

Instruments and tools of evaluation

(Originally published in Portuguese as Técnicas e instrumentos de avaliação, as translated by Heraldo Marelin Vianna. Brasília: Universidade de Brasília.)

Eric L. Dey and Joseph M. Fenty
Center for the Study of Higher and Postsecondary Education
University of Michigan

Focus and content

This learning module is intended to provide the learner with a basic introduction to the tools and instruments commonly used in undertaking an evaluation. The focus of this module is two-fold. First, attention will be given to the basic tools of evaluation, including overviews of the various methodological approaches to evaluation and basic principles of research design. The second primary focus of this module will be on identifying important issues related to the selection and construction of instruments, tests, and other tools and techniques used to collect data to be used in evaluation efforts in colleges and universities.

It is important to place the topics of tools and techniques within the larger context provided by evaluation needs. As outlined by Patton (1990), important questions that are useful in guiding any evaluation effort include:

- (1) Who is the information for and who will use the findings?
- (2) What kinds of information are needed?
- (3) How is the information to be used? For what purposes is evaluation being done?
- (4) When is the information needed?
- (5) What resources are available to conduct the evaluation?
- (6) Given answers to the preceding questions, what methods are appropriate?

Answers to these questions will determine the kinds of data that will be most useful in a particular evaluation. The challenge in evaluation is getting the best possible information to the people who need it – and then getting those people to actually use the information in appropriate ways for intended purposes. (Patton, 1990, 12-13)

The goal of this module, using Patton's phrasing, is to provide an introductory overview of tools and techniques which can be used in an evaluation to obtain the best possible information under the constraints at hand.

Finally, since the intent of this module is to provide a general overview of evaluation tools and techniques, it is important to understand that there are numerous interrelated topics which are not covered. For example, analytical techniques such as statistics for quantitative analysis and interpretative approaches to understanding qualitative data are not covered in detail in this module, but are necessary complements to the tools covered herein.

General methodological approaches to evaluation

In planning an evaluation, there are a number of basic decisions which need to be made. Among the most important of these decisions is the general methodological approach to be used in undertaking the evaluation, since the selection of an approach influences many subsequent decisions including the selection of appropriate evaluation tools. There are many factors which can and do influence the selection of a particular methodological approach – evaluation goals, available resources, personal preference and expertise of the evaluator, and the like – and these factors are an important consideration in tailoring the approach to the program being evaluated.

Evaluations are usually approached using methodologies which traditionally have been described as either quantitative or qualitative in nature (Patton, 1990). In this context, a quantitative approach can be described as one in which an evaluator is concerned with judging a program based upon a limited number of measures which have been identified in advance of the evaluation and which are described in numerical terms. Such approaches are typically designed so that the results are collected from relatively large numbers of program participants or stakeholders who are, through random sampling procedures, intended to be statistically representative of others who are not participating in the evaluation. Another emphasis in quantitative approaches to evaluation is on collecting information which can readily be compared to data from similar efforts. For example, a quantitative evaluation might be concerned with determining the participation rate for a particular program in one university, and then comparing this rate to the average rate across all universities. The emphasis is thus on developing generalizable information.

In contrast, a qualitative approach to evaluation is one which emphasizes the collection of rich, detailed information from a small number of program participants and stakeholders. Rather than limiting the scope of the evaluation to measures which have been determined in advance of the evaluation effort, qualitative evaluation approaches typically seek to identify themes which emerge from discussions with the participants and stakeholders. In this way, qualitative approaches strive “to understand the meaning of a program and its outcomes from the participants’ perspectives. The emphasis is on detailed description and on in-depth understanding as it emerges from direct contact and experiences with the program and its participants” (Herman, Morris, & Fitz-Gibbon, 1987, p. 21). For example, a qualitative evaluation might be based upon interviewing a dozen or so

program participants to find out what the program being evaluated has meant to them, and identifying ways in which it might be improved. The emphasis here, then, is upon developing information for understanding particular cases rather than collecting generalizable, comparable data.

It is convenient to think of quantitative and qualitative approaches to evaluation as being completely distinct, but this is in fact a false dichotomy. These approaches do not have to be mutually exclusive, and the reality is that most evaluations can benefit from incorporating both perspectives in the evaluation plan (Patton, 1990). For example, Table 1 of Module VI in this series (Perspectives and models by Joan S. Stark) arranges selected evaluation models on a continuum which runs from the most qualitative to the most quantitative. Thus, while a naturalistic approach would favor qualitative methods, an evaluation might also benefit from some quantitative evidence concerning the program being evaluated. Likewise, an evaluator using an Input-Environment-Output model to identify the statistical linkages between various parts of the program being evaluated might benefit by drawing upon interviews from program participants to shed light on the numerical results.

