

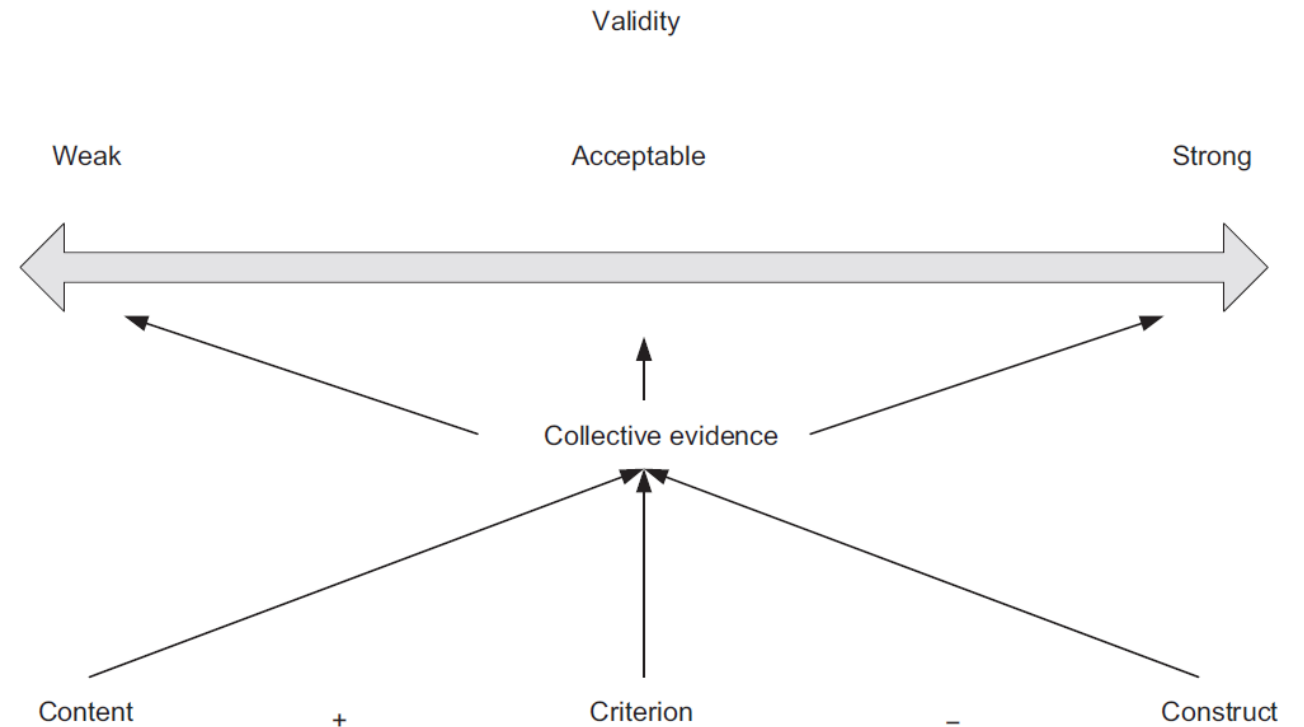
Validade

Reverendo alguns conceitos de validades

Me. Gabriel Tiraboschi

Conceito clássico de validade

- A validade é a característica do teste de quão bem o teste ou instrumento mede de maneira acurada o que ele deveria medir.
- A validade era dividida em 3 tipos.



- Evidence based on expert judgments of the relationship between parts of the test and the construct.

- External variables that include criteria that the test is expected to predict as well as relationships to other tests hypothesized to measure the related or different constructs.

- Analysis of the internal structure of a test indicates the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based.

