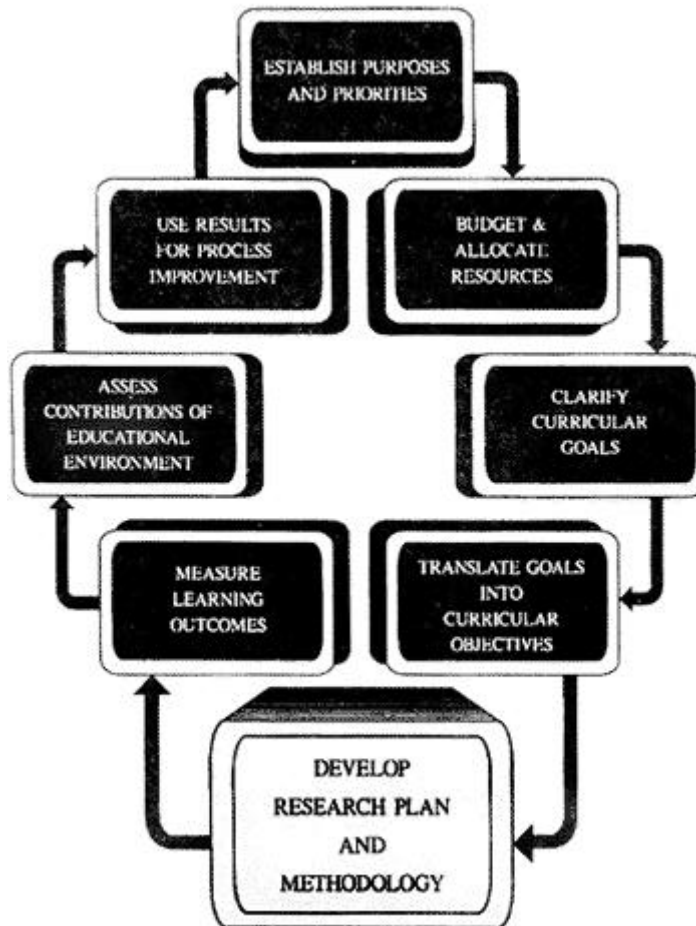


Assessment for the New Curriculum: A Guide for Professional Accounting Programs

Chapter 8 Developing the Research Plan and Methodology



The assessment research plan translates the purpose of assessment ([Chapter 4](#)) into a method or methods of inquiry to fulfill that purpose. The assessment committee determines the nature and scope of the inquiry, then selects appropriate research methods, analytical techniques, variables to be included, measures needed, subgroups to be studied, and sample sizes.

This chapter outlines options and offers suggestions for determining what is necessary, sufficient, and feasible within the resources of the department. It is intended to clarify the role of traditional research design in assessment, and to suggest research approaches and issues especially relevant in the accounting context. [Chapter 9](#) follows with guidelines for selecting and/or developing measurement instruments. Specifically, this chapter:

- Identifies criteria for selecting research designs and methods
- Outlines basic paradigms for educational assessment
- Identifies design issues especially relevant to educational assessment
- Suggests designs that allow the assessment program to develop gradually
- Offers practical, credible ways to obtain data for assessment
- Notes ethical issues related to educational research

8.1 Criteria for Selecting the Research Approach

A department contemplating an assessment initiative can set its sights on publication-quality research, on informal studies for internal use, or on a degree of sophistication intermediate between these two extremes. A primary concern in assessment is whether the benefits will justify the costs involved.

Most fundamentally, the design should yield results that serve the purposes identified for assessment, which in turn should be formulated to respond to the information needs of key stakeholders (as described in [Chapter 4](#)). The scope and formality of the inquiry depend on the purposes of the assessment, who will use the results, and what is at stake. This section proposes five criteria for selection of research methods tailored to departmental needs.