For the purpose of this module, the most important aspect of understanding the differences between these general methodological approaches is that they influence the degree to which a particular tool or instrument is appropriate. In the case of qualitative evaluation, for example, the evaluator serves as the data collection instrument. By interacting with participants and other program stakeholders through interviews or focus groups, the qualitative evaluator works to generate raw data to be used in subsequent analyses. In contrast, the quantitative evaluator tends to use very structured techniques and instruments for data collection. Although quantitative evaluators can also collect data through

interviews, these interviews tend to be very different than the conversations generated through a qualitative interview. Quantitative interviews are typically designed to elicit from the respondent a specific set of pre-specified responses, which can then be assigned quantitative codes. Structured survey forms – administered through the mail or distributed in educational settings – can be used to accomplish many of the same goals.

Research design issues

Before moving on to address the different techniques, it is important to briefly address the subject of research design. Research design is an especially important consideration for quantitative approaches to evaluation for it directly influences the nature and the quality of the conclusions which can be drawn and the comparisons which can be made. Design issues are also important in qualitative evaluation, especially as is related to the selection of a particular qualitative approach, and are reviewed by Patton (1990).

In order to illustrate the main research design issues that need to be considered when planning a quantitative evaluation, an example will be developed using the model outlined by Astin (1991; see also Campbell & Stanley, 1966)). As noted in Module VI, Astin's Input-Environment-Outcome (I-E-O) model tends to be among the most quantitative approaches to evaluation. Although it does not necessarily have to be implemented using quantitative evaluation tools, in practice it typically is.

The I-E-O approach is deceptively simple. The essential idea is that all educational programs can be thought of as having three basic components – inputs, environments, and outcomes – and that data needs to be collected from each of

these three components in order to fully evaluate an educational program and its effectiveness. If we apply the I-E-O model to evaluations directed at educational programs we can define these components as follows:

Outcome refers to the characteristics of the student that the educational program either does influence or attempts to influence as measured after exposure to the educational program.

Input refers to the characteristics of the student at the time of initial entry into the educational program.

Environment refers to the various aspects of the educational program that are designed to influence student outcomes, including the structure of institution which operates the educational program, its policies and practices, and the nature of the teachers and students who are associated with the program.

Implicit in these definitions is that the collection of data on student inputs and outcomes needs to occur at different points in time. This is known as a longitudinal design, in which students (or, more generally, program participants) are queried at the beginning of the educational program and at least one time more after they have had exposure to the program, typically after completing the program. Change can then be inferred by comparing the responses of each of the students at the two (or more) time points. Since student outcomes are usually related to the characteristics that they bring to the program (e.g., their input characteristics), it is important to control for the effects of inputs before attempting to identify the effects of the program environment.

The I-E-O model thus focuses on the possible effects of environments on outcomes. Understanding environmental effects is, of course, a common goal of most evaluation efforts, but environmental effects are also important from an educational perspective since the environment includes those aspects of the

student's experience that the program can directly change. This design, then, is designed to identify better ways to structure programs and educational environments in order to maximize their effectiveness.

As an example, let's assume we are interested in evaluating a new curriculum designed to increase the amount of knowledge that students have about organic chemistry. One simple approach would be to identify or create a test which would serve as a measure of student knowledge in this area, and then have students complete the test at the end of two chemistry courses – one which used a traditional chemistry curriculum and one which used a new curriculum. By comparing the test results from both courses we could determine whether or not the students in the course using the new curriculum had higher average scores. If this were the case, this would appear to be evidence that the new curriculum was in fact more effective at stimulating student knowledge of organic chemistry, but there are other possibilities we need to consider as well. It may be that students in the course using the new curriculum knew more about organic chemistry before the course started than those in the regular course, and that this along may explain the post-course score differences. Without information about student inputs, we cannot rule out this possibility. It may also be that the instructor using the new curriculum was a more effective instructor than the one teaching the regular course; reassigning the instructors might have yielded the opposite results. Without information about environmental factors, we cannot rule out this possibility. This, then, is an example of what Astin calls an outcomes only approach to evaluation, where in the absence of knowledge about student inputs (e.g., knowledge before the course) or environmental factors (e.g., instructor effectiveness) the results might be misleading.

It should be noted that outcomes only approaches can yield valid results under certain conditions. For example, if the evaluation encompassed a larger number courses, students and instructors might be randomly assigned to either the traditional or new chemistry course, thereby reducing the possibility that either the students or instructors associated with either approach would be different in important ways. This sort of approach – randomization of participants – is close to classical experimental designs used in many fields, but has been difficult to accomplish in most educational settings due to the ways in which schools and colleges are structured and operated. Without randomization of participants or knowledge about student inputs and educational environments, outcome-only approaches yield relatively weak information about program effectiveness.

Slightly stronger designs combine outcome information with either – but not both – input or environment components of the I-E-O framework. For example, a input-outcome design could measure knowledge about organic chemistry at two time points to establish the amount of change, but without additional information on environmental factors it is impossible to perfectly attribute change in knowledge to the curriculum (as opposed to other environmental factors). Environment-outcome designs provide additional details about the environmental factors which are related to outcomes, but in the absence of input characteristics we cannot directly infer that any outcome differences are attributable to these environments since they may simply be the result of differences among students upon entry into the course.

