

Evaluating Student Learning

Assessment of Outcomes

General types of assessment:

Summative versus Formative evaluation

- Different functions: Summative assessments measure “completed” learning and are useful to assess student success on outcomes at the completion of the course or program.
- Formative assessments measure learning “in progress” and allow the teacher and learners to modify their actions to improve future success in the course/program.
- These types of assessment have different affective impact on students.

Norm-referenced (curved) versus criterion-referenced evaluation

- May include "extra credit" points
- Norm-referenced evaluations put students in competition with their classmates. It may be based on pre-determined rubrics (e.g. the best effort in a particular group = A+ and other grades are proportional) or may be done by looking for "breaks in the distribution".
- Criterion-referenced evaluation sets criteria based on either generally accepted standards or an understanding of what is needed by students in the future (e.g. in subsequent classes for which the current class is a prerequisite).

Subjective versus "countable" scoring

- Some disciplines and/or assessment techniques lend themselves to “countable” scoring (e.g. multiple choice tests in science).
- Some disciplines and/or assessment techniques may be hard to score by a count, and are more often scored subjectively (e.g. creative writing essays graded via a rubric)

Educational measurement

Validity: does the assessment tool measure what it is meant to measure?

- Content validity
- Alignment with objectives
- Expectations of students
- Range of methods
- Barriers to cheating

Reliability: are results consistent in matched groups?

- Consistency of how questions are "read"
- Consistency in grading
- Unambiguous directions

Practicality: is the assessment feasible in terms of time and resources?

- Skill and time needed to produce assessment instrument and analyze outcomes
- Resources (e.g. for a practical exam) needed to use the assessment tool

Positive effect: does it increase learning and success on outcomes?

- Does it measure surface or deep learning? Does it encourage deep learning?
- Will the results be used to improve the teaching/learning process in the future?
- Does the use of the assessment lead to dialog about what and how our students are learning?

I. Direct observation of skills or products

A. Definition:

This includes (1) any assessment tool that involves directly observing students as they utilize the skills taught in the class and (2) evaluation of student work that was produced by directly using the skills taught in the class. [Note: Exam questions are often indirect measures – that doesn't make them less valid, it just puts them in another section of this paper.]

B. Examples of types of assessment tools in this category:

Role plays, presentations, etc.

- These are the best way to assess some outcomes related to complex skills because they are direct measure of the skill.
- Reliability problems may interfere with the use of these techniques for outcomes assessment. Detailed rubrics, checklists, and rating forms, are helpful in solving this problem.

Measures of student participation in activities on campus and use of campus resources:

- These are useful to assess student service outcomes and may provide direct assessment of some skills (e.g. ability to fill out financial aid forms, ability to utilize campus resources, etc.)
- Can be indirect measures, rather than direct measure of student skills. For example, measures of student activity/participation on campus may provide indirect measures of college-wide or program outcomes related to student interaction with the college.

Guided discussion

- Direct observation of student response during guided discussion in groups can make evident both thinking skills and foundation knowledge.
- Guided discussion is relatively difficult to grade or to use to produce quantitative outcomes assessments. It can be used to provide a reflective qualitative assessment of some outcomes if rubrics are used to define levels of success for the group.

Performances, structured practical assessments, and demonstrations of specific skills:

- These effectively assess skills which are usually psychomotor in nature (e.g. lab equipment use, dance steps, construction techniques).
- There may be validity problems. Validity may be increased by specifying criteria on a rubric, providing clear instructions for the tasks, and allowing a task to be performed multiple times.

Portfolios

- Collection of student work in portfolios is an important student centered assessment in the arts, and has been used in other disciplines.
- Portfolios may have trouble with reliability and validity in some disciplines. In this case, it is helpful to (1) have a clear evaluation system and rubrics (2) have students do more than just "turn in" items, e.g. have them explain and evaluate the importance of each item in the portfolio.

Projects, reports, papers, etc.

- The assessment tools in this category include research papers, lab reports, inquiry projects, creative projects, term papers, etc. These are common tools in evaluating student learning in many disciplines and, thus, there are many models to follow.
- Because they typically represent a major amount of work, these assessment tools can reflect more than one skill – in that case, a rubric may be needed to map outcomes to work product.

C. Summary of strengths and weakness of tools in this category

Strengths:

- Provide direct measures of student skills and can involve "authentic" problems (similar to those done in the professional pursuit of the discipline).

Challenges:

- May have reliability and/or validity problems unless good rubrics are used since evaluation of student work of some types may have a strong subjective component.

Use as outcomes measure

- Works well for assessment of psychomotor skills (e.g. use of lab equipment, ability to fill out financial aid forms, music performance, public speaking abilities, etc.)

Possible documentation

- Requires careful implementation in order to make it easier to produce quantitative data for outcomes assessment reporting.