Four criteria proposed by the Joint Committee on Standards for Educational Evaluation (1981) offer a starting point:

- *Relevance to Policy Decisions and Planning*: The methodology chosen should yield results that will help to answer the policy or planning questions that prompted the study in the first place.
- *Feasibility*: The scope of the plan should fall well within the resources and time available.
- *Credibility*: The technical quality of the study should be sufficient to allow reasonable certainty in drawing conclusions based on the results obtained.
- *Propriety*: The approach should conform to legal and ethical standards for the conduct of research on human subjects (adapted from Davis, 1989, p. 18).

Assessment differs from traditional educational evaluation in one important respect: it should be designed to *assist the learner*. This aspect of the design affects both student and faculty willingness to participate. The assessment design should therefore respond to an important additional criterion:

- *Benefit to Participants*: The approach should provide some immediate benefit for all those who participate, but especially the students and faculty most directly involved.

Together, these criteria suggest a flexible approach to assessment research design. For example, the "credibility" criterion suggests that good design is important, but the "feasibility" criterion will often preclude obtaining publication-quality data. The sampling plan and methods should nonetheless meet credibility standards of the department for internal use, since the results may influence allocation of resources.

Credibility is enhanced by "triangulation," that is, the use of data from several sources to enhance interpretation of data from any one source and to strengthen the plausibility of inferences from all sources. Triangulation cannot overcome all the weaknesses of non-experimental designs, nor can it compensate for poorly designed instruments. However, triangulation does help faculty to interpret data by providing more than one perspective on the question at hand.

The assessment committee can facilitate integration of the criteria by recommending approaches that incorporate assessment measures within the normal requirements of instruction. Referred to as "course-embedded" assessment, this strategy offers a practical way to obtain data within a variety of research designs. Because it makes assessment an integral part of teaching, course-embedded assessment facilitates faculty participation and interest (Ewell, 1991b). It also benefits participating students since they receive instructor feedback on the work they produce for assessment. Evaluating students' work on clearly identified performance criteria also benefits them by promoting the important lifelong learning skill of self-assessment. These benefits are increased when the professor introduces the performance criteria and discusses their application to students' work (Loacker and others, 1984; Loacker, Cromwell, and O'Brien, 1986).

When seeking a research approach that responds to the five criteria, the assessment committee should consider lessons learned by the organizers of the Harvard Assessment Seminar:

An incorrect expectation was that larger-scale, elaborate studies would be especially interesting to most participants [in the Seminar]. We misjudged this badly. and now believe that less can be more. Sometimes a small effort with a quick turnaround, if well done, is the most effective research of all. This is especially true when the findings from a project may

affect a policy decision and the person in charge of policy has specially requested research to help shape the decision. (Light and others, 1990, p. 236)

8.2 Research Options for Educational Assessment

Two modes of inquiry contribute to a balanced and informative assessment portfolio:

- *Concurrent* inquiry to monitor and respond to program outcomes and client satisfaction during implementation
- *Retrospective* inquiry to determine program effectiveness and identify contributing factors, generally upon completion of a program cycle

Concurrent inquiry is similar to "formative" evaluation (that is, contributing to the development and improvement of the program). *Retrospective* inquiry is similar to "summative" evaluation (providing data to judge the merits of the program) because it is implemented at the end of a program cycle. However, retrospective inquiry should contribute to continuing improvement of the program in future cycles. The appropriate balance between concurrent and retrospective inquiry depends on the purposes and audiences for the assessment.

Within these two modes of inquiry, research design options for educational assessment fall into four general paradigms: descriptive, relational, experimental, and quasi-experimental studies (adapted from Light, Singer, and Willett, 1990; see also Williams and others, 1988):

- *Descriptive* research designs examine student characteristics, the educational environment, or learning outcomes separately and at a single point in time. Descriptive studies may use either qualitative or quantitative methods, or a combination of the two.

Descriptive studies answer questions such as, "What are our graduates' strengths and weaknesses with respect to a particular learning outcome?" or "What skills related to use of information technology do our students have on entry to the program?"

- *Relational* studies examine the degree of association between student characteristics, the educational environment, and/or learning outcomes. Relational models include cross-sectional, simple correlational, and multivariate designs (ANCOVA, multiple regression, causal modeling).

