Meta-Assessment: Evaluating Assessment Activities
Author(s): John C. Ory
Source: *Research in Higher Education*, Vol. 33, No. 4 (Aug., 1992), pp. 467-481
Published by: Springer
Stable URL: http://www.jstor.org/stable/40196044
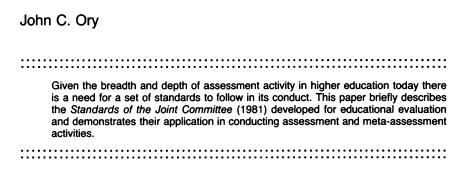Accessed: 08-02-2017 15:17 UTC

META-ASSESSMENT:
Evaluating Assessment Activities

John C. Ory

....
....

Given the breadth and depth of assessment activity in higher education today there is a need for a set of standards to follow in its conduct. This paper briefly describes the *Standards of the Joint Committee* (1981) developed for educational evaluation and demonstrates their application in conducting assessment and meta-assessment activities.

....
....

Several observations about the Assessment Movement in higher education prompted the writing of this paper. The first observation is the tremendous amount of interest in and conduct of assessment activities in the last five years. Evidence of this interest and activity is well documented in recent articles about the Assessment Movement (Halpern, 1987; Banta, 1988; Blumenstyk, 1988; Ewell, 1985; Gray, 1989). Cited are state legislative mandates for assessment activities, requests for assessment information by accreditation agencies, large attendances at national assessment conferences, the creation of assessment offices at many colleges and universities, and numerous books and articles published about assessment activities.

Observation number two is the movement's expanding definition of assessment. The initial focus of the current assessment movement was measuring student outcomes for the purpose of student development. Recently the scope of assessment has been expanded to include a broader view of purposes and processes. Today's campus assessment activities focus on students as well as faculty, programs, and the institution as a whole.

Explaining the shift in emphasis, Gray (1989) writes:

The leaders of the assessment movement have realized that they must embrace a broader view of assessment because it allows them to consider an appropriately di-

John C. Ory, Office of Instructional Resources, 307 Engineering Hall, 1308 W. Green St., University of Illinois, Urbana, IL 61801.

verse set of reasons for doing assessment; it provides the framework for conducting more comprehensive, thorough, and valid assessments; and it offers a structure for including a wide group of stakeholders. (p. 1)

Driving the Assessment Movement and shaping its definition are the demands for accountability or institutional effectiveness made by state legislators, regents, university/college administrators, taxpayers, parents, and students. A poor national economy, escalating costs of higher education, and public concern over the basic skills possessed by college graduates are some of the reasons cited (Westling, 1988) for the increasing demand for accountability in higher education.

My third observation came as a result of a survey (Ory and Parker, 1989) we conducted to determine the extent of assessment activities at other large universities. We found a diverse group of individuals responsible for campus assessment activities. Assessment was the responsibility of active and retired faculty across various disciplines, campus administrators, student affairs personnel, testing office staff, management information specialists, and educational researchers.

My fourth and final observation is one I share with Barbara Davis and several colleagues trained in educational evaluation. Davis (1989) writes, "In the rush to meet external demands for assessment, those involved in assessment have overlooked what the field of evaluation can contribute to their endeavors" (p. 6). There is very little mention of evaluation theory and practice in the assessment literature, yet for those of us trained in educational evaluation there is a strange sense of déjà vu as we read about the Assessment Movement.

The feverish pace required to respond to legislative mandates and the assignment of assessment responsibilities to individuals untrained in educational evaluation have resulted in the reinventing of some evaluation wheels and the making of old and familiar evaluation mistakes. For example, a common problem in assessment is the failure of faculty to accept assessment results and to use them to improve departmental curriculum. The utilization of evaluation literature (Braskamp and Brown, 1980; Patton, 1986) identifies general conditions and behaviors that promote the use of evaluation results. Many of these suggestions could be followed to encourage the use of assessment findings. Davis (1989) cites other areas of evaluation research that may be of interest to the Assessment Movement, including the existence of evaluation models, methodology for conducting case studies and naturalistic inquiry, criteria for judging merit and worth, or strategies for acknowledging and serving various stakeholders.

The purpose of this paper is to emphasize one other area of evaluation literature that should not be overlooked by the Assessment Movement. The literature

on meta-evaluation, or the evaluation of an evaluation, seems pertinent given the similarity between the conditions that exist today in the Assessment Movement and those that existed in the 1960s when evaluators began to write about meta-evaluation—namely a diverse range of evaluation/assessment activities being conducted in a variety of ways by a diverse group of individuals.

