

Statistical Validity Pitfalls

Dr. Julie A. Adams
Electrical Engineering and Computer Science Department
Vanderbilt University
Nashville, TN USA

Garbage In, Garbage Out!



“The key to success is knowing what people want. Too bad it isn’t knowing what people don’t want.”

Evaluation Validity

- Evaluation **validity** is concerned with the correspondence between how representative the evaluation results are of how the evaluated activities will be performed in the real world.
 - “What is true about behavior for one time and place may not be universally true” (Maxwell and Delaney, 2004)

Types of validity

- Statistical conclusion validity
- Internal validity
- Construct validity
- External validity

Statistical Conclusion Validity

- “was the original statistical inference correct?”
 - Did the investigators arrive at the correct conclusion regarding whether or not a relationship between the variables exists or the extent of the relationship?
 - Not concerned with the causal relationship between variables, but whether or not there is any relationship, either causal or not.

Statistical Conclusion Validity

- Type I Error
 - Conclude that a relationship exists between two variables, when in fact there is no relationship.
- Type II Error
 - Conclude that there is no relationship when one exists.
- The power of the analysis focuses on the sensitivity or ability to detect a relationship.

Statistical Conclusion Validity

- Threats to statistical validity
 - **Liberal biases**: being overly optimistic regarding the existence of a relationship or exaggerating its strength.
 - **Conservative biases**: being overly pessimistic regarding the absence of a relationship or underestimating its strength.
 - **Low power**: the probability that the evaluation will result in a Type II error.

Statistical Conclusion Validity

- Threats to statistical validity

<i>Threats leading to overly conservative bias</i>	<i>Remedies</i>
Small sample size	Increase sample size
Increased error from irrelevant, unreliable, or invalid measures	Improve measurements
High variability due to participant diversity	Control individual differences: control for covariates; using a design that blocks, matches, or uses repeated measures.
Violation of statistical assumptions	Transform data or use different analysis methods.
<i>Threats leading to overly liberal bias</i>	
Repeated statistical test	Use adjusted test procedures
Violation of statistical assumptions	Transform data or use different analysis methods
Biased estimates of effects	Use corrected values to estimate effects in population

Internal Validity

- “Is there a causal relationship between variable X and variable Y, regardless of what X and Y are theoretically supposed to represent?”
 - If a variable is a true independent variable and the statistical conclusion is valid, then internal validity is largely assured.
 - The concern of internal validity is causal in that we are asking what is responsible for the change in the dependent variable.

Internal Validity Threats

<i>Threats</i>	<i>Definition</i>
Selection Bias	Participant characteristics confounded with treatment conditions because of use of intact or self-selected participants, or more generally, whenever predictor variables represent measured characteristics as opposed to independently manipulated treatments.
Attrition	Differential drop out across conditions at one or more time points that may be responsible for differences.
Testing	Altered performance as a result of a prior measure or assessment instead of the assigned conditions.
Regression	The changes over time expected in the performance of participants, selected because of extreme scores on a variable, that occur for statistical reasons but may incorrectly be attributed to the intervening condition.
Maturation	Observed changes as a result of ongoing, naturally occurring processes rather than condition effects.
History	Events, in addition to an assigned condition, to which participants are exposed between repeated measurements that could influence performance.

(Maxwell and Delaney, 2004)

Metrics for HRI Workshop
March 12, 2008

Construct Validity

- “Given there is a valid causal relationship, is the interpretation of the constructs involved in that relationship correct?”
- The **problem**: there is a possibility “that the operations which are meant to represent a particular cause or effect construct can be construed in terms of more than one construct, each of which is stated at the same level of reduction.”

Construct Validity Threats

- **Experimenter bias:** The experimenter transfers expectations to the participants in a manner that affects performance for dependent variables.
- **Condition diffusion:** The possibility of communication between participants from different condition groups during the evaluation.
- **Resentful demoralization:** A group that is receiving nothing finds out that a condition (treatment) that others are receiving is effective.
- **Inadequate preoperational explication:** The construct under consideration does not assess what you want and incorporates similar constructs that should be distinguished from the desired construct.
- **Mono-operation bias:** The use of only a single dependent variable to assess a construct may result in under representing the construct and containing irrelevancies.

External Validity

- “Can the finding be generalized across populations, settings, or time?”
 - A primary concern is the heterogeneity and representativeness of the evaluation sample population.

Garbage In, Garbage Out!



“The key to success is knowing what people want. Too bad it isn’t knowing what people don’t want.”

References

- V. J. Gawron, (2000). *Human Performance Measures Handbook*, Lawrence Erlbaum Associates
- R. J. Grissom & J. J. Kim (2005) *Effect Sizes for Research*, Lawrence Erlbaum Associates
- S. E. Maxwell & H. D. Delaney (2004) *Designing Experiments and Analyzing Data: A model comparison perspective*, Second Edition, Lawrence Erlbaum Associates
- J. L. Myers & A. D. Well (2003) *Research Design and Statistical Analysis*, Second Edition, Lawrence Erlbaum Associates
- K. R. Murphy & B. Myers (2004) *Statistical Power Analysis*, Second Edition, Lawrence Erlbaum Associates
- J. Nielsen, (1993). *Usability Engineering*. Morgan Kaufman
- C. D. Wickens, J. D. Lee, Y. Liu, S. E. Gordon Becker, (2004) *An Introduction to Human Factors Engineering*, Second Edition, Pearson/Prentice Hall.