

Advanced Tests and Measurement

Validity



4/8/2003

PSY 721 Validity

1

What Is It?

The degree to which an inference from a test score is appropriate or meaningful.

- A test may be valid for one application but invalid for another.
- A test's validity is limited by its reliability.

4/8/2003

PSY 721 Validity

2

Types We Will Discuss

1. Face validity
2. Content validity
3. Criterion related validity
 - Concurrent
 - Predictive
4. Construct validity



4/8/2003

PSY 721 Validity

3

Type 1. Face Validity

The extent to which a test looks like it measures what it says it measures.

4/8/2003 4

Issues

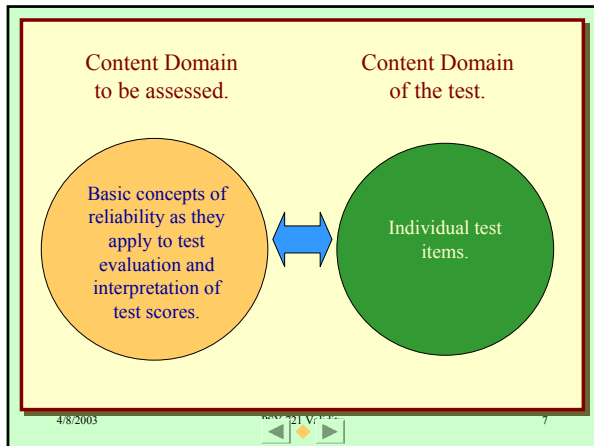
1. **Superficial.**
2. **Because it looks good doesn't mean it is good.**
3. **Because it looks weird doesn't mean it is weird.**

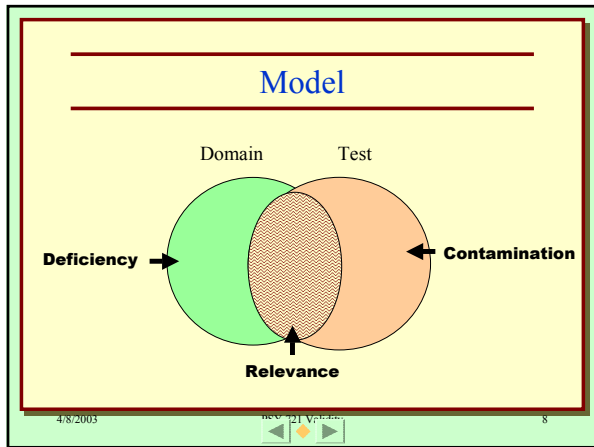
4/8/2003 5

Type 2. Content Validity

Showing that the behaviors sampled by the test are a representative sample of the attribute being measured.

4/8/2003 6





What Good Is It?

Does the test cover a representative sample of the skills, abilities, knowledge, and/or behaviors relevant to the construct being measured?

4/8/2003 4/8/2003 9

Concerns/Issues

1. Did the test items cover the Content Domain?
2. Did the test include items that were irrelevant to the content domain?
3. Were important aspects of the Content Domain missed by test items?
4. How to determine where 'good' is?

4/8/2003 10

Types of prediction

Clinical *Expert interpretation based on logical integration and interpretation of the test data.*

Actuarial *Statistical assessment using some empirically derived mathematical formula.*

4/8/2003 11

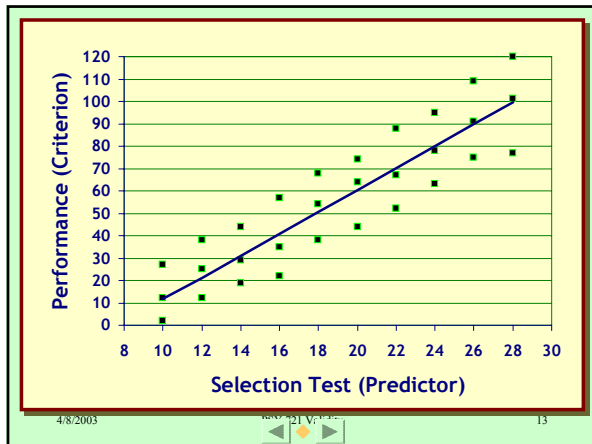
Type 3. Criterion Related Validity

Criterion *A standard or measure of the accuracy of a decision or behavioral prediction.*

