Organizing Assessment Data Collection and Evaluation Processes

Designing a plan to collect and analyze information about student learning is about more than designing an assessment calendar—knowing what data to collect, where to collect that information, how the data will be collected, and when to collect it. While these considerations are critical to sustaining an effective departmental assessment process, they are not sufficient to derive a collective understanding of what students are learning and how to collectively respond to the data. The collective development of the elements—learning outcome statements, assessment methods, standards, and criteria—ensures that the departmental (or institutional) intention for assessment is clear. What do the members of the department or unit desire to know about student learning, its progression, and the achievements of students?

Several tasks are involved in the collection of student learning evidence and the eventual use of the results to inform educational practices:[5]

- Reach a consensus about the sampling method for the student population based on what the department or unit want to know about student learning.
- Identify the contexts and occasions for the collection of evidence for student learning using direct and indirect measures.
- Identify the assessment method(s) to be used to collect evidence of student learning.
- Score the student work products, projects, performances, or responses by applying consistent and agreed-upon criteria or standards.
- Analyze and represent the accumulated data in a manner that promotes collective interpretation of the results.
- Collectively interpret the results and make decisions based on the data, which may include but is not limited to modifying or innovating pedagogy, the curriculum, criteria, standards, or requirements, or developing alternative approaches to providing services or support programming.
- Reinitializing the assessment cycle processes to evaluate the efficacy of the innovations, adaptations, or modifications that were a result of the data collection and analysis.

Sampling

A program or unit may need to engage in sampling if the assessment of all students is impossible or overly burdensome. There are situations when all students can be assessed, such as when dedicated time is provided for assessment, such as at the end of a course or a program or on a day dedicated to such work. In other cases, sampling from a student population may be more feasible and realistic, such as when a program or unit serves a high volume of students and/or has very limited resources to engage in the comprehensive assessment of the student population. Identifying how the unit will collect data from a representative sample of students allows the unit to make inferences about the larger student population based on the assessment results.

Three random sampling processes are briefly described below:

1. *Simple Random Sampling*—In this sampling strategy, each student has an equal, random chance of being selected. The method collects a representative sample of students from the population

by virtue of the randomness of the sample, thereby reducing selection bias as each student has an equal chance of being selected. The use of random number tables or a generator is a critical element of this sampling strategy.

2. *Stratified Random Sampling*—Before applying random sampling, as in the first example, this strategy first categorizes students within the larger population. From each of these categories independent, random samples of individual students are selected. This method allows for greater control over the representativeness of the sample by ensuring that the overall sample reflects the proportion of the different categories of student in the population. Possible criteria for categorization may include but are not limited to:
    - Educational background
    - First-generation status
    - Age (e.g. traditional versus non-traditional students)
    - Underrepresented student populations
    - Levels of academic readiness
    - Levels of initial placement
    - Major program of study

3. *Cluster Random Sampling*—In this method, clusters (heterogenous groups) are defined and are representative of the student population. In cases where the random sampling of individuals may be difficult or impossible, cluster sampling can provide a means of collecting data. Where random sampling selects individuals at random, cluster random sampling selects groups at random. For example, a department offers many sections of the same class each semester. While it might be overly burdensome, perhaps impossible, to randomly sample and assess individuals across all sections of the course, it will be more manageable to randomly sample the course sections and assess those students as a cluster.

These are a few brief examples of sampling strategies. For more information concerning sampling strategies and protocols, please contact the Director of Assessment.

Remember: the validity of the inferences made about a student population from the assessment of a sample directly depends upon how representative the sample is of the population. The larger the random sample, the greater the confidence in the generalizability of the results.

Note: it is highly recommended that departments and units leverage technology, where possible, to maximize the collection and analysis of student learning data. Sometimes the application of functions in a learning management system (such as Canvas) can make the assessment of a student population much more manageable and reduce or obviate the need for sampling protocols.

