ASSESSMENT VS. RESEARCH

WHY WE SHOULD CARE ABOUT THE DIFFERENCE

Should we call our efforts to learn about programs and outcomes "research" or "assessment"? The name can make all the difference. And trying to adhere to the standards of research may get in the way of doing effective assessment.

BY M. LEE UPCRAFT AND JOHN H. SCHUH

AT FIRST GLANCE, a discussion of the differences between research and assessment may seem about as relevant to most people in higher education as debating how many angels can dance on the head of a pin. Some would assert that this topic is fodder for various educational research journals rather than *About Campus*. But discerning the differences between research and assessment may well determine if the assessment movement will have a long-term, significant impact on policy and practice in higher education or land in the graveyard of other fads. More specifically,

assessment studies must be conducted in ways that are credible with their intended audiences, including faculty and administrators who may well determine whether these studies will influence decision making, policy, and practice. Thus, providing some insight into how assessment studies become credible with these audiences is necessary. To frame this discussion, we have chosen to compare and contrast assessment studies with research studies, and we have endeavored to explain how these forms of inquiry are different even though they employ similar methodology.

Perhaps a personal anecdote will drive home our point. A few months ago, when we were working on the manuscript for *Assessment Practice in Student Affairs*,

For a few days after we received this review, our bruised professional egos went through the gamut of emotions.

we received a very mixed prepublication review. Although many of this anonymous reviewer's suggestions were very helpful in refining the manuscript, much of the criticism focused on our apparent flagrant violation of several canons of conventional educational research. The reviewer suggested that if we wanted to avoid becoming the laughingstock of the higher education research community, we should either (1) abandon the manuscript altogether or (2) start all over and involve another author "who knows how to design and conduct rigorous social science research."

For a few days after we received this review, our bruised professional egos went through the gamut of emotions, from hurt to anger to denial to defensiveness. With that self-therapy behind us, we were determined to do a more thorough analysis of the reviewer's comments. We considered several possibilities. First, we all know people in higher education for whom the very term *social science research* is an oxymoron, who believe that no study could possibly control for all the variables that influence human behavior. Another possibility was that the reviewer was so solidly grounded in quantitative research that he or she found our advocacy of qualitative research, by definition, a major mistake, seeing this methodology as lacking rigor and credibility. Although these two possibilities might certainly explain why some audiences may criticize our work, they did not seem to be the case with this reviewer.

We considered a third possibility: perhaps this reviewer found little or no difference between research and assessment. The reviewer may have been using con-

Lee Upcraft is assistant vice president emeritus for student affairs at Penn State University, where he was also senior research scientist at the Center for the Study of Higher Education until his retirement.

John Schuh is professor and chair of educational leadership and policy studies at Iowa State University. Previously he held administrative and faculty positions at Arizona State University, Indiana University, and Wichita State University.

Note: We would like to receive feedback about the articles we publish. If you send an e-mail to an author, please consider sending a copy to *About Campus* managing editor Paula Stacey at pstacey@jbp.com or posting your letter on our Web site under Letters to the Editor (http://www.josseybass.com/go/aboutcampusforum).

ventional social science research criteria to evaluate assessment studies, and in the process, sinking a knife into the heart of most assessment efforts. Better to have no information at all in making decisions and developing policy and practice than to rely on so-called fatally flawed assessment studies that fail to meet conventional social science research criteria.

We did, in fact, make many revisions based on the reviewer's comments. But we decided that, in some cases, using traditional social science criteria, such as carefully determining the psychometric properties of assessment instruments or conducting stratified random sampling to identify potential subjects to evaluate assessment studies, was inappropriate. What are the differences between these two approaches to gathering and interpreting information? And why is it inappropriate to use some research criteria to evaluate assessment studies?

THE DIFFERENCES BETWEEN RESEARCH AND ASSESSMENT

IN OUR BOOK, *Assessment in Student Affairs*, we defined *research* as "any effort to gather evidence which guides theory by testing hypotheses" and *assessment* as "any effort to gather, analyze, and interpret evidence which describes institutional, departmental, divisional, or agency effectiveness" (p. 21). When comparing research and assessment in his book *Assessing Student Learning and Development*, Dary Erwin argues that although they share many processes, research and assessment differ in two important respects:

- Assessment guides good practice, whereas research guides theory and tests concepts.

- Assessment typically has implications for a single institution, whereas research typically has broader implications for higher education.

The distinctions that Erwin pointed out cannot be overemphasized, because they provide a crucial point of departure between assessment projects and research studies. Assessments use research methods, but they have

Does this mean that, by definition, assessments and evaluations are flawed and therefore not reliable? Definitely not!

very different reasons for being conducted. Assessments are undertaken to guide practice. As a consequence of the assessment's findings, practice is adjusted. Research is framed by theory. As a consequence of a study's findings, the theory may be reconceptualized, affirmed, or perhaps even rejected until another investigation is undertaken.