CRITERION VALIDITY

extent to which item / scale is associated with a criterion

➤ Compare measurement of interest to a “gold standard” for construct

➤ Predictive



➤ Concurrent



➤ Retrospective



CONSTRUCT VALIDITY

extent to which item / scale assesses the intended construct

➤ Convergent

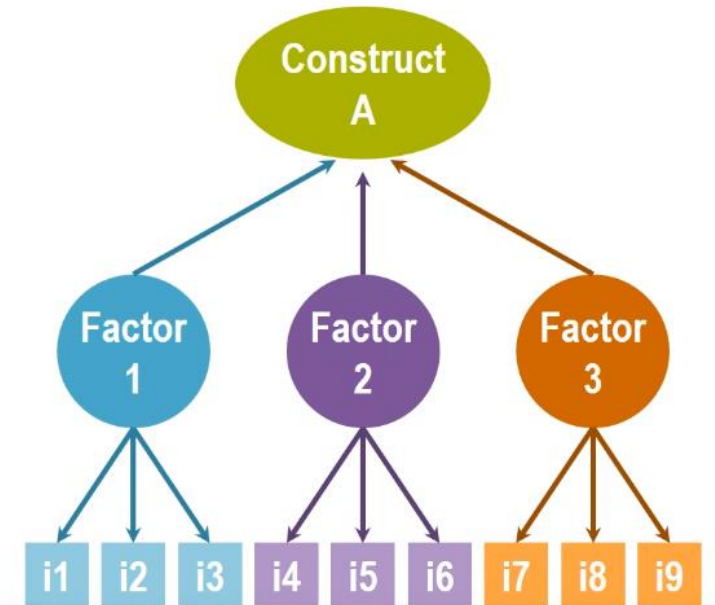
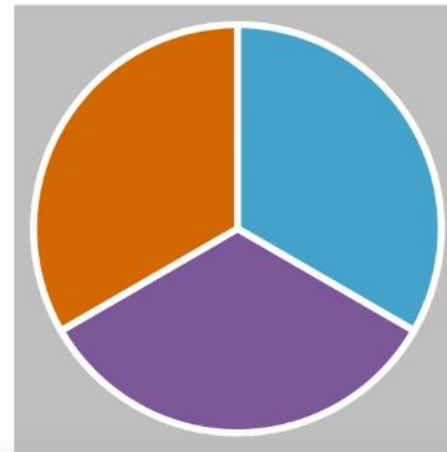


➤ Discriminant



➤ Factorial / Structural

❖ Structure of Data = Theoretical Structure



Content Validity

Does the content of the measure represents the universe of content of the concept being measured?

Items of the Measure

Meaning of the Concept

If all dimensions are not represented

LOW CONTENT VALIDITY

Some aspects will be missing

Universe of Questions



Conceito atual de validade

- Conceito unitário
- Ao invés de tipos, agora são maneiras diferentes de acumular evidências de validade
- A validade não é mais para o teste e sim para a interpretação do teste
- **The Standards for Educational and Psychological Testing** (Standards) is a joint product of the American Educational Research Association (**AERA**), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). Published collaboratively by the three organizations since 1966, it represents the gold standard in guidance on testing in the United States and worldwide.

Conceito atual de validade

TABLE 1 Tracing Historical Trends in the Concept of Validity

1974 Standards <i>(Validity as Three Types)</i>	1985 Standards <i>(Validity as Three Interrelated Types)</i>	1999 Standards <i>(Validity as Unitary Construct)</i>
Content validity	Content-related validity	Validity evidence based on test content
Criterion validity	Criterion-related validity	Validity evidence based on relations to other variables
Construct validity	Construct-related validity	Validity evidence based on response processes
		Validity evidence based on internal structure
		Validity evidence based on consequences of testing