Relational studies answer questions such as, "What subgroup(s) of students would be excluded from the program if we required demonstrated competency in writing?" or "What are the differential effects of an innovative program for students with different characteristics?"

- *Experimental* studies use random assignment to experimental and control groups to establish *causal* links between the educational environment (independent variables) and learning outcomes (dependent variables). The most common design is the "pretest-posttest control group design"; however, the "posttest-only control group design" is equally valid if students have been randomly assigned to groups (Campbell and Stanley, 1966).
- *Quasi-experimental* designs are often necessary in educational settings, where it is difficult to assign students randomly to groups. Quasi-experimental designs reduce threats to validity and reliability, for example, by using repeated measurement with an experimental condition introduced at one or more points of measurement (the "time series" design; Campbell and Stanley, 1966) or by including student characteristics as covariates in the analysis to reduce the effects of self-selection bias (Light and others, 1990).

Experimental and quasi-experimental studies answer questions such as, "Which of three instructional methods has the greatest impact on professionally related writing skills of students in upper-division accounting courses?"

Descriptive designs are most often associated with concurrent (formative) inquiry. Retrospective inquiries employ the full range of design paradigms, but especially the correlational and experimental or quasi-experimental methods.

8.3 Design Issues in the Assessment of Learning Outcomes

Familiarity with research design principles and statistical methods is an important asset brought by accounting faculty to the task of assessment. Accounting faculty may be less familiar, however, with design issues associated with studies of learning outcomes. This section identifies several such issues.

Issue 1: Interactions between student characteristics and the educational environment: Educational research designs frequently take into account the interaction between student characteristics and the educational environment. Researchers today are less likely to ask, "Which method is best?" and more likely to ask, "Which method is best for whom?" (Snow and Peterson, 1980, p. 2). For example, students who transfer into accounting are more interested in expanded learning outcomes (such as those advocated by the AECC) than those who began their college careers as accounting majors (Inman, Wenzler, and Wickert, 1989). A study of students' responses to an innovative junior-year program might therefore take into account both transfer status and initial interest in the expanded outcomes.

Many student characteristics can affect the outcomes achieved by a program. For example, prior knowledge (both general and task-specific) is a good predictor of students' performance on learning tasks. This well-established relationship underlies the use of measures such as admission test scores and GPA as covariates in studies of college outcomes (see for example, Astin, 1993; Pascarella and Terenzini, 1991). Similarly, prior interest in the subject is an important predictor of satisfaction (Marsh, 1980). Accordingly, self-reported motivation to take the course is used as a control variable in a widely used nationally normed instrument for assessing students' responses to instruction (Center for Faculty Evaluation and Development, 1975).

Less familiar to many faculty but important in fostering lifelong learning are characteristics such as students' preferred strategies for learning and their motivational orientation. For example, students who are motivated to learn independently benefit more from innovative learning experiences than those motivated to learn by conforming (Domino, 1971). In one study, assigning students to a section designed for their motivational orientation would have increased the scores of 44% of the sample by 12 to 25 percentile points. An additional 10% of the students would have improved by as much as 40 percentile points (Peterson, 1979).

Learning styles also influence students' ability to benefit from different educational environments. For example, "Sensors," who process information in terms of concrete details, prefer an emphasis on factual information and standardized procedures. "Intuitives," who process information in terms of connections and possibilities, prefer learning environments that encourage them to develop their own ideas (using the Myers-Briggs Type Inventory; Schroeder, 1993). In a recent study comparing CAI and lecture methods in elementary accounting, Sensors performed better with lecture instruction than CAI, while the reverse was true for Intuitive students (Ott, Mann, and Moores, 1990). Most accounting students prefer the Sensing mode (Geary and Rooney, 1993).

Including student characteristics in the research design improves the chances that program effects on learning outcomes will be detected rather than averaged out. More importantly, use of student characteristics in designing instruction may increase overall student success in the program. At the same time, students should be challenged to expand their repertoire of learning strategies for greater professional adaptability.

