

	<h2>VALIDITY</h2> <p>Cal State Northridge Psy 427 Andrew Ainsworth PhD</p>
--	--

	<h3>VALIDITY - DEFINITIONS</h3> <ul style="list-style-type: none">● The extent to which a test measures what it was designed to measure.● Agreement between a test score or measure and the quality it is believed to measure.● Proliferation of definitions led to a dilution of the meaning of the word into all kinds of “validities”
--	--

	<h3>VALIDITY - DEFINITIONS</h3> <ul style="list-style-type: none">● Internal validity - Cause and effect in experimentation; high levels of control; elimination of confounding variables● External validity - to what extent one may safely generalize the (internally valid) causal inference (a) from the sample studied to the defined target population and (b) to other populations (i.e. across time and space). Generalize to other people<ul style="list-style-type: none">▪ Population validity - can the sample results be generalized to the target population▪ Ecological validity - whether the results can be applied to real life situations. Generalize to other (real) situations
--	---

VALIDITY - DEFINITIONS

- ◎ Content validity - when trying to measure a domain are all sub-domains represented
 - When measuring depression are all 16 clinical criteria represented in the items
 - Very complimentary to domain sampling theory and reliability
 - However, often high levels of content validity will lead to lower internal consistency reliability

VALIDITY - DEFINITIONS

- ◎ Construct validity - overall are you measuring what you are intending to measure
 - Intentional validity - are you measuring what you are intending and not something else. Requires that constructs be specific enough to differentiate
 - Representation validity or translation validity - how well have the constructs been translated into measureable outcomes. Validity of the operational definitions
 - Face validity - Does a test “appear” to be measuring the content of interest. Do questions about depression have the words “sad” or “depressed” in them

VALIDITY - DEFINITIONS

- ◎ Construct Validity
 - Observation validity - how good are the measures themselves. Akin to reliability
 - Convergent validity - Convergent validity refers to the degree to which a measure is correlated with other measures that it is theoretically predicted to correlate with.
 - Discriminant validity - Discriminant validity describes the degree to which the operationalization does not correlate with other operationalizations that it theoretically should not correlate with.

VALIDITY - DEFINITIONS

- Criterion-Related Validity - the success of measures used for prediction or estimation. There are two types:
 - Concurrent validity - the degree to which a test correlates with an external criteria that is measured at the same time (e.g. does a depression inventory correlated with clinical diagnoses)
 - Predictive validity - the degree to which a test predicts (correlates) with an external criteria that is measured some time in the future (e.g. does a depression inventory score predict later clinical diagnosis)
- Social validity - refers to the social importance and acceptability of a measure

VALIDITY - DEFINITIONS

- There is a total mess of “validities” and their definitions, what to do?
- 1985 - Joint Committee of
 - AERA: American Education Research Association
 - APA: American Psychological Association
 - NCME: National Council on Measurement in Education
- developed *Standards for Educational and Psychological Testing (revised in 1999)*.

VALIDITY - DEFINITIONS

- According to the Joint Committee:
- Validity is the evidence for inferences made about a test score.
- Three types of evidence:
 - Content-related
 - Criterion-related
 - Construct-related
- Different from the notion of “different types of validity”

VALIDITY - DEFINITIONS

- ◉ Content-related evidence (*Content Validity*)
 - Based upon an analysis of the body of knowledge surveyed.
- ◉ Criterion-related evidence (*Criterion Validity*)
 - Based upon the relationship between scores on a particular test and performance or abilities on a second measure (or in real life).
- ◉ Construct-related evidence (*Construct Validity*)
 - Based upon an investigation of the psychological constructs or characteristics of the test.

VALIDITY THAT ISN'T

- ◉ Face Validity
 - The mere appearance that a test has validity.
 - Does the test look like it measures what it is supposed to measure?
 - Do the items seem to be reasonably related to the perceived purpose of the test.
- ◉ Does a depression inventory ask questions about being sad?
 - Not a "real" measure of validity, but one that is commonly seen in the literature.
 - Not considered legitimate form of validity by the Joint Committee.

CONTENT-RELATED EVIDENCE

- ◉ Does the test adequately sample the content or behavior domain that it is designed to measure?
- ◉ If items are not a good sample, results of testing will be misleading.
- ◉ Usually developed during test development.
 - Not generally empirically evaluated.
 - Judgment of subject matter experts.