Um pequeno parênteses: Validade de face

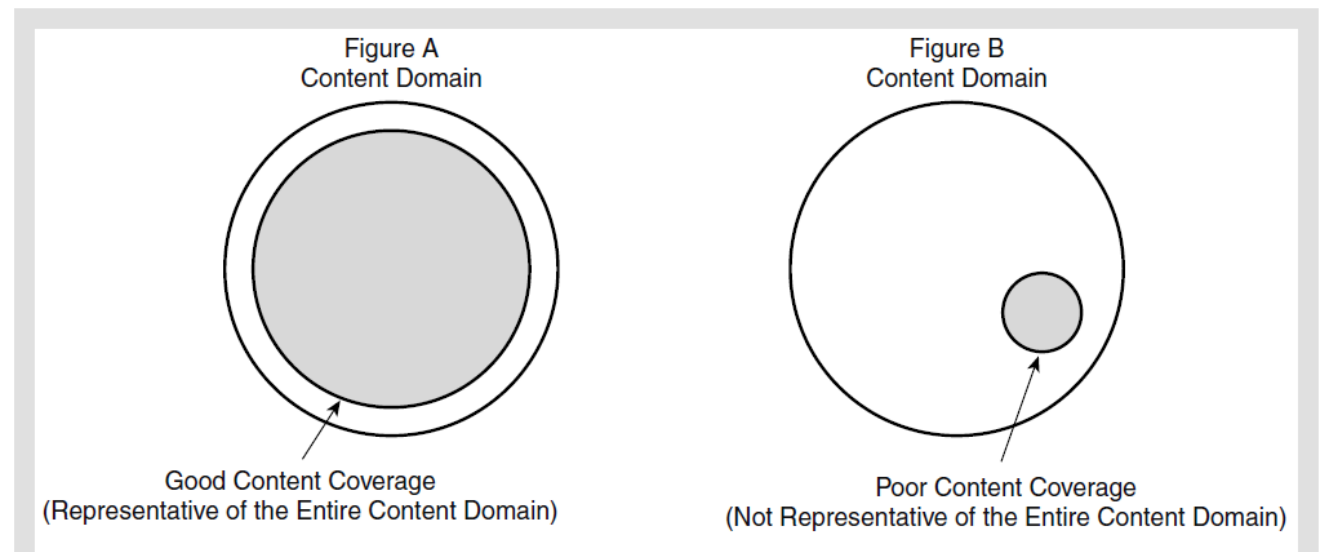
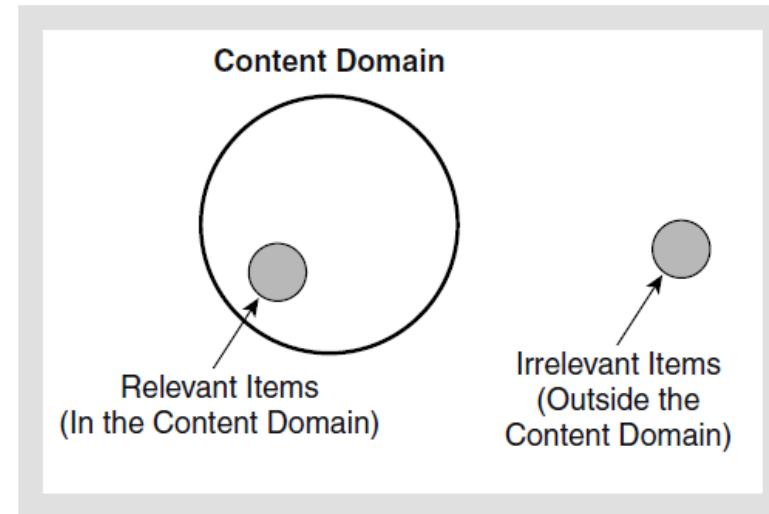
- O teste aparenta medir o que deveria medir?
- Diferente de validade de conteúdo pois é para pessoas leigas (não-experts)
- As pessoas tendem a aderir mais aos testes e se sentirem motivadas com alta validade de face

Um pequeno parênteses: fidedignidade

- Fidedignidade diz respeito a quão preciso o teste é, ou seja, quanto que suas medidas são consistentes
- A fidedignidade é um pré-requisito para que haja validade, mas só fidedignidade não é o suficiente (e.g. medir cabeça para inteligência)