Issue 2: The trend toward "Naturalistic" modes of inquiry: To date, few departments have undertaken formal experimental studies of curricular innovations and outcomes. Among the AECC grant institutions, the University of North Texas (Bayer and others, 1993) and Arizona State University (McKenzie, 1991) were noteworthy for their use of experimental designs with pre- and posttesting and comparison groups.

Discussing the feasibility of a traditional control-group design, BYU faculty involved in the AECC-funded junior year core program comment:

Desirable as it may seem at first glance, this type of design would break down were it invoked in evaluating the junior core program. Since the technical competencies in the new program are somewhat different from those in the traditional program, there is no way to obtain a comparable control group. Further, we could not simply administer a pre-test at the beginning

of the year and a post-test at the end. Without a control group, there is no way to attribute any change to the program. Finally, even if a suitable control group could be found and the new group showed greater gains, we would have no way of deciding what aspect of the program made the difference. Too many variables are at work in a new program, not the least of which is the Hawthorne effect (BYU Vol. I, 1992, p. 61).

The desire to benefit students and increase feasibility has led faculty to adopt classroom-centered, naturalistic assessment strategies such as capstone experiences, portfolios, and use of faculty-designed, course-embedded instruments rather than standardized achievement test. A trend toward the use of naturalistic approaches is increasingly evident in surveys of current assessment practice (Ewell, 1991b). The assessment methodologies used by BYU faculty, described in Section 8.4, illustrate this trend.

Issue 3: Design for results with practical significance: Research that is not directly useful in program planning, or that yields minimal results, may lead faculty to question the value of the assessment program, or to conclude that little can be done to improve students' skills. Yet the underlying problem may be that the intervention was not broad enough in scope to yield a visible result.

Effecting meaningful change in learning outcomes may require a far greater degree of change in the program than faculty expect. Curricular changes such as reorganizing topics or adding modules or even courses may have limited impact on learning outcomes. For example, ethics instruction in a single course rarely produces significant gains on measures of ethical reasoning (Conry and Nelson, 1989; Ponemon, 1993). Critical thinking is similarly resistant to brief instructional interventions (Kurfiss, 1988; McMillan, 1987). Increasing writing *assignments* without a corresponding increase in writing instruction may have only limited impact on measures of writing skill. Achieving results of practical significance may require both a change of curricular emphasis and a qualitative change in the way students are taught.

Early results from the University of Southern California's Year 2000 Curriculum Project suggest the degree of change necessary to achieve attention-getting results. Early indicators of this program's impact include:

- Increased number of applications for admission to the accounting program
- Increased enrollment in the program
- Reduced drop rates (down to 3% with *tougher* grading standards)
- More diverse students (attracted from other majors, for example, political science students drawn initially by the course's inclusion of government examples)

Faculty have informally noted increases in students' "intellectual aggressiveness," "teamwork and communication skills," and "awareness of business issues" (Pincus, in press). These impressions of important learning outcomes could be verified, for example, by examining student portfolios and by surveying employers of alumni.

It would be impossible to isolate a single cause of these changes. The course has been modified in at least the following ways:

- A focus on the user, not the preparer, and on concepts and tools rather than rules
- An integrated approach to accounting education, introducing basic concepts and issues across all the functional areas of accounting—including systems, tax, auditing, financial and management accounting
- An accent on contemporary examples and current events involving international and domestic business, non-profit and government organizations
- An emphasis on skill development, as well as technical accounting knowledge—including group assignments, written and oral presentation assignments, electronic research assignments, and assignments concerned with ethics and values
- Course and instructor materials that support a change to an interactive learning environment (excerpted from Pincus, in press)

Virtually any feasible design will leave unanswered important questions about "what works" in such a complex program. Still, the USC program advances discussion of curricular and pedagogical innovation by demonstrating that a dramatic departure from normal practice can indeed have an

immediate impact on students' success in and response to the course.

8.4 Design Options for Gradual Evolution of the Assessment Program

The designs chosen in the early stages of assessment should permit gradual evolution of the assessment program. Some suggestions follow.

