



## More on Validity

In the “Validity and Reliability: Quantitative Research Study Buddies” Research Snack, we made the point that validity relates to the accuracy of a research study and the appropriate interpretation of the results. The essential question that underpins validity is: are you measuring what you purport to measure, and appropriately interpreting the results?

The term validity is used in a couple of distinct ways in engineering education research:

1. To describe the quality of an instrument or items in the instrument. An instrument could be a survey, rubric, test or other measurement system. There are several types of validity that relate to instrument design, development and testing (see table below). These types of validity are most frequently associated with instruments used for observational (correlational) studies.
2. To describe the quality of a study design (typically an experiment or quasi-experiment) and the way it is run. Most often explained as “threats” to validity, these are internal and external factors in the research that may impact the interpretation, or generalizability, of the results.

1. Types of validity used in instrument development	
Face validity	Do the participants interpret the items and instructions on the instrument as intended? <sup>1</sup> Is your instrument clear and transparent to the population of interest?
Construct validity	Do the items in your instrument reflect key dimensions of the construct as characterized by the research on this topic?  This type of validity is substantially related to the conceptual framework used in your research and builds on a literature review. An instrument that has construct validity is tuned to the conceptual or theoretical framework you are using, and it is clear in your reporting how it relates to the framework and covers the major aspects of the constructs you are purporting to measure.
Content validity	Do the items measure the intended content/construct? Does the instrument cover the key aspects of the content/construct? If an instrument does not cover the major aspects of the construct effectively, the result is “construct underrepresentation” or “construct deficiency”.
Criterion validity	Sometimes this type is split into two: predictive validity and concurrent validity. Both are related to the concept of triangulation: do you have other evidence that suggests your instrument is accurate? Or is there some objective criterion you can use for comparison with your results to strengthen confidence in your instrument?  <u>Predictive validity</u> : Do the results from your instrument accurately predict some other independent measure? For example, if you design a tool to measure students’ design ability, do the results predict their grades in a subsequent design course?  <u>Concurrent validity</u> : Do the results from your instrument concur with another, independent measurement of the same population (i.e., convergent evidence)? Or, possibly, do your results diverge from results from some other, independent measurement in an expected way (i.e., discriminating evidence)?

<sup>1</sup> An “item” may be a question on a survey or test. But not all items are questions. For example; an item may be a criteria on a rubric, or a statements on a survey that you are asking participants to rank.

2. Types of validity related to the context and participants, typically in an experiment  
(Creswell has an excellent section on this (p. 169-172), which I highly recommend reading)

Internal validity	Relates to the selection of the participant group and their experiences, feelings, and so on, and how these may affect (i.e., threaten) an “objective” measure.
External validity	Relates to external factors that limit the generalizability of the results or interpretation of the results: <ul style="list-style-type: none"> <li>- Ecological validity: is the context you used for observations or other aspects of the study authentic (e.g., observations in a real class or a mock class)? Can you generalize your results to other contexts, or are the results highly context specific?</li> <li>- Historical validity: are the results generalizable to future situations? Or will changes that occur over time mean your findings become less meaningful/accurate?</li> <li>- Population validity: are your results generalizable to other populations? Or specific to the specific group who participated (e.g., at one university, one discipline, etc.)</li> </ul>

There are other types of validity discussed in the literature, and sometimes they are organized somewhat differently, but this table covers the ones mentioned most frequently.

Common question: Do I need to address every type of validity in my study design and write-up? Answer: No, not necessarily. Most studies and write-ups do not address every type of validity explicitly. However:

1. Even if you don’t include a write-up on every aspect of validity, think through these types, and the overall question of validity in your study. Reflect, document (make notes) and talk with collaborators about the validity of your study. This will help you design and conduct better research. It will also help you answer questions about your research when you present it.
2. Do include the most critical aspects of validity in your write-up. This is an important aspect of research and something that reviewers will look for in your work.

The table shows types of validity that are relevant for instrument development and types of validity relevant for experiments. You might ask: these all seem relevant for both observational and experimental work – Is this categorization really so distinct? Answer: No. This is the categorization used by Creswell, but in fact these types of validity are important considerations in all types of research. That said, some types will be more important to a particular study than others; so identify which ones best align with your work.

It is worth noting that not every source uses “types of validity”. The AERA Standards for Testing, which specifically pertains to the development and use of educational and psychological tests, discourages the use of “types of validity”. Rather they make the point that validity resides in the interpretation of the results. That said, their description of the development of tests references many of the concepts identified in the table.

No matter how it is framed, validity is a critical aspect of engineering education research.

**Author: Susan McCahan, University of Toronto**

Acknowledgement: Thank you also to Brian Frank (Queen’s U.), Sean Maw (U. Saskatchewan), and Cindy Rottmann (U. Toronto) for contributing to this snack.

**References:**

“Standards for Educational and Psychological Testing”, AERA, APA and NCME, 2014.  
 J.W. Creswell and J.D. Creswell, “Research Design: Qualitative, Quantitative and Mixed Methods Approaches”, 5<sup>th</sup> ed., Sage, 2018.  
 A. Field, “Discovering Statistics using IBM SPSS Statistics”, 4<sup>th</sup> ed., Sage, 2013.  
 R. Heale and A. Twycross, “Validity and reliability in quantitative studies,” Evid Based Nurs, Vol. 18(3), 2015.  
 B.M. Moskal, J.A. Leydens, M.J Pavelich, “Validity, Reliability and the Assessment of Engineering Education”, JEE, 2002.

Except where otherwise noted, this guide document is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Distributed: May 31, 2022