By employing an evaluation research design encompassing all the components of the I-E-O framework, the evaluator able to reduce the prospects of making these particular kinds of errors. Other approaches can do this as well (including the pure experimental design mentioned above), but it should be recognized that in many circumstances the ideal may not be feasible. Nevertheless,

it is important in any evaluation to attempt to collect information which will be useful in helping eliminate alternate explanations for any observed differences in individual outcomes.

Tools for evaluation

In identifying appropriate tools for conducting and evaluation, care should be taken to adhere where ever possible to established standards for evaluation projects (The Joint Committee, 1994). From the general standards of utility, feasibility, propriety, and accuracy follow a number of specific standards which relate most directly to the tools of evaluation used for the purpose of collecting data:

Defensible Information: The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

Valid Information: The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

Reliable Information: The information gathering procedures should be chosen and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

Systematic Information: The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

These standards are important for they allow those outside of the evaluation to develop confidence in the information used in the evaluation. An important part of the defensibility standard is providing sufficient detail about the information sources so that a stakeholder using the evaluative results can consider whether the most important information sources that were practicably available were used in the

evaluation effort. In turn, the validity standard ensures that the information used in the evaluation is sufficient to develop trustworthy conclusions, while the reliability standard guarantees that this information is both consistent and free from influences associated with other factors in the evaluation. Finally, the systematic information standard reminds the evaluator of the need to review and verify the quality of the information being used for the purpose of evaluation.

Selecting appropriate measurement instruments

In addition to the general methodological approach being used for the evaluation and established evaluation standards, a primary consideration in identifying what might be an appropriate tool or technique to use in an evaluation can be determined by carefully examining the nature and structure of the program and the goals that the program has for those who are participating in it. Although a more complex taxonomy might be developed, educational programs are usually directed at influencing one or more of the following: knowledge, academic performance, behaviors, attitudes, and beliefs. Information from each of these domains can be collected using a variety of tools, but some tools are more commonly used for certain purposes than others. To return to the previous example, knowledge about organic chemistry could be assessed by including a number of factual questions about chemistry on a survey or through an in-depth interview of a student conducted by a knowledgeable chemist, but this sort of knowledge is most commonly assessed using a test or examination (which might have been developed specifically for the evaluation or which might be available from a testing organization). There is no “one best approach” to the selection of such tools, as the suitability of a approach is dependent upon each particular combination of circumstances.

Tools commonly used in quantitative approaches to evaluation

Examinations for evaluating knowledge

In educational settings, the need to evaluate academic performance and knowledge-based outcomes is an important one and is a task which is usually accomplished using some form of a test or examination. An early decision related to the selection of one kind of test over another is whether to use a standardized test available from testing organizations or to develop one locally. Standardized tests often have advantages which test vendors are quick to enumerate. Advantages most often cited include established reliability and validity; available comparative and normative data based on national samples; and the relative ease of administration and scoring.

On the other hand, critics have highlighted several disadvantages which evaluators should be aware of to make a wise decision about whether a standard test will meet their evaluation needs. For example, Jacobi, Astin, and Ayala (1987) argue that the decision to use standardized instruments may actually stymie important discussions about educational goals and objectives that are part of most evaluations. Moreover, standardized instruments tend to focus on minimum standards rather than recognizing standards of the local institution. The method the testing organization uses to report test results may also limit the usefulness of standard instruments; the failure to supply absolute scores for each test item or topical area may make it difficult to advise students about academic weaknesses which deserve attention. These disadvantages have been cited by critics of standardized tests as justification for developing local instruments to assess cognitive outcomes.

In contrast, locally-developed tests can have the advantage of directly responding to institutional goals and priorities. The process of identifying academic objectives and goals as part of an evaluation process by requiring the participation of faculty, administrators, and staff is a necessary part of the development of local instrumentation, and may well represent the greatest benefit of this approach. The collaborative effort may spur useful discussions about the program and its goals.

Critics of locally-developed tests have identified several disadvantages. Designing a locally designed instrument to assess student outcomes can be both expensive and time consuming. These instruments can take several years to develop and may initially lack adequate reliability and validity. Furthermore, because locally designed test are typically unique and new, cross-sectional and longitudinal comparisons are often unavailable.

The following is a brief discussion of standardized examinations that are commonly used in the United States. Several indices and critiques of published tests available in the United States have been published, and include works by Sweetland and Keyser (1991), Keyser and Sweetland (1987), and Murphy, Conoley, and Impara (1987).

Tests designed for lower-division college students:

The American College Test (ACT) and the Scholastic Aptitude Test (SAT) are general tests of academic performance and aptitude, the results of which are typically used to determine if a student will be admitted to a particular college.

College-Level Examination Program (CLEP) is a general exam designed to measure materials typically covered in the first two years of

college. The test covers five areas: English composition, humanities, mathematics, natural sciences, and the social sciences and history.

Tests designed for upper-division college students:

The Graduate Record Examination (GRE) results are generally used to determine admission to graduate school. This examine provides scores for verbal, quantitative, and analytical reasoning abilities.

The Miller Analogies Test (MAT) is designed to measure verbal reasoning skills. The results are often used to assist college decide whether a student will be admitted to postgraduate study, especially in the field of education.

