

# 1 Fiabilidade e Validade

Muitos estudos na área das ciências da saúde (incluindo fala e audição) envolvem a realização de testes (questionários) e/ou medição de características fisiológicas. Associados a estes testes e medições encontramos dois aspectos de relevo: a **fiabilidade** e a **validade**.

Fiabilidade significa precisão do método de medição e pode ser averiguada através da análise da consistência ou estabilidade desse método. Um método (teste ou instrumento de medida) fiável não deve produzir resultados significativamente diferentes se for repetido sobre o mesmo indivíduo.

Um teste ou instrumento de medida dizem-se válidos se conseguirem traduzir de forma correcta a grandeza que pretendem medir. Por exemplo o número de anos e meses de vida de uma pessoa constitui uma medida válida da sua idade, o mesmo já não acontece se usarmos a sua estatura para medir a idade.

Enquanto a fiabilidade diz respeito à consistência ou estabilidade de uma medida, a validade diz respeito à sua veracidade. Uma medida pode ser muito fiável (precisa) mas pode estar errada e portanto ser inválida. Portanto fiabilidade não implica validade mas é um requisito para avaliar a validade. Ou seja, uma medida para ser válida deve antes de mais ser fiável. Consequentemente devemos primeiro avaliar a fiabilidade dos instrumentos (ou métodos) de medida e só depois avaliar a validade dos mesmos.

## 1.1 Fiabilidade

Existem diversos factores que influenciam a fiabilidade de um método. Por exemplo, a pessoa sobre a qual se está a efectuar as medições pode ter reacções diferentes de dia para dia. Por outro lado o método pode apenas medir uma parte do fenómeno de interesse e não servir para caracterizar de forma global esse fenómeno.

Existem várias formas de averiguar a fiabilidade de um método (teste ou medição):

1. Pode-se repetir exactamente o mesmo teste ou medição sobre os mesmos indivíduos e comparar os resultados. Este processo é designado por *test-retest*.
2. Podem-se aplicar dois testes ou medições supostamente equivalentes e comparar os resultados.
3. Podem-se subdividir os testes ou medições em duas partes equivalentes (nem sempre isto é possível) e examinar a consistência dos resultados nessas duas partes. Este processo é designado por *split-half reliability*.

Existem basicamente dois processos para quantificar a fiabilidade de um teste ou medição:

1. o primeiro consiste em avaliar a variabilidade das medidas após sua repetição sobre os mesmos indivíduos. A medida de variabilidade mais utilizada é o desvio padrão. No entanto, é raro podermos aplicar o mesmo teste ou medição aos mesmos indivíduos repeti-

das vezes para podermos obter uma estimativa do desvio padrão. Assim, desenvolveram-se métodos para obter estas estimativas sem ter que se repetir os testes ou medições sobre os mesmos indivíduos;

2. o segundo e o mais comum consiste no cálculo de coeficientes de fiabilidade a partir de medições repetidas ou de comparações *split-half*. Tipicamente, com base em dois conjuntos de medidas (ou porque o mesmo teste foi aplicado duas vezes sobre o mesmo indivíduo ou porque o teste foi dividido em duas partes) calcula-se um coeficiente de correlação adequado ao tipo de medidas em causa.

Uma medida pode então dizer-se fiável se o desvio padrão for reduzido ou se o coeficiente de fiabilidade (correlação) for elevado.

Uma forma de fiabilidade que muitas vezes é avaliada é a chamada fiabilidade inter-observador (*inter-observer reliability* ou *inter-judge reliability*). Neste caso pretende-se averiguar se diferentes observadores (avaliadores) ao utilizarem os mesmos métodos de medição (testes ou medições) obtêm resultados consistentes. Trata-se de um caso particular de fiabilidade descrita acima.

As medidas de associação (correlação) mais utilizadas para medir a fiabilidade de um teste ou medição encontram-se descritas no capítulo da Associação entre variáveis.

## 1.2 Validade

A validade de um teste (ou instrumento de medida) não é simples de averiguar na maioria das situações. Duma forma geral podemos dizer que quanto mais directa for a forma de medir o fenómeno em causa mais válido será o método utilizado. Por exemplo, se observarmos o que uma pessoa come durante uma refeição temos uma medida mais válida do seu consumo de calorías do que se lhe perguntarmos à posteriori o que comeu. Para termos medidas válidas é portanto conveniente considerar diferentes métodos de medição e procurar avaliar a sua validade comparativa. Infelizmente, porque é difícil avaliar a validade dos métodos, muitas vezes assume-se a validade até que alguém afirme em contrário.

Schweigert (1994) distingue três tipos de validade: validade de **critério** (*criterion validity*), **conceptual** (*construct validity*) e **facial** (*face validity*). Note-se no entanto que esta classificação não é única e existem outras formas de categorizar os tipos de validade de um método (ver por exemplo Ventri & Schiavetti (1986).