Descriptive Studies: Descriptive studies can yield immediate program information and can also serve as a baseline for future longitudinal comparisons. Many institutions conduct annual surveys of students and graduates that include self-reported gains on a variety of learning outcomes along with measures of client satisfaction. Including such questions, along with identification of the student by major, is an inexpensive way to obtain student feedback on the program. The results will usually suggest areas for further study.

Course-embedded measures can be combined to obtain a profile of students' accomplishments over time using the "portfolio assessment" method. The professor assigns projects related to targeted objectives and provides prompt feedback to individual students using performance criteria for that objective. Students compile their work into portfolios for subsequent review by a faculty subcommittee. The subcommittee reviews a selected sample of portfolios in a program-level study of students' strengths and weaknesses on the targeted objectives.

In accounting, the portfolio could include one or more case studies, an edited paper, a significant individual research project, and a cooperative group project. Or a single, major case study can be used to assess several skills such as writing, complex problem solving, and ethical reasoning (a high-stakes assessment unless used in concert with other measures). Students' work in two or three courses can be included in the portfolio to allow assessment of skills across a range of content areas.

The credibility of findings from the portfolio approach is enhanced by building "interrater- reliability" among those who will judge the portfolios ([Chapter 9](#)) and using a systematic sampling procedure to select assignments or portfolios for review. Systematic sampling, rather than attempting to review all portfolios, also enhances feasibility.

When using the portfolio review model, faculty may question student ownership of the work submitted. One solution is to include samples of work completed both during and outside of class (Belanoff and Elbow, 1986). An example in accounting would be an in-class essay demonstrating the ability to interpret financial data or spontaneously analyze a complex accounting situation.

Building on Descriptive Foundations: In the descriptive study outlined above, the first set of portfolios reveal students' current capabilities, whether sophomores, juniors, or seniors. Later, the assessment committee can add data for other student groups, and continue to collect data over a period of years. This strategy eventually allows for analysis of developmental change in individual students and trends in program outcomes across cohorts. The initial descriptive study therefore evolves into a relational study. If innovations are introduced, their impact can be assessed using the quasi-experimental time series design (Campbell and Stanley, 1966).

Validation Studies: Another option is to validate an instrument designed to measure students' progress on a high-priority outcome of the program. Performance assessments in particular warrant validation. For example, the assessment committee could recommend a study to determine whether a performance measure used to predict graduates' professional success adds value (such as unique diagnostic insight) compared to more readily accessible measures such as faculty ratings. A different type of validation study would be necessary to determine whether a measure is biased toward or against a particular group of students, for example women or international students. (For additional suggestions see Light and others, 1990). Validation studies can result in a useful contribution to the profession as well as to departmental understanding of its assessment measures.

Immediate Feedback Studies for Monitoring Instructional Innovations: When the faculty implement major changes in curriculum and/or instruction, concurrent inquiry with quick turnaround time can be essential to strengthen the program and prevent major problems from developing.

BYU's "ethnographic" approach satisfied the faculty's need for quick turnaround of data. The faculty

wanted to monitor their new curriculum while it was being implemented so that they could make mid-program adjustments if necessary. They included a variety of qualitative, concurrent inquiry methods in their design:

- Regular sack-lunch discussions with students about the program
- Videotapes and observations of actual class sessions
- Exchange and study of faculty teaching plans
- Descriptions of office hour visits
- Examination of samples of student work
- Frequent meetings and retreats to discuss the program

The BYU faculty supplemented this qualitative approach with a traditional exit examination to assess learning outcomes. Their approach illustrates a mix of concurrent and retrospective, formal and informal assessment strategies. Assessment became an integral part of program planning and improvement.

Relational Studies: Today, relational studies frequently use multivariate analytical techniques to identify the relative weight of factors contributing to a specified learning outcome. Such studies can provide valuable insight regarding the role of student characteristics and features of the educational environment in achievement of a particular outcome. Data for these studies can be stored on a departmental database. Building and gradually enhancing the database gives accounting faculty a flexible, familiar tool for tracking students, monitoring program outcomes, and exploring questions about the program's impact on students.