The Educational Testing Service (ETS) Academic Profile is a test designed to measure reading, writing, and math ability; and critical thinking in the context of three major discipline groups: humanities, social sciences, and natural sciences.

The Graduate Management Admissions Test (GMAT), Medical College Admission Test (MCAT), and the Law School Admission Test (LSAT) are designed to predict success in the first year of graduate schools in management, medicine, and law, respectively.

Tests designed for all levels of college students:

The American College Testing College Outcomes Measures Project (COMP) is designed to measure the general abilities and competencies typically associated with general and liberal education. Three content areas (functioning within social institutions, using science and technology, and using the arts) and three process areas (communicating, solving problems, and clarifying values) are assessed.

Watson-Glaser Critical Thinking Appraisal - this test is used to assess students' critical thinking abilities. Five sub-tests (inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments) to measure abilities.

Cornell Critical Thinking Test - this test is designed to measure critical thinking abilities for college students and older adults.

Surveying attitudes, values, and behaviors

In addition to being directed at influencing the knowledge of participants, many educational programs are also expected to bring about change in participants' attitudes, values, and behaviors. Moreover, student attitudes, values, and behaviors can have an important effect on knowledge-based outcomes as well. If the evaluator is to accurately evaluate the impact of a program on participants it is vital that this information be collected and analyzed.

A number of methods of data collection are available to the evaluator to collect information in this area, with a survey approach being quite popular. A

survey can serve as an efficient and effective means of identifying the impact of educational programs and institutional environments on students, faculty, staff, and employers. Surveying these various groups can also enhance the credibility of the evaluation by demonstrating that important voices have been heard and considered.

The basic purpose of a survey of the type being discussed here is to collect information that will be used to describe a population, or group, which meets certain criteria defined by the evaluation (e.g., students participating in the program being evaluated). Data collection is usually done by selecting and contacting a sample of individuals who are part of the larger population, and then using their responses to infer things about the larger population. In cases where the population of interest is small, surveys often seek to gather information from the entire population. A population survey can, however, be used with any size population, but the sampling approach is commonly used to reduce evaluation costs.

Once the population and sample have been identified, there are a number of forms that a survey can take. Information can be gathered through questionnaires distributed in classrooms or through the mail. Information can also be gathered using face-to-face interviews or over a telephone. Regardless of form, evaluators should keep in mind that a survey is an attempt to interview a number of individuals, and should be structured in the fashion of an interview even if said interviews are not conducted in person.

The next step in using a survey as part of an evaluation is to identify and/or develop the questions which will be presented to respondents. The choices here parallel those discussed above with respect to using a standardized or locally-developed tests. In the United States, for example, a number of published questionnaires are available which can be used in general evaluations of

undergraduate education. These include student and faculty surveys published by the Cooperative Institutional Research Program operated by the UCLA Higher Education Research Institute and the College Student Experiences Questionnaire developed by C. Robert Pace and published by Indiana University.

In addition to complete instruments, there are numerous scales and indexes available which can be incorporated into locally-developed questionnaires. For example, if a program was designed to improve an individual's self-esteem, it would be possible to identify a number of established ways of measuring self-esteem and incorporate one of them rather than simply creating a new measurement method (Miller, 1991). Of course, as with standardized tests and examinations, published questionnaires which are suitable for particular purposes may not always be available, but a number of questionnaires should be reviewed as they are a good source for generating ideas for developing more appropriate questionnaires.

The development of survey forms is mostly considered an art form, but there is an abundance of technical and practical literature on the topic as well. General suggestions include the following:

The questionnaire should be simple, intelligible, and clearly written.

Good questions are those that use simple language, common concepts, and manageable tasks.

Specific questions are better than general ones. The more general the question, the wider the range of interpretations it may be given. In contrast, wording that is specific and concrete is more apt to communicate uniform meaning.

Finally, in constructing or using questionnaires, evaluators should limit the use of open-ended questions (i.e., those without pre-specified answers from which the respondent is asked to choose) for they are hard to use within a quantitative evaluation framework. In contrast, open-ended questions are important elements of

qualitative approaches to evaluation and should therefore be encouraged if that is the general methodological approach being used.

Tools commonly used in qualitative approaches to evaluation

Qualitative interviewing and focus groups

In contrast to the approach to interviewing adopted by an evaluator using a quantitative approach, the interviewing strategy of the qualitative evaluator is designed to encourage directed, but free-flowing, conversations. Interviews of this nature can be conducted with individuals or in a group setting, with the latter often referred to as a “focus group” or “focused group interview.” Regardless of format the intent is the same: To generate information which can be used to understand the program being evaluated from the perspective of the participant. In doing so, the qualitative evaluator uses the language of the participant (rather than responses to pre-defined categories offered via a questionnaire) as the raw data upon which the analysis will be conducted.

In developing protocols for these kinds of activities, it is important to keep a number of considerations in mind. Even though one intent of this technique is to generate a free-flowing conversation with the respondent, the conversation needs to be focused so as to generate information that will be useful for the evaluation. Given this, it is important for the evaluator to structure the interview so that it is limited to a relatively small number of issues. At the same time, the goal of these activities is to seek information that is deeper, and more profound than that which is accessible through other means. This implies the presence of a skilled-interviewer or moderator who can work to elicit information of use to the evaluation.