1. A **validade de critério** é o grau com que um método de medição se correlaciona com outros métodos já estabelecidos para o mesmo fenómeno. Existem dois tipos de validade de critério: **predictiva** (*predictive validity*) e **concorrente** (*concurrent validity*).

- 1.1. Validade predictiva é o grau com que o resultado de um teste (ou medida) prevê o comportamento futuro do indivíduo. Por exemplo o resultado de um teste à inteligência (IQ) pode prever o sucesso escolar de uma criança. Se guardarmos os

resultados dos testes de inteligência e os correlacionarmos com os resultados escolares obtidos posteriormente podemos ter uma ideia do grau de concordância das duas medidas. Em geral para avaliar a validade preditiva de um método é necessário conduzir um estudo paralelo onde se guardam algumas das medidas do fenómeno a prever e posteriormente se correlacionam com outras medidas obtidas directamente após a ocorrência do mesmo. Isto requer disponibilidade temporal (há que esperar pela ocorrência do fenómeno) e disponibilidade de unidades experimentais destinadas exclusivamente a este fim.

- 1.2. Validade concorrente é o grau com que um novo método se correlaciona com outro já existente e tido como válido.
2. A validade conceptual relaciona-se com o método de medição de um conceito (ou fenómeno). Se um método se destina a medir um certo conceito, então ele deve correlacionar-se fortemente com outros métodos existentes para o mesmo conceito. No entanto, se o investigador concluir que o método também se correlaciona com métodos de medição de outros conceitos que não se consideram relacionados com o conceito em causa, então é necessário ter cuidado com o que de facto se está a medir.
3. Validade facial diz respeito ao grau com um método aparenta medir aquilo que de facto pretende medir. Por exemplo, um exame sobre literatura portuguesa terá uma forte validade facial se for baseado em excertos de literatura portuguesa. Este tipo de validade parece ser o que tem menos importância na investigação científica. Isto acontece porque um método pode ter elevada validade facial e no entanto não possuir nenhum outro tipo de validade ou fiabilidade. Por outro lado, um método pode não ter validade facial e no entanto ser bastante válido e fiável. No que respeita a inquéritos, é usual os inquiridos exigirem algum tipo de validade facial sob o risco de não responderem ao inquérito por acharem que não tem nada a ver com a questão em causa.

Do que acabámos de descrever podemos concluir que para medir a maioria dos conceitos de validade há que aplicar medidas de associação (ou correlação) tais como as propostas para medir fiabilidade.

Também encontramos o conceito de **validade** associado a estudos observacionais (em vez de testes ou instrumentos de medida específicos) e nesse caso há que classificar a validade em dois tipos: **interna** e **externa**.

A validade interna depende da capacidade do estudo realmente responder às questões propostas inicialmente. Ela mede até que ponto os resultados do estudo são o produto das variáveis que foram seleccionadas, observadas e medidas e não o fruto de outras variáveis que não foram tratadas.

A validade externa mede até que ponto os resultados obtidos pelo estudo podem ser generalizados para outras situações com outros indivíduos.

Um estudo pode ter validade interna mas não ter validade externa. Já um estudo que não tenha validade interna nunca poderá ter validade externa.

## 2 Sensibilidade e especificidade

Para terminar iremos referir dois conceitos que surgem por vezes na literatura das ciências da saúde incluindo as da fala e da audição: sensibilidade e especificidade.

Muitas vezes aplica-se um teste de diagnóstico para averiguar a presença de determinada patologia num paciente. Quando o teste produz um resultado positivo conclui-se a favor da presença e quando produz um resultado negativo a favor da ausência. Estas conclusões podem obviamente estar erradas e interessa avaliar a probabilidade de estes erros ocorrerem.

A *sensibilidade* de um teste mede a sua capacidade de detectar a presença da doença correctamente. Chama-se *positive predictive value* à percentagem de resultados positivos verdadeiros que o teste fornece.

A *especificidade* de um teste é a capacidade de este detectar a ausência da patologia. Chama-se *negative predictive value* à percentagem de resultados negativos verdadeiros que o teste fornece.

## Bibliografia

- Schweigert, W. (1994) *Research methods and statistics for psychology*, Brooks/Cole Publishing Company.
- Ventri, I. & Schiavetti, N. (1986) *Evaluating Research in Speech Pathology and Audiology*, Macmillan.
- Marks, R. (1994) *Designing a Research Project: the art of doing Science*, unpublished (Prof. Ana Mendes).
- Schavelson, R. (1988) *Statistical Reasoning for the Behavioral Sciences*, Allyn and Bacon.