CONTENT-RELATED EVIDENCE

- ◉ To develop a test with high content-related evidence of validity, you need:
 - good logic
 - intuitive skills
 - Perseverance
- ◉ Must consider:
 - wording
 - reading level

CONTENT-RELATED EVIDENCE

- ◉ Other content-related evidence terms
 - **Construct underrepresentation: failure to capture important components of a construct.**
 - Test is designed for chapters 1-10 but only chapters 1-8 show up on the test.
 - **Construct-irrelevant variance: occurs when scores are influenced by factors irrelevant to the construct.**
 - Test is well-intentioned, but problems secondary to the test negatively influence the results (e.g., reading level, vocabulary, unmeasured secondary domains)

CRITERION-RELATED EVIDENCE

- ◉ Tells us how well a test corresponds with a particular criterion
 - criterion: behavioral or measurable outcome
 - SAT predicting GPA (GPA is criterion)
 - BDI scores predicting suicidality (suicide is criterion).
- ◉ Used to “predict the future” or “predict the present.”

CRITERION-RELATED EVIDENCE

- Predictive Validity Evidence
 - forecasting the future
 - how well does a test predict future outcomes
 - SAT predicting 1st yr GPA
 - most tests don't have great predictive validity
- decrease due to time & method variance

CRITERION-RELATED EVIDENCE

- Concurrent Validity Evidence
 - forecasting the present
 - how well does a test predict current similar outcomes
 - job samples, alternative tests used to demonstrate concurrent validity evidence
- generally higher than predictive validity estimates

QUANTIFYING CRITERION-RELATED EVIDENCE

- Validity Coefficient
 - correlation between the test and the criterion
 - usually between .30 and .60 in real life.
 - In general, as long as they are statistically significant, evidence is considered valid.
- However,
 - recall that r^2 indicates explained variance.
 - SO, in reality, we are only looking at explained criterion variance in the range of 9 to 36%.
- Sound Problematic??

EVALUATING CRITERION-RELATED EVIDENCE

- Look for changes in the cause of relationships. (*third variable effect*)
 - E.g. *Situational factors during validation that are replicated in later uses of the scale*
- Examine what the criterion really means.
 - Optimally the criterion should be something the test is trying to measure
 - If the criterion is not valid and reliable, you have no evidence of criterion-related validity!
- Review the subject population in the validity study.
 - If the normative sample is not representative, you have little evidence of criterion-related validity.

EVALUATING CRITERION-RELATED EVIDENCE

- Ensure the sample size in the validity study was adequate.
- Never confuse the criterion with the predictor.
 - GREs are used to predict success in grad school
 - Some grad programs may admit low GRE students but then require a certain GRE before they can graduate.
 - So, low GRE scores succeed, this demonstrates poor predictive validity!
 - But the process was dumb to begin with...
- Watch for restricted ranges.

EVALUATING CRITERION-RELATED EVIDENCE

- Review evidence for validity generalization.
 - Tests only given in laboratory settings, then expected to demonstrate validity in classrooms?
 - Ecological validity?
- Consider differential prediction.
 - Just because a test has good predictive validity for the normative sample may not ensure good predictive validity for people outside the normative sample.
 - External validity?

CONSTRUCT-RELATED EVIDENCE

- ◉ Construct: something constructed by mental synthesis
 - What is Intelligence? Love? Depression?
- ◉ Construct Validity Evidence
 - assembling evidence about what a test means (and what it doesn't)
 - sequential process; generally takes several studies

CONSTRUCT-RELATED EVIDENCE

- ◉ Convergent Evidence
 - obtained when a measure correlates well with other tests believed to measure the same construct.
 - Self-report, collateral-report measures
- ◉ Discriminant Evidence
 - obtained when a measure correlates less strong with other tests believed to measure something slightly different
 - This does not mean any old test that you know won't correlate; should be something that could be related but you want to show is separate
 - Example: IQ and Achievement Tests

VALIDITY & ESTIMATES

- ◉ Standard Error of Estimate:

$$s_{est.} = s_{y-\hat{y}} = s_y \sqrt{(1-r^2) \left(\frac{N-1}{N-2} \right)}$$

- $s_{est.}$ standard error of estimate
- s_y standard deviation of the test
- r validity of the test

- ◉ Essentially, this is regression all over again.

VALIDITY AND RELIABILITY

◎ Maximum Validity depends on Reliability

$$r_{12\max} = \sqrt{r_1 r_2}$$

- $r_{12\max}$ is the maximum validity
- r_1 is the reliability of test 1
- r_2 is the reliability of test 2

VALIDITY AND RELIABILITY

Reliability of Test	Reliability of Criterion	Maximum Validity (Correlation)
1	1	1.00
0.8	1	0.89
0.6	1	0.77
0.4	1	0.63
0.2	1	0.45
0	1	0.00
1	0.5	0.71
0.8	0.5	0.63
0.6	0.5	0.55
0.4	0.5	0.45
0.2	0.5	0.32
0	0.5	0.00
1	0.2	0.45
0.8	0.2	0.40
0.6	0.2	0.35
0.4	0.2	0.28
0.2	0.2	0.20
0	0.2	0.00
1	0	0.00
0.8	0	0.00
0.6	0	0.00
0.4	0	0.00
0.2	0	0.00
0	0	0.00