Focused group interviews use techniques to those used in one-on-one interviews, but are conducted using groups of individuals who are interested in – or have had some involvement with – the topic being discussed. It is, however, important to recognize that there is more of a difference between individual and group interviewing than the number of participants. As noted by Brown (quoted by Patton, 1990, p. 17):

Groups are not just a convenient way to accumulate the individual knowledge of their members. They give rise synergistically to insights and solutions that would not come about without them.

This realization again underscores the need for skilled interviewers and moderators to promote this process. As with the construction of questions for surveys, this can be seen as an art form which can be improved by following suggestions based on empirical research.

If done well, the data produced by these techniques will be very rich, and give those reading the evaluation a sense of the experiences of the participants. Moreover, by being unstructured, those being interviewed can readily qualify their responses and identify for the interviewer important contingencies which influenced the nature of their experiences.

Observing behavior

Observations of participants is another common implementation of qualitative approaches to evaluation. This approach stems from the realization that “there are limitations...to how much can be learned from what people say. To understand fully the complexities of many situations, direct participation in and

observation of the phenomenon of interest may be the best research method” (Patton, 1990, p. 25). The purpose of this technique is to generate detailed and in-depth data which can be used to give the reader of the evaluation report a “sense of place” which complements and extends that which can be generated solely from qualitative interviewing.

A number of observational approaches can be used, with participant observation perhaps being the most common. In participant observation, the evaluator visits the site of interest and, by participating in the unfolding events, makes detailed notes about the setting and the participants within it. These notes become, in the words of Patton (1990, p. 26) “the eyes, ears, and perceptual senses for the reader. The descriptions must be factual, accurate, and thorough without being cluttered by irrelevant minutiae and trivia. The basic criterion to apply to a recorded observation is the extent to which the observation permits the reader to enter the situation under study.” Other forms of observational data collection approaches share similar goals, but vary in terms of the role adopted by the evaluator: Rather than being a participant in the unfolding events, the evaluator adopts the role of an onlooker.

Portfolio approaches to evaluating academic performance

Used as part of efforts directed at evaluating student learning, the use of student portfolios is becoming increasingly popular in postsecondary institutions in the United States. Portfolios are typically a compilation of various artifacts which are collected throughout the student’s university experience along with several self-reflective essays written especially for the portfolio. These artifacts are then used to demonstrate specific skills, competencies, and values which are consistent with

university and program goals and objectives. As described here, portfolios are especially useful in evaluating the broad impact of educational programs, although similar concepts could be adapted to other evaluative purposes.

An important objective of using an evaluation tool such as portfolios is to help students improve their ability to self-assess their own work. Students are often asked to introduce each item in the portfolio to explicate the nature of the artifact and the type of development the assignment facilitated. Several institutions also require that students include a reflective essay which allows the her or him to discuss how the university experience has changed their lives.

Whether intentional or not, the use of portfolio evaluation can also serve as a learning opportunity for faculty members. In essence, the evaluation effort becomes directly linked with programmatic improvement efforts. The student portfolio is often evaluated by the instructor or a committee to provide important feedback to the students. Portfolios are evaluated in terms of the academic quality, construction, and efficacy of the career-related materials. Faculty who review portfolios from the perspective provided by another discipline often come away with an enhanced appreciation of the student's academic experience as well as ideas for cross-disciplinary collaboration.

A recent book by Banta (1996) provides a number of institutional case studies which demonstrate the variety of ways in which colleges and universities in the United States have used the portfolio approach in evaluating and assessing student academic performance:

King's College, Wilkes-Barre, Pennsylvania. King's College uses a portfolio of writings developed in a required course to determine the writing and thinking ability of English majors. Students in an advanced writing class are asked to select three of six papers for further revision and editing. In addition to these papers,

student are required to include a reflective essay and a 'meta-text' which articulates the student's approach to writing and the writing process. At the end of the semester the students present their portfolios to the class. In addition, the faculty member who teaches the class evaluates the portfolio, the evaluation is shared with the student responsible for the portfolio and other professors.

Many positive results have been reported. Students learn to revise their writing through an organized approach. They have learned to take ownership of their learning by becoming actively involved in the development of their personal portfolio. Faculty have indicated that the portfolio has improved their ability to evaluate students.

Samford University, Birmingham, Alabama. At Samford University, faculty in the School of Education require students to maintain a "professional portfolio" to: (1) give students a vehicle by which they can market their skills; and (2) to provide a process by which students can reflect upon and evaluate their own learning. The actual portfolios are created during the student's final year in college. The portfolio contains artifacts which demonstrate a well-developed teaching philosophy and competent curricular decision-making. The portfolio also contains a reflective journal to record the student's thoughts, feelings, and self-evaluations of growth during the student's teaching experience. Using an established set of criteria, the portfolio is evaluated by a professor and an external reviewer. The student uses the portfolio to talk about their experiences and accomplishments. The portfolio approach has improved the ability of students to communicate. This approach has improved the ability of students to articulate the dimensions of their teaching experience to faculty, peers, and potential employers.