Examples from CRC:

II. Assessment via Exam and Homework

Questions:

A. Definition:

This category of tools includes (1) exams and quizzes ranging from blue-book essay exams, to multiple choice objective exams, to oral exams and (2) homework problem sets that are composed of items similar to those found on exams.

B. Examples of types of assessment tools in this category:

Essays (blue book exams, long essays as homework, etc.)

- Essays can measure outcomes linked to critical thinking and writing abilities, or those reflecting recall, analysis, evaluation, and synthesis skills.
- These are time consuming to grade, and, since marking may be subjective; rubrics are nearly essential.

Short-answer (including simple calculation) questions.

- These are moderately easy to grade and not too time consuming to write. Rubrics are helpful, especially when there is a need to be very clear about what alternatives of the answer are acceptable.
- It has been suggested that these are "not suitable for testing complex learning outcomes" (Barbara Gross-Davis, Tools for Teaching). I disagree, but note that it takes some thought to write items that measure critical thinking skills or problem solving.

Objective (multiple choice, true/false, matching) questions.

- These provide a high reliability of scoring and are easy to grade, but are time consuming to write well.
- Considerable skill is required to write objective tests that measure critical thinking skills or problem solving. Context-dependent questions often help with this.

Oral exams

- If questions are unrestricted, the reliability is very low and good rubrics and techniques that increase examiner consistency (check lists, multiple observers, etc.) are important.
- This technique can be used to assess outcomes related to interpersonal skills (e.g. interview skills) and speaking skills (e.g. foreign language classes).

C. Summary of strengths and weakness of tools in this category

Strengths:

- Less likely to have reliability and/or validity problems than other types of assessment tools.
- Commonly used and recognized in many disciplines; many models available.

Challenges:

- Requires careful planning in order to provide direct measures of student skills or involve “authentic” problems (similar to those done in the professional pursuit of the discipline).

Use as outcomes measure

- Exam or homework questions can be linked directly to outcomes for assessment purposes. Each question can be linked to a specific outcome or objective and, thus, an item analysis can provide an outcome assessment.
- In a course, mapping of content to outcomes may allow each exam to focus on one or a few outcomes. Alternatively, the final exam or final homework set can be used as a summative assessment of all course outcomes.

Possible documentation

- Data collection for outcomes assessment analysis may involve collection of student work examples or quantitative data based on item analysis of exams, quizzes, etc.
- Some types of exams allow for many questions and countable scoring allows relatively easy data tracking for outcomes assessment
- Requires careful implementation in order to make it easier to produce quantitative data for outcomes assessment reporting.

Examples from CRC:

III. Student self-assessments

A. Definition:

This category of tools includes a variety of tools that ask students to assess themselves in terms of how well they can perform the skills and abilities learned in the activity or course. It also includes measures of student self-confidence regarding their ability to use the skills learned in the activity or course.

B. Examples of types of assessment tools in this category:

Classroom assessment techniques such as short surveys immediately following an activity:

- These provide quick feedback and allow for modifications of teaching strategies in a rapid turn-around basis.
- While good for formative assessment, these tools may not be as useful for summative assessment.

Student journals or reflective essays

- These provide rich information on student progress and allow the assessment of affective outcomes in a detailed way.
- Only a few disciplines (e.g. ECE) have a tradition of using these tools.

Surveys conducted at the completion of a course, activity, or program.

- These are very useful for summative assessment, particularly of program outcomes.
- Long surveys are time intensive and require substantial resources to implement.

Student self-grading of an assignment or project:

- Students benefit from learning how to assess their progress and can apply this to future work.
- Validity can be major problems with student self assessments unless students are trained in how to assess their own success.

Individual discussions between a student and a teacher:

- One-on-one discussions provide a rich source of affective information.
- It requires substantial planning to assess outcomes using these tools.

C. Summary of strengths and weakness of tools in this category

Strengths

- Students benefit from learning how to assess their progress via self assessment, so these tools can be used as a formative assessment. As such they can be used to help generate a rubric for future assessments and to provide feedback or outcomes assessment without producing grades (e.g. as is done with sports teams).
- Using this type of assessment can influence a teacher to make the learning outcomes more explicit to students, which can enhance instructional quality and learning
- The method can be fairly quick and easy to implement

Weaknesses

- Validity and reliability can be major problems with student self assessments unless students are trained in how to assess their own success.

Use as outcomes measure

- Including a self-assessment on an exam (perhaps for bonus points) can be an easy way to collect this type of information.
- Administering surveys throughout or at the end of the course provides a lot of quantifiable data.
- Administering surveys in capstone classes of a vertical program can serve both program and course assessment functions.

Possible documentation

- Collated survey results can be collected and saved.

Examples from CRC