The debate over these distinctions has included some very strong differences of opinion among social science researchers themselves. Lee Cronbach was one of the first to distinguish between research and evaluation. In *Designing and Evaluating Social Programs*, he wrote

> Designing evaluative investigation is an art. The central purpose of evaluation differs from that of basic social science research, and evaluations fit into different institutional and political contexts. The strategy of evaluative research therefore requires special consideration. Logic is necessarily the same in all disciplined inquiry, but the translation of logic into procedure should depend upon context, purpose, and the expected payoff. Many recommendations appropriate for long-term programs of scientific research are ill suited to evaluation. Hence, general writings on design and scientific method are inadequate to guide the evaluator. For any evaluation many good designs can be proposed, but no perfect ones. (pp. 1–2)

DOING ASSESSMENT IN THE REAL WORLD

DOES THIS MEAN that, by definition, assessments and evaluations are flawed and therefore not reliable? Definitely not! Assessment and evaluation studies are useful and should be done even when they do not adhere strictly to the standards of social science research. Peter Rossi and Howard Freeman make the distinction between "perfect" and "good enough" assessments in their book, *Evaluation:* "In many circumstances, it is difficult or impossible to conduct impact evaluations using what are in ideal terms, the best possible designs." What are the circumstances that can impinge on how we

design assessment studies? Most of us to some degree have to consider the following:

Resource limitations. Most student programs and services lack the human and financial resources to conduct perfect assessments. Overworked and harried educators are asked to participate in assessment studies in addition to their other responsibilities at the worst possible moment, such as budget preparation time or an especially stressful part of the academic calendar such as midterm examinations. Few staff feel qualified to conduct credible assessment studies because they took their last research methods course many years before, and they realize that they are rusty. Further, assessments cost money, but seldom is enough money available to construct the perfect study. Thus, modifications made on the basis of limited resources often affect assessment designs. Although faculty might argue that they, too, suffer from resource limitations, the fact is that inquiry is an important element in their job descriptions and their academic institutions provide them time to develop research studies that fit within the expectations established for their scholarship; in addition, they are able to compete for support that will help facilitate their research projects.

Time limitations. Often, investigators must make decisions, address policies, or solve a problem before they can implement the perfect assessment design The best example in higher education may be retention studies. A well-controlled longitudinal study is ideal, yet few institutions can afford to wait for five or six years of study to address retention problems. Our experience is that the window of opportunity to influence policy and practice may be open for as little as a month and rarely more than a year. Assessment designs must be modified to fit more realistic time expectations. In contrast, researchers can design their studies to provide sufficient time to complete them.

Organizational contexts. Organizations are not static; they are in a constant state of change, and their assessment needs may vary over time. Assessment needs might change drastically as a result of new leadership that requires different evidence of effectiveness, sees new

problems, or devalues old problems. In this changing environment, investigators must make changes in initial assessment agendas, perhaps further modifying the perfect study. Research studies are almost always independent of changing organizational contexts.

Design limitations. Although investigators may conceive of assessment with the best intentions, problems can arise in the implementation phase of the project that they must overcome. In a quantitative assessment design, for example, a perfectly drawn random sample may not yield usable responses from elements of the population that are of particular interest, such as members of historically underrepresented groups; the overall response rate might not be as high as desired, and thus the statistical analysis may be limited or the sampling error increased (or both); or survey instruments may have certain psychometric shortcomings.

In a qualitative study, while individual interviews may be the preferred data collection mechanism, focus groups may be a more efficient way of gathering data. But fewer people than expected may show up to participate in focus groups. Or the interview protocol may not yield the desired information. Or the interviewers may fail to perform effectively. Or something as simple as a malfunctioning tape recorder may limit precise analyses of participant voices. So again, investigators must make compromises. The social science researcher may have the luxury to start over when flaws emerge in the implementation of the project; the assessment investigator is under more pressure to salvage a project and report results acknowledging design implementation flaws.

Political contexts. Social scientists attempt to conduct research that is, to the extent possible, apolitical, and they often have the luxury of conducting studies that search for the truth, no matter where it leads. Assessment, on the other hand, virtually always occurs in a political context that investigators must take into account in designing the assessment. We discussed the politics of assessment in an article in the September/October 2000 issue of *About Campus,* in which we asserted that "all assessment is political" (p. 14) and suggested several strategies for managing these political realities. Assessment designs can and often do reflect local political realities. For example, an assessment study under discussion may never begin; a study in progress may be discontinued; and in rare instances, investigators may keep confidential a study already completed because the results may be politically or ideologically unacceptable to policymakers. In contrast, researchers are likely to be shielded from political influences on their research.