Rappahannock Community College, Glens, Virginia. At Rappahannock Community College the portfolio has been used to provide useful feedback to

students. Students are required to include in their portfolio a resume, completed graduation application, and cover letter. These documents and others are reviewed by the Citizen's Advisory Committee, a committee designed to provide important feedback to students from professionals working in the field.

The portfolio approach has created an opportunity for the Citizen's Advisory Committee provide to comments to students about the quality and efficacy of their educational projects and resume. This approach has strengthened the ties between the business community and in so doing has help keep the course and program offerings relevant.

Northeast Missouri State University, Kirksville, Missouri. Northeast Missouri State University uses the portfolio to evaluate learning objectives such critical thinking, interdisciplinary synthesis, scientific and mathematical reasoning, aesthetic appreciation, and co-curricular learning. Although the portfolio is not collected until the senior year, students are encouraged from the first academic term on to save their work. Faculty may use the portfolio during the sophomore year to review the student's writing ability. During the senior year, the University Portfolio Committee (UPC) articulates and reports the items that students must include in the portfolio. Faculty who teach the senior seminar work with the students to ensure that the portfolio specifications are met. Since the ability to self-evaluate one's work is an important portfolio objective, each item included in the portfolio must be evaluated by the student. In addition, students are required to include a cover letter which describes their overall academic experience.

Students are instructed to make a copy of the portfolio, at university expense, so its contents can be reviewed by the faculty. Just after spring graduation, a select group of faculty are paid to review the portfolios. These faculty are drawn from many disciplines, thus they often face the challenge of evaluating the student's

academic work in another field. Faculty report that this experience has given them a better appreciation for the university experience, from the prospective of the student. In addition, faculty often come away with a better understanding of where improvements may be made in the curriculum.

The use of portfolios at Northeast Missouri State University is seen as an opportunity to engage in active learning for both students and faculty. Students learn to reflect upon and to self-evaluate their academic experience. Faculty, on the other hand, have the opportunity to holistically examine the curricular experience and to provide important and relevant feedback to their students. The portfolio evaluation experience has helped the faculty understand the effects of the curriculum and adjust course offerings.

Unobtrusive methods

Although varying widely in terms of approach and technique, all of the evaluation tools described above share in common the following characteristic: They require the evaluator to interact at some level with program participants in order to collect the data being sought. This is an important consideration since the presence of an evaluator can inadvertently alter the data being collected. Given this, a final group of tools and techniques needs to be mentioned, and which is not closely aligned with either quantitative or qualitative approaches. These have been described as “unobtrusive methods” by Webb, Campbell, Schwartz, and Sechrest (1981), and are based on data pulled from physical traces, archives, and observations.

In introducing the idea of nonreactive research, Webb, Campbell, Schwartz, and Sechrest (1981, p. 2) provide the following illustrative examples:

The floor tiles around the hatching-chick exhibit at Chicago’s Museum of Science and Industry must be replaced every six weeks. Tiles in other parts of the museum need not be replaced for years. The selective

erosion of tiles, indexed by the replacement rate, is a measure of the relative popularity of exhibits.

Library withdrawals were used to demonstrate the effect of the introduction of television into a community. Fiction titles dropped, nonfiction titles were unaffected.

Attitudes toward race in two colleges were compared by noting the degree to which students of different races clustered in lecture halls.

Unobtrusive methods of collecting evaluation data can probably not be the only tool used for collecting evaluation data simply because these kinds of data tend to be neither comprehensive nor definitive. Nevertheless, such approaches – limited only by the evaluators creativity – can be very useful supplements to data collected through more common means.

Concluding comment

The instruments and tools of evaluation are constantly in the process of being revised and improved through practice. The intent of this learning module has been to provide a basic introduction to common tools and techniques used in evaluation, and to highlight concerns and considerations about their use. As noted throughout the module, the applicability of any tool for a particular purpose is dependent upon a great number of factors, and these factors need to be carefully considered at the outset of any evaluation. However, it should also be remembered that considered creativity in developing and using evaluation tools is the basic process through which professional skills – and evaluation tools – improve.

References cited

- Astin, A.W. (1991). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. New York: Macmillan.
- Banta, T.W. (1996). Assessment in practice: Putting principles to work on college campuses. San Francisco: Jossey-Bass.
- Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand-McNally.
- Cook, T.D., & Campbell, D.T. (1979). Quasi-Experimentation. Chicago: Rand-McNally.
- Herman, J.L., Morris, L.L., & Fitz-Gibbon, C.T. (1987). Evaluator's handbook. Newbury Park, CA: Sage Publications.
- Jacobi, M.A., Astin, A.W., & Ayala, F. (1987). College student outcomes assessment: A talent development perspective. ASHE-ERIC Higher Education Report No. 7. Washington, D.C.: Association for the Study of Higher Education.
- Keyser, D.J., & Sweetland, R.C. (Eds.) (1987). Test critiques compendium: Reviews of major tests from the Test critiques series. Kansas City, MO: Test Corporation of America.
- Murphy, L.L., Conoley, J.C., & Impara, J.C. (1987). Tests in print IV: An index to tests, test reviews, and the literature on specific tests. Lincoln, NE: Buros Institute on Mental Measurements, University of Nebraska Press.
- Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd edition). Newbury Park, CA: Sage Publications.
- Sweetland, R.C. & Keyser, D.J. (Eds.) (1991). Tests: A comprehensive reference for assessments in psychology, education, and business. Austin, TX: Pro-ed.

The Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards (2nd edition). Newbury Park, CA: Sage Publications.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1981). Unobtrusive measures: Nonreactive research in social sciences. Boston: Houghton Mifflin.

Readings and resources

General methodological approaches and research design

- Astin, A.W. (1991). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. New York: Macmillan.
- Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand-McNally.
- Cook, T.D., & Campbell, D.T. (1979). Quasi-Experimentation. Chicago: Rand-McNally.
- Creswell, J.W. (1994). Research design: Qualitative and quantitative approaches. Newbury Park, CA: Sage Publications.
- Herman, J.L., Morris, L.L., & Fitz-Gibbon, C.T. (1987). Evaluator's handbook. Newbury Park, CA: Sage Publications.
- Light, R.J., Singer, J.D., Willett, J.B. (1990). By design: Planning research on higher education. Cambridge, MA: Harvard University Press.
- Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd edition). Newbury Park, CA: Sage Publications.
- Rossi, P.H., & Freeman, H.E. (1985). Evaluation: A systematic approach. Newbury Park, CA: Sage Publications.
- Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage Publications.
- The Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards (2nd edition). Newbury Park, CA: Sage Publications.

Testing and examinations

- Adelman, C. (Ed.) (1988). Performance and judgment: Essays on principles and practice in the assessment of student learning. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Banta, T.W. (1996). Assessment in practice: Putting principles to work on college campuses. San Francisco: Jossey-Bass.
- Jacobi, M.A., Astin, A.W., & Ayala, F. (1987). College student outcomes assessment: A talent development perspective. ASHE-ERIC Higher Education Report No. 7. Washington, D.C.: Association for the Study of Higher Education.
- Keyser, D.J., & Sweetland, R.C. (Eds.) (1987). Test critiques compendium: Reviews of major tests from the Test critiques series. Kansas City, MO: Test Corporation of America.
- Morris, L.L., Henerson, M.E., & Lindheim, E. (1987). How to measure performance and use tests. (Program Evaluation Kit #7). Newbury Park, CA: Sage Publications.
- Murphy, L.L., Conoley, J.C., & Impara, J.C. (1987). Tests in print IV: An index to tests, test reviews, and the literature on specific tests. Lincoln, NE: Buros Institute on Mental Measurements, University of Nebraska Press.
- Sweetland, R.C. & Keyser, D.J. (Eds.) (1991). Tests: A comprehensive reference for assessments in psychology, education, and business. Austin, TX: Pro-ed.

Developing survey instruments

- Converse, J.M., & Presser, S. (1986). Survey questions: Handcrafting the standardized questionnaire. Newbury Park, CA: Sage Publications.
- Fowler, Jr., F.J. (1995). Improving survey questions: Design and evaluation. Newbury Park, CA: Sage Publications.

Henerson, M.E., Morris, L.L., & Fitz,-Gibbon, C.T. (1987). How to measure attitudes.

(Program Evaluation Kit #6). Newbury Park, CA: Sage Publications.

Miller, D.C. (1991). Handbook of research design and social measurement (5th

edition). Newbury Park, CA: Sage Publications.

Rossi, P.H., Wright, J.D., & Anderson, A.B. (1983). Handbook of survey research.

Orlando, FL: Academic Press.

Salant, P., & Dillman, D.A. (1994). How to conduct your own survey. New York:

John Wiley & Sons.

SPSS. (1995). SPSS survey tips. Chicago: SPSS.

Qualitative interviewing and focus groups

Krueger, R.A. (1994). Focus groups: A practical guide for applied research (2nd

edition). Newbury Park, CA: Sage Publications.

Morgan, D.L. (1988). Focus groups as qualitative research. Newbury Park, CA: Sage

Publications.

Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd edition).

Newbury Park, CA: Sage Publications.

Rubin, H.J., & Rubin, I.S. (1995). Qualitative interviewing: The art of hearing data.

Newbury Park, CA: Sage Publications.

Stewart, D.W., & Shamdasani, P.N. (1990). Focus group: Theory and practice.

Newbury Park, CA: Sage Publications.

Unobtrusive measurement

Sechrest, L. (1979). Unobtrusive measurement today. San Francisco: Jossey-Bass.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1981). Unobtrusive measures: Nonreactive research in social sciences. Boston: Houghton Mifflin.

Learning activities

The activities that follow are important supplements to the materials contained in the module, and should be done after studying the module and the recommended readings.

Activity 1 General methodological approaches and evaluation standards

Identify an educational program at your institution that has recently been evaluated, and carefully review the final report of the evaluation.