A recent study illustrates the value of using relational methodology and data from multiple institutions. Data were obtained from three institutions with varying proportions of minority students (primarily African-American). The researchers used regression and analysis of variance to determine the predictive power of high school grade-point average (HSGPA) and students' expected grades for minority and "majority" students (male and female), using withdrawal after the third week and course grades as the dependent variables (Carpenter, Friar, and Lipe, 1993).

The researchers found that minority males were most likely to withdraw. For these students (unlike majority students), withdrawing from the course was unrelated to HSGPA but strongly related to expected grades; actual grades, however, were less strongly related to expectations for minority students when compared to majority students. These findings suggest that efforts to retain African-American students might begin by helping them develop realistic expectations, then provide academic support to increase their chances of success.

Experimental and Quasi-Experimental Studies: Results from descriptive and relational studies often suggest hypotheses about program changes that will improve learning outcomes. The study just described, for instance, might lead faculty to propose a program to address the expectations brought to the institution by students of color. Small-scale pilot studies using experimental methods are useful for testing the effectiveness of such innovations. Other changes that lend themselves to experimental study include new applications of technology or an enhanced writing or speaking component. As noted in Section 8.4, inclusion of student characteristics adds depth to the study and increases the chances of a meaningful result (for example, see Ott and others, 1990).

Research conducted for purposes of educational assessment will rarely satisfy traditional criteria for research quality. Nonetheless, well-planned pilot studies, descriptive and relational studies, and highly focused experimental and quasi-experimental designs can provide timely and relevant information that is significantly more reliable and valid than anecdotal evidence and impressionistic reports.

8.5 Practical Ways to Obtain Data for Assessment

Often the most perplexing challenge in educational assessment is how to obtain a reasonably representative or complete sample of students. The use of course-embedded measures is one important response to this challenge, one of few strategies ideally suited to assessment of learning outcomes. Other potentially useful data collection strategies are suggested below:

- Use electronic mail networks and bulletin boards to surface students' questions,

understanding of the subject, and/or responses to instruction while the course or innovation is in progress.

- Regularly distribute brief, anonymous program questionnaires through the faculty. Use a simple "report card" format. Or pose questions focused on outcomes ("What is the most important concept you have learned in this course so far?" or on the educational environment ("What one thing would you change about the program if you could? What aspect of the program most helps you learn?"). Ask faculty to allocate 5–10 minutes of class time to the questionnaires every 3–4 weeks. Faculty can scan results for their students and make brief reports at a department meeting.
- Set up sack lunch meetings to discuss the program.
- Have teams of students make 20-minute presentations at a series of faculty-student luncheons "by invitation only." Make it an honor to participate. Encourage attention to expanded learning outcomes such as creativity, teamwork, group interaction, and relevance. Videotape the presentations to assemble a panorama of student performances. Let the audience provide brief written feedback.
- Set up focus groups in which students must respond to a current issue in accounting. List concepts and resources used by the group, identify approaches they take to the problem, and note how they interact with each other.
- Include self-reports of progress on key learning outcomes as part of the petition to graduate or an automated registration system.

Faculty, students, instructional resource personnel, and practicing professionals can suggest additional methods.

8.6 Ethical Standards for Research with Human Subjects

The American Psychological Association has established ethical standards for research with human subjects. Issues that are often salient in educational research are the right to privacy, voluntary participation, and the right to expect benefits of participation that outweigh the risks (Joint Committee, 1981, as cited in Davis, 1989). Especially relevant to relational and longitudinal studies such as those described above is the need to obtain permission from students prior to accessing their records.

Research involving students is subject to institutional review procedures for the use of human subjects. Educational studies are often considered exempt from formal review, but should be disclosed to the appropriate institutional review body. Most institutions have a policy on the use of human subjects and may also have a human subjects review committee. Because these policies and procedures are required for Federal funding of grants and contracts, information can usually be obtained from the institutional office that administers externally sponsored projects.