

Five Models for Assessing Student Learning

What follows is a summary of the most effective and efficient ways to assess student learning.

Improved Learning

The basic assessment question to be answered is the degree to which improved learning has been achieved as a result of the course redesign. Answering this question requires comparisons between the student learning outcomes associated with a given course delivered in its traditional form and in its redesigned form.

I. Establish the method of obtaining data

A. Pilot Phase

This comparison can be accomplished in one of two ways:

1. Parallel Sections (Traditional and Redesign)

Run parallel sections of the course in traditional and redesigned formats and look at whether there are any differences in outcomes—a classic "quasi-experiment."

2. Baseline "Before" (Traditional) and "After" (Redesign)

Establish baseline information about student learning outcomes from an offering of the traditional format "before" the redesign begins and compare the outcomes achieved in a subsequent ("after") offering of the course in its redesigned format.

B. Full Implementation Phase

Since there will not be an opportunity to run parallel sections once the redesign reaches full implementation, use baseline data from a) an offering of the traditional format "before" the redesign began, or b) the parallel sections of the course offered in the traditional format during the pilot phase.

The key to validity in all cases is a) to use the same measures and procedures to collect data in both kinds of sections and, b) to ensure as fully as possible that any differences in the student populations taking each section are minimized (or at least documented so that they can be taken into account.)

II. Choose the measurement method

The degree to which students have actually mastered course content appropriately is, of course, the bottom line. Therefore, some kind of credible assessment of student learning is critical to the redesign project.

Five measures that may be used are described below.

A. Comparisons of Common Final Exams

Some projects use common final examinations to compare student learning outcomes across traditional and redesigned sections. This approach may include sub-scores or similar indicators of performance in particular content areas as well as simply an overall final score or grade. (Note: If a grade is used, there must be assurance that the basis on which it was awarded is the same under both conditions—e.g., not “curved” or otherwise adjusted.)

1. Internal Examinations (Designed by Faculty)

Parallel Sections Example: “During the pilot phase, students will be randomly assigned to either the traditional course or the redesigned course. Student learning will be assessed mostly through examination developed by departmental faculty. Four objectively scored exams will be developed and used commonly in both the traditional and redesigned sections of the course. The exams will assess both knowledge of content and critical thinking skills to determine how well students meet the six general learning objectives of the course. Students will take one site-based final exam as well. Student performance on each learning outcome measure will be compared to determine whether students in the redesigned course are performing differently than students in the traditional course.”

Before and After Example: “The specifics of the assessment plan are sound, resting largely on direct comparisons of student exam performance on common instruments in traditional and re-designed sections Sociology faculty have developed a set of common, objective, questions that measure the understanding of key sociological concepts. This examination has been administered across all sections of the course for the past five years. Results obtained from the traditional offering of the course will be compared with those from the redesigned version.”

2. External Examinations (Available from Outside Sources)

Parallel Sections Example: “The assessment plan involves random assignment of students to “experimental” (redesign) and “control” (traditional) groups operating in parallel during the pilot phase of implementation. Assessment will measure student success against established national (ACTFL) guidelines, including an Oral Proficiency Interview that has been widely validated and is also in use in K-12 settings. This will allow the university to compare results of the redesign to baseline literature about results of traditional pedagogy, to compare the added effect of use of multimedia to the same material delivered conventionally, and to gauge the effect of new remediation strategies on student performance.”

Before and After Example: “The centerpiece of the assessment plan with respect to direct measures of student learning is its proposed use of the ACS Blended Exam in Chemistry in a before/after design—administered to students in both traditional and redesigned course environments. A well-accepted instrument in chemistry, the ACS Exam has the substantial advantage of allowing inter-institutional comparisons according to common standards.”

B. Comparisons of Common Content Items Selected from Exams

If a common exam cannot be given—or is deemed to be inappropriate—an equally good approach is to embed some common questions or items in the examinations or assignments administered in the redesigned and traditional delivery formats. This design allows common baselines to be established, but still leaves room for individual faculty members to structure the balance of these finals in their own ways where appropriate. For multiple-choice examinations, a minimum of twenty such questions should be included. For other kinds of questions, at least one common essay, or two or three problems should be included.

Parallel Sections Example: “The primary technique to be used in assessing content is common-item testing for comparing learning outcomes in the redesigned and traditional formats. Traditional and redesigned sections will use many of the same exam questions. Direct comparisons on learning outcomes are to be obtained on the basis of a subset of 30 test items embedded in all final examinations.”

Before and After Example: “The assessment plan must address the need to accommodate a total redesign in which running parallel sections is not contemplated. The plan calls for a “before/after” approach using 30 exam questions from the previously delivered traditionally-configured course and embedding them in exams in the redesigned course to provide some benchmarks for comparison.”

