

*Traga o livro  
sobre testes*

UFRGS - FACULDADE DE EDUCAÇÃO

DEPARTAMENTO DE ESTUDOS ESPECIALIZADOS/LABORATÓRIO DE MEDIDAS EDUCACIONAIS

CURSO DE EXTENSÃO EM AVALIAÇÃO E TECNOLOGIA DE TESTES

---

DOCUM. Nº 7

PROCEDIMENTOS CIENTÍFICOS  
NA  
CONSTRUÇÃO DE TESTES

PORTO ALEGRE, MAIO/JUNHO/1981

## VALIDADE

- A validade de um instrumento é determinada pelo grau com que mede o que pretende medir.
- A validade de um instrumento está relacionada aos seus propósitos.
- Os 4 tipos de validade são: (1) de conteúdo, (2) *prediz atuação futura* preditiva, (3) concorrente e (4) de construto.
- Um instrumento de avaliação apresenta validade de conteúdo quando:
  - os resultados estão em consonância com os objetivos
  - a amostra é representativa do universo a ser medido
- A validade de conteúdo resulta do julgamento de especialistas em relação à representatividade dos conteúdos e à relevância dos objetivos.
- O controle da validade de conteúdo, em parte, pode ser realizada a partir da tabela de especificações.
- A comparação entre levantamento de objetivos e conteúdos pre-estabelecidos e de objetivos e conteúdos emergentes do instrumento de avaliação possibilita verificar as discrepâncias e corrigi-las, garantindo maior validade.

## F I D E D I G N I D A D E

- A fidedignidade de um instrumento de avaliação é determinada pela coerência com/que mede aquilo que mede.
- A fidedignidade de um instrumento de avaliação está relacionada às variações das medições em torno do valor verdadeiro.
- Um instrumento de avaliação apresenta alta fidedignidade, quando as variações das medições em torno do valor verdadeiro são pequenas.
- Um instrumento de avaliação apresenta baixa fidedignidade, quando as variações das medições em torno do valor verdadeiro são altas.
- O desvio padrão é o elemento estatístico que expressa o grau das variações das medições em torno do valor verdadeiro.
- Os procedimentos para determinar o grau de fidedignidade de um instrumento de avaliação consistem na:

- aplicação do instrumento em duas ocasiões a um mesmo grupo de indivíduos, que representa o grupo para o qual é destinado;

- aplicação de duas formas equivalentes do instrumento, em uma ou duas ocasiões, ao mesmo grupo de indivíduos que representa o grupo para o qual é destinado;

- divisão do instrumento em duas partes e aplicação numa única ocasião; (pares x ímpares)

- utilização de técnicas correlacionais para verificar o coeficiente de correlação.

• O coeficiente de correlação expressa o grau de fidedignidade do instrumento de avaliação.

### Estabilidade e consistência interna

Consistência interna:

$$KR_{20} = \frac{K}{K-1} \left( 1 - \frac{\sum p^2}{\sigma^2} \right)$$

(aceita ou erra)

$$KR_{21} = \frac{K}{K-1} \left[ 1 - \frac{\bar{x} \cdot \left( 1 - \frac{\bar{x}}{K} \right)}{\sigma^2} \right]$$

(para TRN)

(admite valores fracionários)

$K$  = n.º de itens  
 $\sigma$  = desvio padrão  
 $\sigma^2$  = variância  
 $p$  = razão de acertos  
 $q$  = razão de erros  
 $\bar{x}$  = média

dupla aplicação do teste

formas paralelas

correlação das duas metades

+1,00 → valor máximo  
0,70 → é bom

$$r_t = \frac{rt \cdot s^2 + (x - c)^2}{s^2 + (\bar{x} - c)^2} \quad ($$

(para TRC)

$s^2$  = variância

$r_t$  → fidedignidade obtida por qq. método.

$c$  → critério em termos de n.º de acertos  
(se foi 80% de 20 acertos →  $c = 16$ )

Fidedignidade { precisa  
consistêcia

Fid. é o fundamento da validade.

Validade — o teste mede o que se quer medir?

### ANÁLISE DE ITENS

- A análise de itens consiste na utilização de procedimentos matemáticos no cálculo dos índices de dificuldades e de discriminação.
- O cálculo dos índices de discriminação e dificuldade é efetuado depois da aplicação do instrumento de avaliação.
- Os índices de discriminação e dificuldade são calculados a partir dos escores dos indivíduos de grupos superiores (27%) e inferiores (27%).
- O índice de discriminação é o quociente do número de erros do grupo inferior menos o número de erros do grupo superior pelo número de sujeitos de qualquer um dos grupos.
- O índice de dificuldade é o coeficiente do número de acertos do grupo superior mais o número de acertos do grupo inferior pelo número de sujeitos dos dois grupos.
- A faixa de valores do índice de dificuldade oscila entre 0,00 e 1,00.
- A faixa de valores do índice de discriminação oscila entre -1,00 e +1,00.
- Um item é mais fácil quanto maior o valor do índice de dificuldade (quanto mais próximo de 1,00).
- Um item é mais difícil quanto menor o valor do índice de dificuldade (quanto mais próximo de 0,00).
- O poder de discriminação de um item é baixo quando o índice de dificuldade situa-se nos extremos próximos a 0,00 ou 1,00.
- A maioria dos itens deve apresentar índices de dificuldades na faixa de 0,40 a 0,60.
- A faixa do índice de dificuldade deve oscilar entre 0,20 e 0,80.
- Um item com índice de discriminação negativo deve ser reformulado ou suprimido.
- O poder de discriminação de um item é maior quando o índice de dificuldade aproxima-se de 0,50.