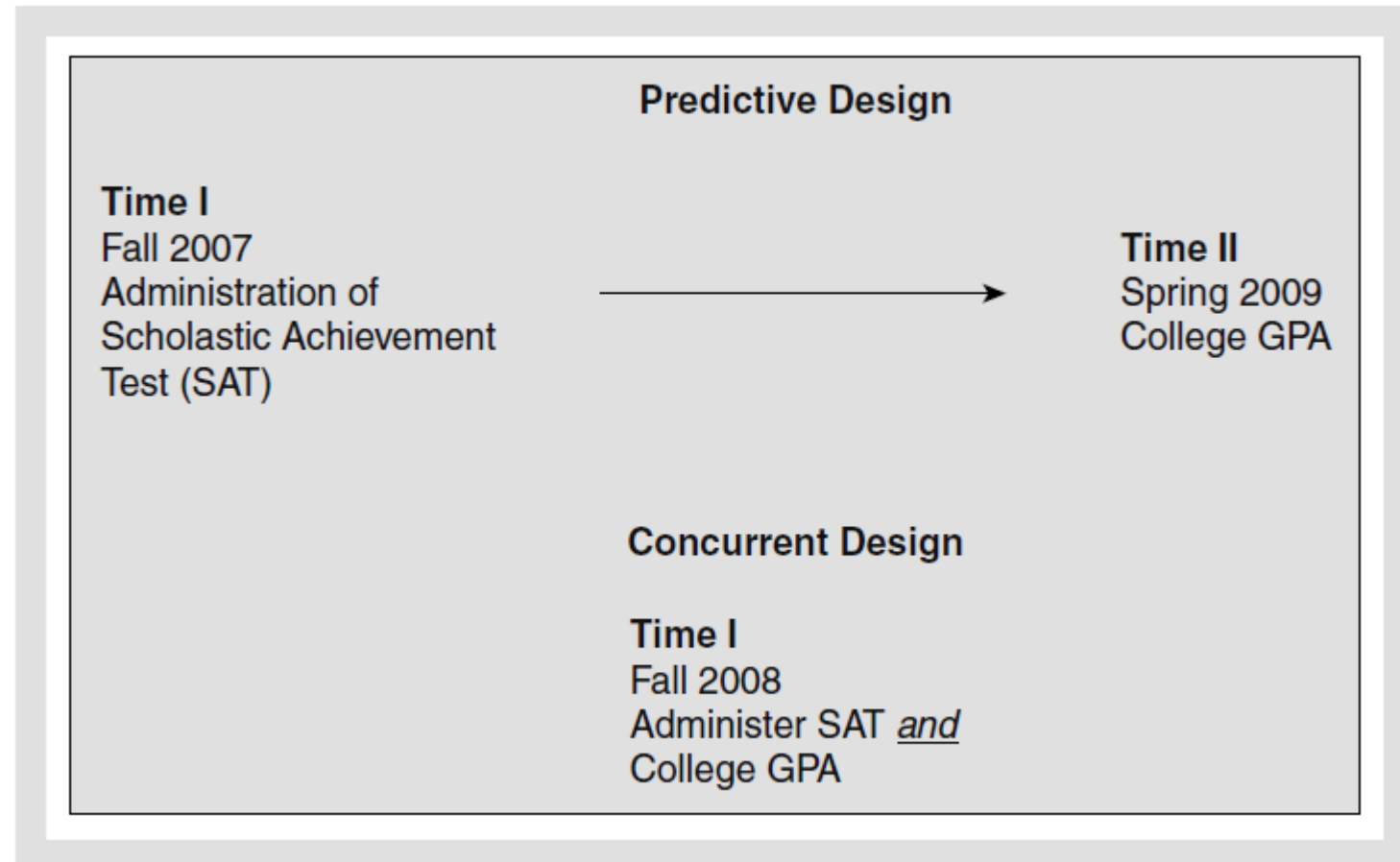
Evidência baseada em conteúdo

- É a relação entre o conteúdo do teste e o constructo ou domínio que o teste teoricamente está medindo
- O teste cobre o conteúdo que deveria cobrir?
- Geralmente é avaliado por especialistas



Evidências
baseadas em
outras variáveis

- Existem diversas modalidades desse tipo de evidência:
 - Evidência de critério
 - Teste de correlação geralmente tido como coeficiente de validade de critério
 - Regressão linear para projetar o quanto o teste pode prever o critério