C. Comparisons of Pre- and Post-tests

A third approach is to administer pre- and post-tests to assess student learning gains within the course in both the traditional and redesigned sections and to compare the results. By using this method, both post-test results and “value-added” can be compared across sections.

Parallel Sections Example: “The most important student outcome, substantive knowledge of American Government, will be measured in both redesigned and traditional courses. To assess learning and retention, students will take: a pre-test during the first week of the term and a post-test at the end of the term. The Political Science faculty, working with the evaluation team, will design and validate content-specific examinations that are common across traditional and redesigned courses. The instruments will cover a range of behaviors from recall of knowledge to higher-order thinking skills. The examinations will be content-validated through the curriculum design and course objectives.”

Before and After Example: “Student learning in the redesigned environment will be measured against learning in the traditional course through standard pre- and post-tests. The university has been collecting data from students taking Introduction to Statistics, using pre- and post-tests to assess student learning gains within the course. Because the same tests are administered in all semesters, they can be used to compare students in the redesigned course with students who have taken the course for a number of years, forming a baseline about learning outcomes in the traditional course. Thus, the institution can compare the learning gains of students in the newly redesigned learning environment with the baseline measures already collected from students taking the current version of the course.”

D. Comparisons of Student Work Using Common Rubrics

Naturally occurring samples of student work (e.g. papers, lab assignments, problems, etc.) can be collected and their outcomes compared—a valid and useful approach if the assignments producing the work to be examined really are quite similar. Faculty must have agreed in advance on *how* student performance is to be judged and on the *standards* for scoring or grading (a clear set of criteria or rubrics to grade assignments.) Faculty members should practice applying these criteria in advance of the actual scoring process to familiarize themselves with it and to align their standards. Ideally, some form of assessment of inter-rater agreement should be undertaken.

Parallel Sections Example: “Students complete four in-class impromptu writing assignments. A standard set of topics will be established for the traditional and redesigned sections. A standardized method of evaluating the impromptu essays has already been established and will be used in grading each assignment. The essays are graded by using a six-point scale. The reliability measure for this grading scale has been established at 0.92. Additionally, each paper is read by at least two readers. The grading rubric will be applied to the four standard writing assignment prompts administered in parallel in simultaneously offered redesigned and traditional course sections.”

Before and After Example: “The assessment plan is quite sophisticated, involving both “before/after” comparisons of student mastery of statistics concepts in the traditional course and the redesigned course. The design itself involves direct comparisons of performance on common assignments and problem sets using detailed scoring guides (many of which were piloted and tested previously and are thus of proven utility). Because the department has already established and benchmarked learning outcomes for statistics concepts in considerable detail, and uses common exercises to operationalize these concepts, the basis of comparison is clear.”

E. Comparisons of Course Grades Using Common Criteria

Course grades may be used as the measure of learning if—and only if—grades are assigned on the basis of comparable performances on common instruments using common grading standards. Faculty must have agreed in advance on standards for scoring or grading.

Parallel Sections Example: “The department utilizes common grading criteria that address topic and purpose, organization and coherence, development, style, and grammar and mechanics. Specific descriptions within each of the areas are provided to distinguish between grades of A, B, C, D, and F, and faculty members are trained in the interpretation of the criteria. The criteria were established collectively and are applied across all sections of College Composition.”

Before and After Example: “Assessment will use before/after comparisons of student performance in the traditional and redesigned settings. The traditional and redesigned sections of the course will use the same textbook assignments and will pursue the same department learning goals. Quizzes, hour exams, and lab assignments will test student knowledge of the same material, and the final exam will include common multiple choice questions for all course sections. Direct measures of achievement will be based on common final examinations that have been administered for many years in the traditional courses. The team will track the proportion of students who receive a C or better to see if student success rates improve.”

Tips

- Avoid creating “add-on” assessments to regular course assignments such as specially constructed pre and post-tests. These measures can raise significant problems of student motivation. It is easier to match and compare regular course assignments.
- If parallel sections are formed based on student choice, it would be a good idea to consider whether differences in the characteristics of students taking the course in the two formats might be responsible for differences in results. Final learning outcomes could be regressed on the following: status (full vs. part-time); high-school percentile rank; total SAT score; race; gender; whether or not the student was taught by a full-time or part-time faculty member; and whether or not the student was a beginning freshman.
- In addition to choosing one of the five required measures, the redesign team may want to conduct other comparisons between the traditional and redesigned formats such as:
 1. Performance in follow-on courses
 2. Attitude toward subject matter
 3. Deep vs. superficial learning
 4. Increases in the number of majors in the discipline
 5. Student interest in pursuing further coursework in the discipline
 6. Differences in performance among student subpopulations
 7. Student satisfaction measures