



## Reliability and Validity in Assessment

Measures that are developed in student learning outcome assessment need to be both reliable as well as valid. There are several aspects of reliability that programs may measure in their assessment processes. First, programs will probably want to have measures that exhibit stability over time. Measuring stability may require collecting information on student learning outcomes in a longitudinal nature. Similar measures can be analyzed over a period of time to assess stability.

In addition to stability, programs may also establish equivalence between measurements. Inter-rater reliability is one of the easiest and most common approaches to measuring equivalence. Inter-rater reliability may be measured by using multiple evaluators to evaluate direct evidence of student work. Some programs accomplish this by requiring first and second readers to evaluate student work utilizing a uniform assessment rubric.

Programs may also wish to evaluate the internal consistency of their assessment measures by correlating the scores on individual indicators of the required competency to be measured. For instance, it is common for programs to utilize several assignments to measure students' ability to lead and manage in public governance. For example, programs may use a course embedded assignment from a Research Methods course and a comprehensive examination question from a Policy course to measure the ability of students to analyze, synthesize, think critically, solve problems, and make decisions. Measuring the internal consistency of these assignments could provide the program with very useful information regarding which assignments truly measure this universal competency.

Validity is also an important consideration for programs in developing assessment measures. Criterion validation is one approach that can be done to compare student performance on course embedded assignments and assessment measures that are collected at the end of the program. For example, programs could use a concurrent approach by comparing student performance on culminating experiences (e.g. comprehensive exams, case studies, etc.) to the student's performance on course embedded assignments. By contrast, a predictive approach could allow programs to use student performance on course embedded assignments to predict performance on the culminating experience in the program.

These are all simple measures to assess the reliability and validity of the assessments utilized by programs. Fortunately, most programs have already collected the data necessary to measure both reliability and validity as part of the normal assessment process.