Research Rundowns >Quantitative Methods > Instrumentation, Validity, Reliability

Part I: The Instrument

Instrument is the generic term that researchers use for a measurement device (survey, test, questionnaire, etc.). To help distinguish between instrument and instrumentation, consider that the *instrument is the device* and *instrumentation is the course of action* (the process of developing, testing, and using the device).

Instruments fall into two broad categories, researcher-completed and subject-completed, distinguished by those instruments that researchers administer versus those that are completed by participants. Researchers chose which type of instrument, or instruments, to use based on the research question. Examples are listed below:

Researcher-completed Instruments	**Subject-completed Instruments**
Rating scales	Questionnaires
Interview schedules/guides	Self-checklists
Tally sheets	Attitude scales
Flowcharts	Personality inventories
Performance checklists	Achievement/aptitude tests
Time-and-motion logs	Projective devices
Observation forms	Sociometric devices

Usability

Usability refers to the ease with which an instrument can be administered, interpreted by the participant, and scored/interpreted by the researcher. Example usability problems include:

1. Students are asked to rate a lesson immediately after class, but there are only a few minutes before the next class begins (problem with administration).
2. Students are asked to keep self-checklists of their after school activities, but the directions are complicated and the item descriptions confusing (problem with interpretation).
3. Teachers are asked about their attitudes regarding school policy, but some questions are worded poorly which results in low completion rates (problem with scoring/interpretation).

Validity and reliability concerns (discussed below) will help alleviate usability issues. For now, we can identify five usability considerations:

1. How long will it take to administer?
2. Are the directions clear?
3. How easy is it to score?
4. Do equivalent forms exist?
5. Have any problems been reported by others who used it?

It is best to use an existing instrument, one that has been developed and tested numerous times, such as can be found in the Mental Measurements Yearbook. We will turn to why next.

Part II: Validity

Validity is the extent to which an instrument measures what it is supposed to measure and performs as it is designed to perform. It is rare, if nearly impossible, that an instrument be 100% valid, so validity is generally measured in degrees. As a process, validation involves collecting and analyzing data to assess the accuracy of an instrument. There are numerous statistical tests and measures to assess the validity of quantitative instruments, which generally involves pilot testing. The remainder of this discussion focuses on external validity and content validity.

External validity is the extent to which the results of a study can be *generalized* from a sample to a population. Establishing eternal validity for an instrument, then, follows directly from sampling. Recall that a sample should be an accurate representation of a population, because the total population may not be available. An instrument that is externally valid helps obtain population generalizability, or the degree to which a sample represents the population.

Content validity refers to the appropriateness of the content of an instrument. In other words, do the measures (questions, observation logs, etc.) accurately assess what you want to know? This is particularly important with achievement tests. Consider that a test developer wants to maximize the validity of a unit test for 7th grade mathematics. This would involve taking representative questions from each of the sections of the unit and evaluating them against the desired outcomes.

Part III: Reliability

Reliability can be thought of as consistency. Does the instrument consistently measure what it is intended to measure? It is not possible to calculate reliability; however, there are four general estimators that you may encounter in reading research:

1. *Inter-Rater/Observer Reliability*: The degree to which different raters/observers give consistent answers or estimates.
2. *Test-Retest Reliability*: The consistency of a measure evaluated over time.
3. *Parallel-Forms Reliability:* The reliability of two tests constructed the same way, from the same content.
4. *Internal Consistency Reliability:* The consistency of results across items, often measured with Cronbach's Alpha.

Relating Reliability and Validity

Reliability is directly related to the validity of the measure. There are several important principles. First, a test can be considered reliable, but not valid. Consider the SAT, used as a predictor of success in college. It is a reliable test (high scores relate to high GPA), though only a moderately valid indicator of success (due to the lack of structured environment – class attendance, parent-regulated study, and sleeping habits – each holistically related to success).

Second, validity is more important than reliability. Using the above example, college admissions may consider the SAT a reliable test, but not necessarily a valid measure of other quantities colleges seek, such as leadership capability, altruism, and civic involvement. The combination of these aspects, alongside the SAT, is a more valid measure of the applicant's potential for graduation, later social involvement, and generosity (alumni giving) toward the alma mater.

Finally, the most useful instrument is both valid and reliable. Proponents of the SAT argue that it is both. It is a moderately reliable predictor of future success and a moderately valid measure of a student's knowledge in Mathematics, Critical Reading, and Writing.

Part IV: Validity and Reliability in Qualitative Research

Thus far, we have discussed Instrumentation as related to mostly quantitative measurement. Establishing validity and reliability in qualitative research can be less precise, though participant/member checks, peer evaluation (another researcher checks the researcher's inferences based on the instrument (Denzin & Lincoln, 2005), and multiple methods (keyword: *triangulation*), are convincingly used. Some qualitative researchers reject the concept of validity due to the constructivist viewpoint that reality is unique to the individual, and cannot be generalized. These researchers argue for a different standard for judging research quality. For a more complete discussion of trustworthiness, see Lincoln and Guba's (1985) chapter.