

Video 4: About 11 minutes long [Already on Test Web page]

The purpose of this video is to introduce to faculty and academic leaders to potential measures for assessing the achievement of program learning outcomes. These steps would also apply in the assessment of course learning outcomes and essential learning outcomes.

Measures

In this video, we will discuss ideas for how you could measure the achievement of program learning outcomes.

Reliability and Validity in Assessment

First, before we cover some potential project ideas, let's review the concepts of reliability and validity.

What is the difference?

Reliability looks at whether several colleagues would agree to whether a student work would actually be a certain score or a different score.

One way to ensure reliability is to work with another colleague or a group of colleagues on what we call a norming session. A norming session is an opportunity to look at some samples of student work and decide what characteristics qualify a student for having achieved a program learning outcome.

Validity will ask, "Is this project measuring what it is intended to measure?"

Validity will look at alignment between the tool used to measure an outcome (for example, a rubric or criteria) and the skills that the students are supposed to have acquired after completing assignments.

Content validity is a reference back to an earlier video on curriculum maps, where we would ask ourselves, "Does this assignment actually align with a specific outcome?"

And Criterion validity will ask ourselves, "Does this assignment actually measure achievement of an outcome?" Even when we use a syllabus to create a curriculum map, sometimes the assignment and outcome mapped aren't a direct match.

Notes about Reliability and Validity

Here is an image summarizing some of the definitions that we covered in the last slide. Let's say the dots represent people coming to the same conclusion. Notice that in this image, we have four targets. In the target on the upper left-hand corner, we have a whole bunch of dots that are clustered together, meaning that the people in this case agree on a certain conclusion. However, the people are nowhere near the bullseye, or the measurement of an outcome. In this example, the assessment is reliable (because people agree on an outcome), but it is not valid (because the assessment didn't actually measure the achievement of an outcome).

In the image of a target in the upper right-hand side of the screen, we see that people are all over the target board. In this case, we can interpret the image to mean that the people's conclusions are nowhere near in agreement, meaning that the assessment is unreliable. The assessment is also not valid because the outcome has not been actually measured.

In the lower left, we see a target where there is a similar situation. The people's conclusions are scattered in one general area, however, they are not similar nor clustered together. Furthermore, they are not near the bullseye, so they are not measuring the outcome they are intending to measure.

Finally, in the lower right, we see a target with a clustered consensus centered on the bullseye. This assessment is both valid and reliable. It is measuring the outcome that was supposed to be measured. The people's conclusions are all clustered around one another.

Now that we have reviewed reliability and validity, we can make some notes about assessments when working on these projects. First, no assessment will be 100% valid. There are too many factors that could interject with the strict interpretation of a program learning outcome.

We can actually combat the risk of an invalid assessment through measuring the same outcome in more than one method over a period of time.

Example: what makes a cookie high quality?

We will review this slide briefly as a review of reliability and validity because it is just a lot of fun. Consider, you are teaching a baking class. Which chocolate chip cookie is scored the highest? If we were in a room full of people right now, different people would select different cookies based off of their own personal tastes and opinions. Same goes for assignments. We need to have conversations with colleagues to make our projects reliable; we need to discuss what we think the highest quality cookie in this case would be. We would also need to specify the criteria that we are looking for to ensure validity. Do we want the cookie to be a certain size? Do we want the cookie to have a specific texture or certain sized chocolate chunks?

Types of Academic Assessments

There are two types of academic assessments. The first type is direct assessment, which we highly encourage to prioritize. Direct assessment looks at whether student work displays achievement of student learning. An example of a direct assessment would be taking a sample of student work and reviewing it against a rubric. Many programs do this type of project when evaluating their capstones and portfolios.

Another example of direct assessment is blueprinting. Blueprinting is the process of assigning an outcome to an exam question and then seeing how students are performing in successfully completing that exam question.

Pre-tests and post-tests measure student learning prior to learning new knowledge and after a lesson or being exposed to learning material. This method could be really helpful in learning how much information students are learning from a learning intervention.

There are also indirect assessments. Indirect assessment focus on people's opinions on their own learning. Indirect assessment has some pros and cons. It can be invaluable to see how students are experiencing the program and whether there are some missing pieces or gaps in the program.

One thing to keep in mind is that indirect assessment can be tricky because sometimes people will have an inaccurate idea of how much they learned or mastered.

Some examples of indirect assessment include conducting focus groups and interviews with students, alumni, and/or instructors/adjuncts. Another example could be surveying participants in co-op's, clinicals, and internships. This survey could include both students and supervisors. Finally, student surveys and exit interviews could be used to review the strengths and challenges of a course or program.

Assessment Does NOT Equal Grades

One misconception is the idea that student grades could be used for assessment. However, we can subconsciously subjectively grade students based off of our memory of their last performance. We also make judgment calls about student assignments based off of the student's identity. This subjectivity is why assessment is so important. It forces us to look at student work, without knowing the student's identity, and forces us to decide whether the work (not associated with a student) is considered high quality or that it achieved outcomes.

Example #1 of an Academic Assessment Project—Rubrics

You can take some sample student assignments (such as student artifacts) and measure how well students are meeting a program learning outcome (using a rubric).

Ensure that your rubric is directly related to the program learning outcome. For instance, if your outcome is about ethics, you may want to review students using a rubric focused more on ethics, than on writing.

AAC&U rubrics are a great example of an assessment rubric. They have been tested for reliability and validity nationwide. We use these rubrics for Essential Learning Outcomes assessment at the University of Hartford; however, these rubrics can be used for assessment of course learning outcomes and program learning outcomes.

Example # 2 of an Academic Assessment Project: Blueprinting

If your program has a lot of exams, you could try to assign a program learning outcome to each question of a specific exam.

This kind of project can be interesting because it can give you further insight about what actually makes up your exam.

It can also provide insight on how well students are answering questions, related to one program learning outcome over another.

Bias and Positionality

While we measure our achievement of student learning outcomes, we also need to be mindful of the potential bias that may be involved.

There is a risk that the instructor has confirmation bias, where they may rate a student based off of existing beliefs about that student, while ignoring contrary information, as illustrated with this image.

Leniency bias may take place if the instructor rates a student favorably when they have room for improvement.

Instructors may also experience the recency effect, where someone is evaluated about how well they did in a recent period.

We also need to consider our privileges when we are reviewing student work. Some students have ready access to technology, while others have to travel to get access or have limited access. Some students came from affluent school districts with well-resourced course offerings while others may have attended schools in districts with more limited resources which impacted the level of coursework offered.

It is also highly, highly recommended to conduct blind assessments, whenever possible. Blind assessments mean that students' identities are unknown. Making sure that student identities are unknown is crucial to ensuring that the product we are evaluating is being judged solely by its merit.

Thank you!

That is our quick video on how to collect measures for your academic program. Please do not hesitate to reach out to the Office of Institutional Effectiveness if you have any questions or concerns!