EXEMPLOs

ITEM X

TOTAL INDIVÍDUOS: 65  
 27% Grupo Superior(GS):18  
 27% Grupo Inferior(GI):18

Respostas	A	B <sup>(1)</sup>	C	D	Erro <sup>(2)</sup>	Omissão <sup>(3)</sup>	Dific.	Discrim.
27% indivíduos com maior nº - de acertos - (total) (GS)	1	17	-	-	-	-	0,72	0,44
27% indivíduos com maior nº - de erros (total) (GI)	5	9	3	1	-	-		

- (1) Resposta certa
- (2) Mais de uma respostas assinalada
- (3) Resposta em branco

INDICE DE DIFICULDADE

$$\frac{NaGS + NaGI}{NGS + NGI} = \text{índ.dific.}$$

$$\frac{17 + 9}{18 + 18} = 0,72$$

INDICE DE DISCRIMINAÇÃO

$$\frac{NeGI - NeGS}{NGS} = \text{índ.discrim.}$$

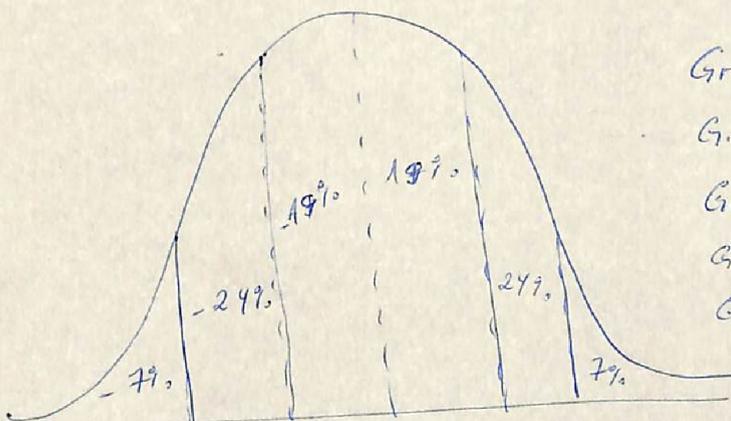
$$\frac{(5 + 3 + 1) - 1,8}{18} = 0,44$$

BANCO DE ITENS

Depois dos itens analisados, aqueles que foram considerados satisfatórios em termos de índice de discriminação e de dificuldade, assim como em termos de construção, podem ser guardados para utilização posterior.

Convém identificar o item quanto ao objetivo ou comportamento que exige, ao conteúdo e aos índices de discriminação e de dificuldade.

Depois de algum tempo o professor terá um número razoável de itens, possibilitando uma organização rápida da prova; também poderá substituí-los por outros, à medida em que ocorrerem reformulações em seu planejamento e sua organização do ensino.



Gr. Sup. - 7% (+1,5σ e mais) <sup>devido padrão</sup>  
 G. M. sup - 24% (+0,5σ até +1,49σ)  
 Gr. M. - 38% (-0,5σ até 0,49σ)  
 Gr. M. inf - 24% (-1,5σ até -0,51σ)  
 G. inf - 7% (-1,51σ até ...)

Considerando-se que a educação é um processo intencional, onde se pretende que o aluno atinja todos ou quase todos os objetivos, que sentido teria os índices de discriminação e de dificuldade?

pequeno grupo de alunos com nota baixa

grande n.º de alunos com notas altas

PROCEDIMENTOS CIENTÍFICOS NA ELABORAÇÃO DE TESTES

Fonte : Curso de Extensão em Avaliação e  
Tecnologia de Testes - Faculdade  
de Educação - UFRGS

V A L I D A D E

- . A validade de um instrumento é determinada pelo grau com que mede o que pretende medir.
- . A validade de um instrumento está relacionada aos seus propósitos.
- . Os 4 tipos de validade são : (1) de conteúdo, (2) preditiva, (3) concorrente e (4) de construto.
- . Um instrumento de avaliação apresenta validade de conteúdo quando :
  - os resultados estão em consonância com os objetivos;
  - a amostra é representativa do universo a ser medido.
- . A validade de conteúdo resulta do julgamento de especialistas em relação à representatividade dos conteúdos e à relevância dos objetivos.
- . O controle da validade de conteúdo, em parte, pode ser realizada a partir da tabela de especificação.
- . A comparação entre levantamento de objetivos e conteúdos pré estabelecidos e de objetivos e conteúdos emergentes do instrumento de avaliação possibilita verificar as discrepâncias e corrigi-las, garantindo maior validade.