Predictor *An assessment tool used to estimate a person's behavior.*

Validity Coefficient *The correlation between test scores (predictor) and the criterion.*

4/8/2003 12




- ### A. Predictive Validation
1. Test all applicants (predictor).
 2. Hire all applicants.
 3. Wait.....
 4. Collect criterion data.
 5. Evaluate the relationship between the predictor and the criterion.

- ### B. Concurrent Validation
1. Get sample of incumbents.
 2. Test sample (predictor).
 3. Get performance data on sample (criterion).
 4. Evaluate the relationship between the predictor and the criterion.

Question?

Which strategy is better and why?



4/8/2003 4:21 Version 16

Comparison

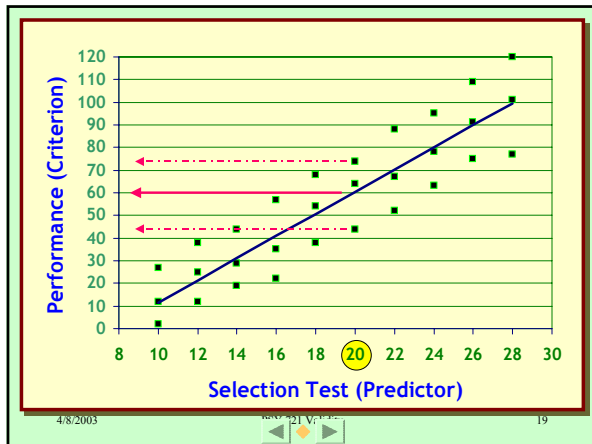
Predictive	Concurrent
Untaminated Sample	Contaminated Sample
Positive Test Attitude	Negative Test Attitude
Full Range of Scores	Restricted Range of Scores
Strong Statistics	Weak Statistics
Takes Time	Little Time
Expensive	Thrifty

4/8/2003 4:21 Version 17

Issues

1. Nature of the sample.
2. Changes over time.
3. Form of the relationship.
4. Is your criterion any good?
5. Standard error of estimate.

4/8/2003 4:21 Version 18



Standard Error of Estimate

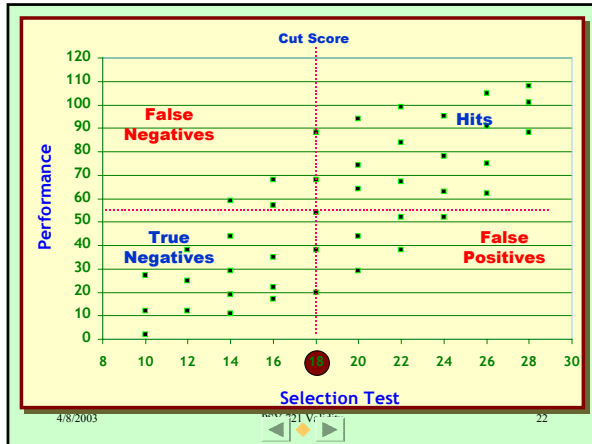
$$SE_{est} = SD_y \sqrt{1 - r_{xy}^2}$$

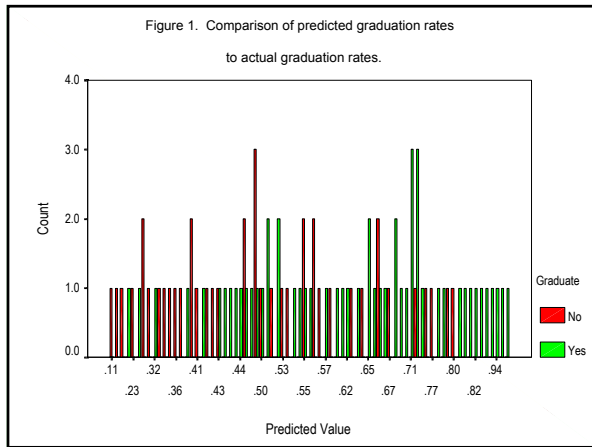
4/8/2003 20

Influence of Increasing r on SE_{est} ($SD = 10$)