Evidence Collection

Establishing a timeline for the collection of evidence is a critical step in the assessment process. A timeline will be designed to address the question(s) the department or unit has regarding student learning. The nature of the evidence you collect will be predicated upon these questions also. As a starting point, examine the program curriculum map. Note the points in the curriculum where students should be able to demonstrate mastery of PLOs by the end of a course. These courses are an effective place to initially focus programmatic assessment work, for they present opportunities to explore the

knowledge and skills of soon-to-be graduates. Pairing these data with assessment data from other parts of the curriculum can provide insights. For example, if a department wanted to track how students are doing at three key points in the curriculum—at the end of a core, gateway 101 course, at a required mid-point class (perhaps a 200- or 300-level course), and at the end of a 400-level capstone, the assessment timeline would focus on these three key points in the curriculum. In this example, the department may have decided to use a set of common assessments corresponding to each of the three courses to be able to compare results across all course sections. A longitudinal analysis could then track student PLO attainment across the curriculum over a time to investigate issues such as the efficacy of course sequencing, student attainment of key learning outcomes, and the need for additional or innovative student support at critical times in the curriculum.

Consider the connections between courses and learning outcomes in curriculum maps. Attempt to reach an internal consensus concerning when students should be able to demonstrate understanding, proficiency, and finally eventual mastery, of those outcomes. Once these milestones in the curriculum have been determined, an assessment timeline can be designed to investigate how effectively students are able to demonstrate their learning at these key points. Decisions about how to collect evidence of student learning will follow. Recall that good assessment practice suggests that each assessed outcome should have at least one direct measure and an indirect measure. For example, a course might pair a summative writing assessment project with a written reflection concerning the skills the students are demonstrating. Students might be asked to self-assess and engage in reflection and metacognition concerning their performance before or after being provided feedback from the direct assessment of their work. In a different example, students might engage in a series of formative assessments (i.e. skill demonstrations graded with a rubric or checklist) prior to a summative, comprehensive certification examination. A mid-semester survey might ask students to respond to a series of open-ended questions about their strengths and weaknesses. In this example, two direct assessments, one internal and another external (skill demonstrations followed by a certification examination), are paired with an indirect assessment—a survey.

The details of an assessment timeline and its methods will be predicated on the needs and interests of the program or unit. Collaborative planning and discussion concerning an assessment timeline, goals for students, measures, and standards or criteria for success are vital for the collective examination and analysis of student learning evidence to affect a shared understanding of student progress. A unit should plan to evaluate all its learning outcomes in one assessment cycle and build in time to come together as a unit for collaborative discussions. Leverage extant opportunities to gather data; most programs already have embedded assessments and opportunities for students to receive and provide feedback. Leveraging or modifying extant processes allows faculty members to use valuable time to plan, implement, and evaluate changes and improvements intended to improve student learning. Dedicating time for collaborative discussions focused on processes and improvements will help to sustain the evidence collection and evaluation work and move assessment forward as an inquiry-based practice.

To reiterate the key steps in the evidence collection planning and implementation process:

- Collaboratively determine the question(s) the department or unit would like to address in its assessment cycle.
- Determine the forms of student learning evidence that will be necessary to address the questions.

- Identify the who, where, and when of an assessment timeline for the collection of evidence by developing or reviewing a curriculum map.
- Identify opportunities to utilize extant student information and assessments and where development is needed to collect reliable, valid evidence.
- Collaborate to articulate common goals for student learning and the standards and criteria that will guide the evaluation of the student learning evidence.
- Communicate internally to ensure all stakeholders understand the intent, responsibilities, and processes for evidence collection.
- Build-in dedicated time for stakeholders to collaborate and communicate in the assessment timeline.
- Implement the timeline to collect the student work, information, and artifacts that will be assessed.

Determining Criteria & Standards

Arriving at a consensus concerning the criteria and standards for programmatic assessment can be a challenge. Instructors have priories for student learning that may differ from each other, and it is for this reason that collaboration is essential. Focus on the learning outcomes the program or unit has articulated.

Begin by determining the minimum level of competency students must be able to demonstrate to have mastered each assessed outcome. Consider: what does success on each outcome look like? How will the members of the unit or department know when they see it? The ability to articulate specific criteria for success is an important starting point. Once these criteria have been determined, consider setting a target for students' collective performance. Consider the question: what proportion of students do we expect to meet the standards for success? Remember that context matters. Students who are about to graduate from a program should have a different standard for success than those who are just beginning or who are in the middle of the program.

Consider the following points regarding the expectations for student success:[1]

- Should every student meet the standard? Would the members of the unit or department be satisfied if 90% of students were able to meet the standard? What is the minimum percentage that can be accepted as successful?
- Would you be satisfied if every student met the minimum standard but no higher?
- Vary your targets depending on the circumstances. Some fields have less tolerance for error than others, but all students should graduate having mastered the fundamental skills.
- Express targets as percentages rather than averages.
- Consider multiple scores: what percentage will score minimally acceptable and what percentage will score above average, proficient, or exemplary?

