), 11. <https://doi.org/10.1186/s41235-021-00350-w>
- Gregory, R. J. (2013). *Psychological testing: History, principles, and applications* (7th ed). Pearson
- Gulliksen, H. (1950). *Theory of mental tests*. Wiley.
- Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282 (1945). <https://doi.org/10.1007/BF02288892>
- Holden, R.R. (2010). Face Validity. In I.B. Weiner & W.E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology*. <https://doi.org/10.1002/9780470479216.corpsy0341>
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, 103, 219-230
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lopes, D. & Pinto, I. R. (2016). Conhecer os métodos quantitativos e qualitativos e suas aplicações em ciências sociais e humanas. In M. V. Garrido & M. Prada (Coord.), *Manual de competências académicas* (pp. 281-342). Sílabo
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and*

meaning. Routledge

- McCrae R. R., Kurtz J. E., Yamagata S., Terracciano A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15(1), 28-50. doi: 10.1177/1088868310366253
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Parameswaran, R., Greenberg, B. A., Bellenger, D. N., & Robertson, D. H. (1979). Measuring reliability: A comparison of alternative techniques. *Journal of Marketing Research*, 16(1), 18-25. <https://doi.org/10.1177/002224377901600103>
- Portney L. G. & Watkins M. P. (2015). *Foundations of clinical research: Applications to practice* (3rd ed.). F.A. Davis Company
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99-103.
- Ryan, J. J., Sattler, J. M., & Lopez, S. J. (2000). Age effects on Wechsler Adult Intelligence Scale-III subtests. *Archives of Clinical Neuropsychology*, 15(4), 311–317. [https://doi.org/10.1016/S0887-6177\(99\)00019-0](https://doi.org/10.1016/S0887-6177(99)00019-0)
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199–223. <https://doi.org/10.1037/1082-989X.1.2.199>
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–293. <https://doi.org/10.2307/1412107>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Trochim, W. M. K. (2006). *The qualitative debate*. Research Methods Knowledge Base. <http://www.socialresearchmethods.net/kb/qualmeth.ph>
- Wechsler, D. (1939). *The measurement of adult intelligence*. Williams & Witkins
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65–82). IAP – Information Age Publishing