Evidências baseadas em outras variáveis

- **Existem diversas modalidades desse tipo de evidência:**
 - **Convergente** : correlação entre o teste e outro teste que mede o mesmo constructo
 - **Divergente** : correlação com teste que mede outro constructo, geralmente um incompatível (e.g. ansiedade e *sensation seeking*)
 - **Diferenças entre grupos**: diferencia grupos que deveriam ter desempenhos diferentes? (e.g. deficientes intelectuais de não-patológicos)

Evidências
baseadas em
outras variáveis

TABLE 3 A Multitrait-Multimethod Matrix

		Self-Report			Parent Rating			Teacher Rating		
		D	H	S	D	H	S	D	H	S
Self-Report Scale	Depression	r_{xx}								
	Hyperactivity	HM	r_{xx}							
	Socialization	HM	HM	r_{xx}						
Parent Rating Form	Depression	r_{xy}	<i>HH</i>	<i>HH</i>	r_{xx}					
	Hyperactivity	<i>HH</i>	r_{xy}	<i>HH</i>	HM	r_{xx}				
	Socialization	<i>HH</i>	<i>HH</i>	r_{xy}	HM	HM	r_{xx}			
Teacher Rating Form	Depression	r_{xy}	<i>HH</i>	<i>HH</i>	r_{xy}	<i>HH</i>	<i>HH</i>	r_{xx}		
	Hyperactivity	<i>HH</i>	r_{xy}	<i>HH</i>	<i>HH</i>	r_{xy}	<i>HH</i>	HM	r_{xx}	
	Socialization	<i>HH</i>	<i>HH</i>	r_{xy}	<i>HH</i>	<i>HH</i>	r_{xy}	HM	HM	r_{xx}

Note: D, H, and S refers to depression, hyperactivity, and socialization, respectively. Monotrait-monomethod values are reliability coefficients (r_{xx}) and monotrait-heteromethod values are validity coefficients (r_{xy}). HM indicates heterotrait-monomethod correlations and *HH* indicates heterotrait-heteromethod correlations.

Evidências baseadas na estrutura interna, processo de resposta e consequências do teste.

- **Evidências baseadas na estrutura interna :** avaliação se a estrutura interna do teste é coerente com a estrutura do constructo. Utiliza-se análise fatorial, que permite verificar a existência de estrutura e constructos latentes nas variáveis do teste.
- **Evidências baseadas no processo de resposta :** avaliação se os comportamentos realizados no teste são referente ao constructo. (e.g. verificar se participantes utilizam raciocínio matemático ou cálculos respetivos – atalhos mentais).
- **Evidências baseadas nas consequências do teste:** o teste gera os benefícios que deveria gerar? Evidência mais polémica pois alguns autores consideram que não é uma evidência de validade.

Ilustração

“Are Psychological Tests as Accurate as Medical Tests? Many, if not most, people assume that tests used by medical professionals are more reliable and valid than those used by psychologists. That is, medical tests such as magnetic resonance imaging (MRI), X-rays, Pap smears, and electrocardiograms provide more reliable and accurate results than common psychological tests such as intelligence, neuropsychological, and personality tests. However, **Meyer et al. (2001)** reviewed research that examined the validity of a large number of medical and psychological tests and concluded that psychological tests often provide results that equal or exceed the validity of medical tests. For example, the Pap smear **test that is used to detect cervical abnormalities produces an effect size of 0.36** whereas the average ability of scores from the **Minnesota Multiphasic Personality Inventory—Second Edition to detect depressive or psychotic disorders has an effect size of 0.37** (larger effect sizes indicate superior validity). Even when you examine the validity of medical and psychological tests to detect the same disorder, psychological tests can provide superior results. For example, the effect size of **MRI results in detecting dementia is 0.57** whereas the **effect size for neuropsychological tests in detecting dementia is 0.68**. This paper reported effect sizes for over 140 medical and psychological tests. So the answer to the question is “Yes, psychological tests can provide information that is as valid as common medical tests.”