So where does this leave us? When does a study become so modified that it should never be done or discarded even if conducted? Rossi and Freeman, while defending the "good enough" principle of assessment, also argue that the investigator has the responsibility to "raise the question whether to undertake the assessment at all, especially if meaningful results are unlikely." These choices, they say, always involve compromise; no single "always best" design exists. The "good enough rule," they say,

> is that the evaluator should choose the best possible design from a methodological standpoint, having taken into account the potential importance of the program, the practicality and feasibility of each design, and the probability that the design chosen will produce useful and credible results. (pp. 220–221)

The cold reality is that decisions will be made, policies developed, and practices implemented regardless of the availability of assessment results. So the question becomes, When it comes to the usefulness of a study for policy and practice, is a study with substantial limitations better than no study at all? The answer, of course,

The social science researcher may have the luxury to start over when flaws emerge; the assessment investigator is under more pressure to salvage a project and report results acknowledging design implementation flaws.

So the question becomes, When it comes to the usefulness of a study for policy and practice, is a study with substantial limitations better than no study at all?

requires a judgment call, based on whether or not those limitations take away the study's credibility and therefore render it useless. Policymakers and assessment investigators should make such judgments on a case-by-case basis, taking into account specific methodological considerations, as well as organizational and political realities. We should remember, however, that a lack of assessment data can sometimes lead to policies and practices based on intuition, prejudice, preconceived notions, or personal proclivities—none of them desirable bases for making decisions. That is the reality of the administrative and political world of higher education.

So we come down on the side of the "good enough" rule but with one clearly important and major caveat: assessment investigators must clearly identify all modifications made when an assessment study is published, cautioning all prospective audiences to take into account the study's various limitations as they decide what credence to give the study. Failure to take this step is not only unethical, it leaves readers to assume that because the investigators did not identify limitations, they must not know them (or worse yet, they made a conscious decision to leave them out), and therefore both the investigators and the study itself lack credibility.

We must address one other issue in discussing the differences between research and assessment: the role of the researcher differs fundamentally from the role of the assessment investigator. In his 1991 article, "Methods for the Experimenting Society," highly respected social science research methodologist Donald Campbell asserted that "the job of the methodologist for the experimenting society is not to say *what is to be done* but rather to say *what has been done*" (author's emphasis, p. 228). This is in sharp contrast to the role of the assessment investigator, who, in an opinion that we share with several assessment experts such as Catherine Palomba and Trudy Banta, is obligated not only to describe what has been done but what should be done, given the findings of a study.

Assessment studies must be done in ways that are credible with their intended audiences, audiences including faculty and administrators who may well determine if these studies will have an impact on making decisions, policy, and practice. One of the reasons that assessment studies may lack credibility is that some audiences may fail to understand the differences between research and assessment and that these two approaches to inquiry are really two sides of the same coin, equally viable for their intended purposes. Failure to understand these differences can sometimes result in a good enough assessment study being marginalized, trivialized, or ignored because the reader erroneously applies the perfect study syndrome.

To avoid this result, assessment investigators must clearly identify the limitations of their studies (and their potential impact on the findings) and caution audiences to take them into account in interpreting the results. Differences between research and assessment: Who cares? All of us who have a vested interest in using assessment results to influence decision making, policy, and practice should care a great deal about these differences, because they may well determine whether assessment makes a difference.

NOTES

Campbell, D. T. "Methods for the Experimenting Society." *Evaluation Practice,* 1991, *12*(3).

Cronbach, L. J. *Designing and Evaluating Social Programs.* San Francisco: Jossey-Bass, 1982.

Erwin, T. D. *Assessing Student Learning and Development: A Guide to Principles, Goals, and Methods of Determining College Outcomes.* San Francisco: Jossey-Bass, 1991.

Palomba, C. A., and Banta, T. W. *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education.* San Francisco: Jossey-Bass, 1999.

Rossi, P. H., and Freeman, H. E. *Evaluation: A Systematic Approach.* (5th ed.) Thousand Oaks, Calif.: Sage, 1993.

Schuh, J. H., and Upcraft, M. L. "Assessment Politics." *About Campus,* 2000, *5*(4), 14–21.

Schuh, J. H., Upcraft, M. L., and Associates. *Assessment Practice in Student Affairs: An Applications Manual.* San Francisco: Jossey-Bass, 2001.

Upcraft, M. L., and Schuh, J. H. *Student Affairs: A Guide for Practitioners.* San Francisco: Jossey-Bass, 1996.