r	r^2	$SD_y \sqrt{1 - r_{xy}^2}$
.90	.81	4.35
.80	.64	6.0
.70	.49	7.1
.60	.36	8.0

4/8/2003 21





Combining Tests

Test Battery *Group of tests used to predict a single criterion.*

Models

- Compensatory** *Strength in one area offsets weakness in another area.*
- Multiple Cutoff** *Minimal level required for one or more critical areas.*

The slide includes a footer with '4/8/2003', navigation arrows, and the number '24'.

Combining Tests, cont.

Multiple Regression

Optimal statistical combination of scores to predict a single criterion.

4/8/2003 4/8/2003 25

Decision Impact

- Selection
- Placement
- Classification

4/8/2003 4/8/2003 26

Type 4. Construct Validity

Demonstration that the test is measuring the hypothetical construct or trait that one claims it is measuring.

4/8/2003 4/8/2003 27

Evidence for Construct Validity

1. **Homogeneity.** Does the test score represent a single construct?
2. **Relationships.** Correlates with other tests in a way that is consistent with the predictions of the construct.
3. **Age.** Scores change as a function of maturation in a way that is consistent with the theory.
4. **Intervention.** Posttest scores change after intervention.
5. **Groups.** Scores from distinctly different groups vary.

4/8/2003 4:21 PM

	Decision Style			
	Rational	Intuitive	Dependent	Avoidant
PI Assertiveness				
AVA Assertiveness				
PI Sociability				
AVA Sociability				
PI Calmness			-.118	-.214
AVA Calmness			-.367	-.410
PI Conformity	.219		-.003	-.269
AVA Conformity			.462	.239

4/8/2003 4:21 PM

Convergent vs. Discriminant Validity

Convergent Validity Demonstrating that the test is related to other tests measuring the same thing.

Discriminant Validity Demonstrating that the test is NOT related to tests with which it should NOT be related..

4/8/2003 4:21 PM

Developmental Changes

- **Some constructs change as a function of age.**
 - Abilities.
 - Intelligence.
 - Cognitive skills.
- **Issues.**
 - Not all change as a function of age.
 - Cultural influences.

4/8/2003



31

Pretest – Posttest Changes

- **Issues.**
 - Experimental design.
 - State vs. Trait.

4/8/2003



32

Distinct Groups

Can the test differentiate between groups that are distinctly different on the construct?



4/8/2003



33

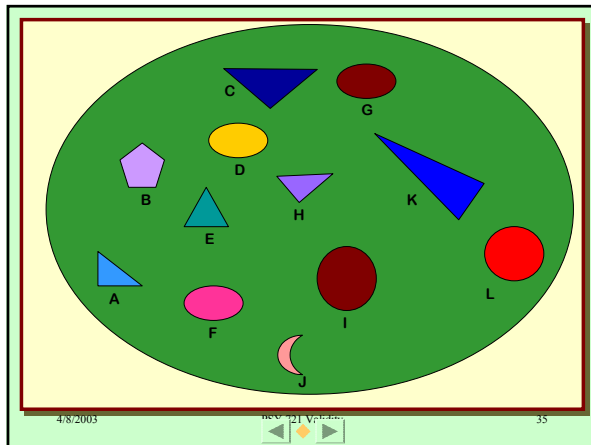
Factor Analysis

Statistical techniques for identifying interrelationships between items with the goal of identifying items that group or cluster together.

4/8/2003



34



4/8/2003



35

Test Bias

Factors inherent in a test that systematically prevent accurate, impartial measurement of one group.

4/8/2003



36