Based on the final report of the evaluation, what was the general methodological approach which was used in conducting the evaluation? Did the evaluators adopt what was primarily a quantitative or qualitative perspective, and to what extent did they report considering alternative approaches?

Using the four evaluation standards noted above as relating most closely to the tools and techniques of evaluation (defensible information, valid information, reliable information, systematic information), create a table which evaluates the evaluation. To what extent were these standards met, and what, if any, changes would you recommend for subsequent evaluations of the program?

Recommended readings:

- Creswell, J.W. (1994). Research design: Qualitative and quantitative approaches. Newbury Park, CA: Sage Publications.
- Herman, J.L., Morris, L.L., & Fitz-Gibbon, C.T. (1987). Evaluator's handbook. Newbury Park, CA: Sage Publications.
- Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd edition). Newbury Park, CA: Sage Publications.
- The Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards (2nd edition). Newbury Park, CA: Sage Publications.

Activity 2
Alternative approaches to evaluation

Working with a colleague, identify an educational program with which you are both familiar. Working independently, develop a 5-10 page evaluation proposal for the educational program; one proposal should adopt a qualitative approach while the other should adopt a quantitative approach.

Exchange proposals, and critique them using the four general evaluation standards of utility, feasibility, propriety, and accuracy.

Using the critiques as a starting point, develop a third, joint proposal which combines elements of both methodological approaches. Thinking now of the program's various stakeholders, which of the three proposals would each of the program's constituencies find the most useful?

Recommended readings:

Creswell, J.W. (1994). Research design: Qualitative and quantitative approaches. Newbury Park, CA: Sage Publications.

Cook, T.D., & Campbell, D.T. (1979). Quasi-Experimentation. Chicago: Rand-McNally.

Light, R.J., Singer, J.D., Willett, J.B. (1990). By design: Planning research on higher education. Cambridge, MA: Harvard University Press.

Activity 3 Measuring performance: Tests and portfolios

Identify an academic discipline or program area in your institution with which you are familiar, and obtain specimen copies of standardized tests which are available in that particular area. Working with a colleague familiar with the selected discipline, review the examinations and reporting options available for each published test.

To what extent might the standardized examinations be useful in evaluating learning in the selected discipline? If there are no examinations available in the area (or if what does exist is not particularly useful for evaluation purposes), what considerations should be addressed in developing an examination locally?

To what extent might a portfolio approach to assessing student learning and development be useful? What trade-offs might be necessary in deciding to use a testing or a portfolio approach?

In considering the likely reactions of students and other stakeholders, would a testing or a portfolio approach be more acceptable to one group or another?

Recommended readings:

- Banta, T.W. (1996). Assessment in practice: Putting principles to work on college campuses. San Francisco: Jossey-Bass.
- Jacobi, M.A., Astin, A.W., & Ayala, F. (1987). College student outcomes assessment: A talent development perspective. ASHE-ERIC Higher Education Report No. 7. Washington, D.C.: Association for the Study of Higher Education.
- Morris, L.L., Henerson, M.E., & Lindheim, E. (1987). How to measure performance and use tests. (Program Evaluation Kit #7). Newbury Park, CA: Sage Publications.

Activity 4 **Measuring attitudes, values, and behaviors: Tests and portfolios**

Identify a recently completed evaluation report which used (at least in part) a survey methodology to collect information, and carefully review the tools and techniques used for the evaluation.

In terms of the kinds of information sought (e.g., attitudes, behaviors, level of satisfaction), what was the primary goal of the questionnaire? In reviewing the questionnaire, what questions seem particularly well-formed and which need additional clarification? Suggest ways of improving the questionnaire using principles found in the readings. Based on the kinds of information sought by the questionnaire, develop an interview protocol that might be used in a qualitatively-focused interview. To what extent do the questions need to be changed? If the interview protocol developed above was to be used in a focus group, what changes would need to be made?

Recommended readings:

- Fowler, Jr., F.J. (1995). Improving survey questions: Design and evaluation. Newbury Park, CA: Sage Publications.
- Henerson, M.E., Morris, L.L., & Fitz,-Gibbon, C.T. (1987). How to measure attitudes. (Program Evaluation Kit #6). Newbury Park, CA: Sage Publications.
- Morgan, D.L. (1988). Focus groups as qualitative research. Newbury Park, CA: Sage Publications.
- Rubin, H.J., & Rubin, I.S. (1995). Qualitative interviewing: The art of hearing data. Newbury Park, CA: Sage Publications.
- Salant, P., & Dillman, D.A. (1994). How to conduct your own survey. New York: John Wiley & Sons.

Activity 5 Unobtrusive measures

Identify a recently completed evaluation and carefully review the report.

To what extent were unobtrusive measures used in the evaluation?

What unobtrusive measures might be gathered to supplement the data used
in the evaluation?

Recommended readings:

Sechrest, L. (1979). Unobtrusive measurement today. San Francisco: Jossey-Bass.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1981). Unobtrusive measures: Nonreactive research in social sciences. Boston: Houghton Mifflin.