F I D E D I G N I D A D E

- . A fidedignidade de um instrumento de avaliação é determinada pela coerência com que mede aquilo que mede.
- . A fidedignidade de um instrumento de avaliação está relacionada às variações das medições em torno do valor verdadeiro.
- . Um instrumento de avaliação apresenta alta fidedignidade quando as variações das medições, em torno do valor verdadeiro, são pequenas.
- . Um instrumento de avaliação apresenta baixa fidedignidade quando as variações das medições, em torno do valor verdadeiro, são altas.

. O desvio padrão é o elemento estatístico que expressa o grau das variações das medições em torno do valor relativo.

. Os procedimentos para determinar o grau de fidedignidade de um instrumento de avaliação consistem na :

- aplicação do instrumento em duas ocasiões a um mesmo grupo de indivíduos, que representam o grupo para o qual é destinado;

- aplicação de duas formas equivalentes do instrumento, em uma ou duas ocasiões, ao mesmo grupo de indivíduos que representa o grupo para o qual é destinado;

- divisão do instrumento em duas partes e aplicação numa única ocasião;

- utilização de técnicas correlacionais para verificar o coeficiente de correlação.

. O coeficiente de correlação expressa o grau de fidedignidade do instrumento de avaliação.

### ANÁLISE DE ITENS

. A análise de itens consiste na utilização de procedimentos matemáticos no cálculo dos índices de dificuldade e de discriminação.

. O cálculo dos índices de discriminação e de dificuldade é efetuado depois da aplicação do instrumento de avaliação.

. Os índices de discriminação e de dificuldade são calculados a partir dos escores dos indivíduos dos grupos superiores (27%) e inferiores (27%).

. O índice de discriminação é o quociente entre o número de erros do grupo inferior menos o número de erros do grupo superior e o número de sujeitos de qualquer um dos grupos.

. O índice de dificuldade é o quociente do número de acertos do grupo superior mais o número de acertos do grupo inferior pelo número de sujeitos dos dois grupos

. A faixa de valores de índice de dificuldade oscila entre 0,00 e 1,00.

. A faixa de valores do índice de discriminação oscila entre -1,00 e 1,00.

. Um item é mais fácil quanto maior o valor do índice de dificuldade ( quanto mais próximo de 1,00 ).

. Um item é mais difícil quanto menor o valor do índice de dificuldade ( quanto mais próximo de 0,00 ).

. O poder de discriminação de um item é baixo quando o índice de dificuldade situa-se nos extremos próximos a 0,00 ou 1,00.

. A maioria dos itens deve apresentar índices de dificuldades na faixa de 0,40 a 0,60.

. A faixa do índice de dificuldade deve oscilar entre 0,20 e 0,80

. Um item com índice de discriminação negativo deve ser reformulado ou suprimido.

. O poder de discriminação de um item é maior quando o índice de dificuldade aproxima-se de 0,50.

**EXEMPLO:**

ITEM X

TOTAL INDIVÍDUOS: 65

27% Grupo superior(GS):18

27% Grupo inferior(GI):18

Respostas	A	B <sup>(1)</sup>	C	D	Erro <sup>(2)</sup>	Omissão <sup>(3)</sup>	Dific.	Discrim.
27 indivíduos com maior nº de acertos - (total) (GS)	1	17	-	-	-	-	0,72	0,44
27 indivíduos com maior nº de erros - (total) (GI)	5	9	3	1	-	-		

(1) resposta certa

(2) mais de uma resposta assinalada

(3) resposta em branco

**ÍNDICE DE DIFICULDADE**

$$\frac{RaGS + SaGI}{RGS + LGI} = \text{Índ. dific.}$$

$$\frac{17 + 9}{18 + 18} = 0,72$$

**ÍNDICE DE DISCRIMINAÇÃO**

$$\frac{RaGI - RaGS}{RGS} = \text{Índ. discrim.}$$

$$\frac{(5 + 3 + 1) - 1}{18} = \frac{8}{18} = 0,44$$

**BAIXO DE ITENS**

Depois dos itens analisados, aqueles que foram considerados satisfatórios em termos de índice de discriminação e de dificuldade, assim como em termos de construção, podem ser guardados para utilização posterior.

Convém identificar o item quanto ao objetivo ou comportamento que exige, ao conteúdo e aos índices de discriminação e de dificuldade.

Depois de algum tempo o professor terá um número razoável de itens, possibilitando uma organização rápida da prova; também poderá substituí-los por outros, à medida em que ocorrerem reformulações em seu planejamento e sua organização do ensino.