Benchmarking student success can be a valuable way to determine assessment targets and goals. Benchmarking is the comparison of our results to one or more external or internal standards. Some examples of useful benchmarks may include:[1]

- Local standards (competency-based/criterion-referenced)—Are students meeting our own standards?
- External standards (competency-based/criterion-referenced)—Are students meeting the standards set by someone else?
- Internal Peer (comparative/norm-referenced)—How do students compare to their peers within our course(s), program, or college?
- External Peer (comparative/norm-referenced)—How do students compare to peers at other colleges?
- Best Practices (best-in-class, internal or external)—How do students compare against the best of their peers?
- Value-Added (growth, longitudinal, improvement)—How much have students grown because of our efforts? Are our students improving? (the comparison of two or more points within the same students)
- Historical Trends (growth, longitudinal, improvement)—How do students compare against the established historical trend? Is our program improving? (the comparison of two or more points between different students)
- Strengths-and-Weaknesses (comparative, improvement)—Where are students' areas of relative strength and weakness? Where should we focus our efforts? (the comparison of subscores within the assessment of each student)
- Capability (potential)—How are students doing as compared to their capabilities? (avoids generalized student comparisons but relies on an accurate assessment of students' starting capabilities)

When examining evidence of higher-order cognitive skills, consider using a rubric as a scoring guide to evaluate the evidence of student learning. Papers, projects, productions, presentations, field experiences and similar forms of demonstrative evidence are often utilized, and these forms of evidence should be assessed using consistent criteria that are clear and transparent to students. Rubrics are a list, chart, or matrix that articulates the criteria that will be used to evaluate the completed student work. A rubric lists the qualities, criteria, or competencies that should be addressed in the student work and often supplies a scoring or grading guide in the form of success categories or descriptors. Here are some advantages of using a rubric for scoring purposes:[1]

- Rubrics can clarify broad or vague learning goals or outcomes for students and instructors.
- Rubrics help students understand the assignment expectations by describing the potential levels of success or failure.
- Rubrics are a teaching tool that can help students improve via feedback and self-evaluation.
- Rubrics can promote improved student performance by articulating instructor values and expectations for the assignment.
- Rubrics make scoring the assessment faster and easier.
- Rubrics also make scoring more reliable, accurate, unbiased, and transparent to students.
- Rubrics improve the ability of students to use the feedback they receive during or after completing the assignment.

- Rubrics also improve the feedback to instructors and staff by establishing clear standards or criteria that can chart progress and the learning growth of students after embedding interventions or innovations.
- Because of these qualities, rubrics improve the transparency of the evaluation process to instructors and students thereby reducing the likelihood of arguments over scores.

Why Grading Is Different Than Assessment

At this point, a couple of questions might arise. Aren't we already doing all of this in our grading? Can't we just use grades as assessment evidence of student learning? The answers to these questions are: yes we should to the first and probably not to the second. There are several reasons why assessment is different from grading, but they do share a few traits: both require evidence; use measures; and are evidence driven. However, the focus of assessment are the skills, knowledge, and values that students can demonstrate as a result of their learning, and the level of success is predicated on students' demonstrations. Grades are often determined by additional factors such as attendance, the timely completion of assignments, technical errors, formatting errors, typos, extra credit, make-up work, or classroom participation. Many extraneous factors can affect grades whereas assessment is unconcerned with these factors. Additionally, students tend to focus on grades as a mark of completion of a course or unit of instruction, and the grade encompasses all the learning, and other factors, that took place in a course. Assessment results are focused, ongoing measures of student learning that can and should be compared over time. Nevertheless, assessment and grading are complementary processes, but the assessment of the student work will be entirely focused on the demonstration of knowledge, skills, and values. Grades are predicated on the criteria residing at the core of assessment but can go beyond these criteria to consider contextual and individualized factors. This is not to suggest that grades lack meaning or are disconnected from learning, quite the opposite, but grades often encompass concerns that reside outside of the assessment of our learning outcomes for students. Analyzing grade trends, particularly Ds, Fs, and withdrawals is certainly important, and this sort of analysis can be informative of student and instructor behaviors and values. However, the content of learning cannot be clearly intuited by examining a grade distribution. A possible exception to this is specification-based grading. For more information, please contact CETL for more information about specification-based grading.