

Lecture 11: Validity

Dewayne E Perry

ENS 623

perry@ece.utexas.edu

All Too True



Validity

→ In software engineering, we worry about various issues:

↳ *E-Type systems:*

- Usefulness - is it doing what is needed
- Embodying important required characteristics - is it doing it in an acceptable or appropriate way

↳ *S-Type programs:*

- correctness of functionality - is it doing what it is supposed to do
- Embodying important required characteristics - are the structures consistent with the way it should perform

Validity

- In empirical work, worried about similar kinds of things
 - ↳ Are we testing what we mean to test
 - ↳ Are the results due solely to our manipulations
 - ↳ Are our conclusions justified
 - ↳ What are the results applicable to
- The questions correspond to different *validity* concerns
- Concerned with the logic of demonstrating causal connections
- Concerned about the logic of evidence

Validity

→ 4 primary types of validity

- ↳ Construct Validity
- ↳ Internal Validity
- ↳ Statistical Conclusion
- ↳ External Validity

→ Comments

- ↳ This organization differs somewhat from R&R
- ↳ Each sequentially dependent on preceding

Construct Validity

- **Are we measuring what we intend to measure**
 - ↳ Akin to the requirements problem: are we building the right system
 - ↳ If we don't get this right, the rest doesn't matter
- **Constructs: abstract concepts**
 - ↳ Theoretical constructions
 - ↳ Must be operationalized in the experiment
- **Necessary condition for successful experiment**
- **Divide construct validity into three parts:**
 - ↳ Intentional Validity
 - ↳ Representation Validity
 - ↳ Observation Validity

Construct Validity

→ Intentional Validity

- ↪ Do the constructs we chose adequately represent what we intend to study
- ↪ Akin to the requirements problem where our intent is *fair scheduling* but our requirement is FIFO
- ↪ Are our constructs specific enough
- ↪ Do they focus in the right direction
- ↪ Eg, is it *intelligence* or *cunningness*

Construct Validity

→ Representation Validity

- ↳ How well do the constructs or abstractions translate into observable measures
- ↳ Two primary questions:
 - Do the sub-constructs properly define the constructs
 - Do the observations properly interpret, measure or test the constructs
- ↳ 2 ways to argue for representation validity
 - **Face validity**
 - ✓ Claim: on the face of it, seems like a good translation
 - ✓ Very weak argument
 - ✓ Strengthened by consensus of experts
 - **Content validity**
 - ✓ Check the operationalization against the domain for the construct
 - ✓ The extent to which the tests measure the content of the domain being tested - ie, cover the domain
 - ✓ The more it covers the relevant areas, the more content valid
 - **Both are nonquantitative judgments**

Construct Validity

→ Observation Validity

↳ How good are the measures themselves

↳ Different aspects illuminated by

- Predictive validity
- Criterion validity
- Concurrent validity
- Convergent validity
- Discriminant validity

Construct Validity

↪ Predictive Validity

- Observed measure predicts what it should predict and nothing else
- Eg, college aptitude tests are assessed for their ability to predict success in college

↪ Criterion Validity

- Degree to which the results of a measure agree with those of an independent standard
- Eg, for college aptitude, GPA or successful first year

↪ Concurrent Validity

- The observed measure correlates highly with an established set of measures
- Eg, shorter forms of tests against longer forms

Construct Validity

↳ Convergent Validity

- Observed measure correlates highly with other observable measures for the same construct
- Utility is not that it duplicates a measure but is a new way of distinguishing a particular trait while correlating with similar measures

↳ Discriminant Validity

- The observable measure distinguishes between two groups that differ on the trait in question
- Lack of divergence argues for poor discriminant validity

↳ R&R discuss various interesting correlations on convergent and discriminant validity among various psychological tests

- In terms of validity, reliability and stability

Internal Validity

- Are the values of the dependent variables solely the result of the manipulations of the independent variables
- Have we ruled out rival hypotheses
- Have we eliminated confounding variables
 - ↳ Participant variables
 - ↳ Experimenter variables
 - ↳ Stimulus, procedural and situational variables
 - ↳ Instrumentation
 - ↳ Nuisance variables

Statistical Conclusion Validity

- Are the presumed causal variable X and its effect Y statistically related
 - ↳ Ie, do they covary
 - ↳ If unrelated then the one cannot be the cause of the other
- 3 questions (sequentially dependent)
 - ↳ Is the study sufficiently sensitive
 - ↳ What is the evidence that they covary
 - ↳ How strongly do they covary

External Validity

→ Two positions

- ↳ The generalizability of the causal relationship beyond that studied/observed
 - Eg, do studies of very large reliable real-time systems generalize to small .COM companies
- ↳ The extent to which the results support the claims of generalizability
 - Eg, do the studies of 5ESS support the claim that they are representative of real-time ultra reliable systems