Worthen and Sanders (1987) describe how in the 1960s "Evaluators began to discuss formal meta-evaluation procedures and criteria [when] writers began to suggest what constituted good and bad evaluations (for example, Scriven, 1967; Stake, 1968, 1970; Stufflebeam, 1968)" (p. 370). Evaluators were trying to help other evaluators by discussing their failures and successes and their criteria for judging both outcomes. I believe that many conversations and correspondence between evaluators in the 1960s and 1970s parallel exchanges made at last year's national assessment conferences between individuals struggling with assessment endeavors. What should we be doing? How should we be doing it? How do we know if we did a good job?

Evaluator concern for a consumer-oriented, professionally developed set of guidelines or evaluation criteria led to the development of standards for judging an evaluation. A profession-wide Joint Committee on Standards for Educational Evaluation published a set of thirty standards in 1981, called the *Standards for Evaluation of Educational Programs, Projects, and Materials* (Joint Committee, 1981). As explained in the introduction of the *Standards*, the standards were developed to provide:

> a common language to facilitate communication and collaboration in evaluation; a set of general rules for dealing with a variety of specific evaluation problems; a conceptual framework by which to study the often-consuming world of evaluation; a set of working definitions to guide research and development on the evaluation process; a public statement of the art in educational evaluation; a basis for self-regulation and accountability by professional evaluators; and an aid to developing public credibility for the educational evaluation field. (Joint Committee, 1981, p. 5)

I believe if one were to replace the word *evaluation* with the word *assessment* in the above paragraph, individuals responsible for conducting campus assessments would see more clearly the relevancy of the evaluation literature to their work. Many of the problems addressed in the introduction are as troublesome for the assessment movement as they are for the field of evaluation. Given the breadth and depth of assessment activity in the nation, I believe there is a need for a set of standards to follow in its conduct. Furthermore, these standards should be used in the evaluation of our assessment efforts, or stated differently, in the conduct of a meta-assessment.

The remaining portion of this paper will briefly describe the thirty *Standards* of the Joint Committee and attempt to demonstrate their application in conduct-

ing a meta-assessment. In doing so, it is not the intent of the author to prescribe a model for conducting assessment. A reviewer of a draft of this paper provides the reasoning behind my decision, "In reading the manuscript . . . I had a feeling that something more concrete was needed in the 'comments.' After thinking about it last night I decided that I was wrong. There are too many possible paths in assessment research to begin setting out concrete recommendations and/or models."

The thirty *Standards* of the Joint Committee are divided into four major categories: utility (Does the evaluation serve practical information needs?), feasibility (Is the evaluation realistic and prudent?), propriety (Does the evaluation conform to legal and ethical standards?), and accuracy (Is the evaluation technically adequate?). Following is a brief description of each standard as written by Blaine Worthen and James Sanders in their textbook *Educational Evaluation: Alternative Approaches and Practical Guidelines* (Worthen and Sanders, 1987, pp. 372–375). The descriptions have been altered by replacing the word *evaluation* with *assessment*. Along with each description is a comment on how the standard is relevant or applicable to conducting assessment activities.

A. UTILITY STANDARDS

The Utility Standards are intended to ensure that an assessment will serve the practical information needs of given audiences. These standards are:

A1. Audience Identification

Description: Audiences involved in or affected by the assessment should be identified, so that their needs can be addressed.

Comment: Assessment can involve and affect a variety of on- and off-campus audiences. Generally speaking, different audiences have different needs. Audiences should be identified before planning an assessment so that their various needs can be determined and addressed. Audiences can include administrators, faculty, high school teachers, students, state boards of education, accreditation agencies, legislators, parents, and taxpayers.

A2. Evaluator Credibility

Description: The people conducting the assessment should be trustworthy and competent to perform the assessment so that their findings achieve maximum credibility and acceptance.

Comment. On many of today's campuses (Ory and Parker, 1989) assessment is being performed by many different types of individuals, including testing

office personnel, campus advisors, active and retired faculty, institutional re-searchers, management information specialists, and administrators. Most indi-viduals involved in campus assessment programs are found to be trustworthy, but many are limited in their assessment skills.

It is not uncommon, for example, for a retired English professor to be asked to direct a review of a new rhetoric curriculum. The professor will be chosen on the basis of his or her content expertise, years of experience, the high regard of colleagues, and amount of available time. After conducting a successful review the professor will most likely be asked to help with other assessment activities, possibly including the assessment of student opinions about the rhet-oric program or the development of placement exams. However, due to a lack of expertise in survey and testing procedures subsequent assessments may not be as successful as the first and the professor's credibility may be questioned. Conversely, situations occur wherein the credibility of a testing expert is chal-lenged not because of lack of technical expertise but because of lack of content knowledge and/or failure to be "of the faculty."

A3. Information Scope and Selection

Description: Information collected should be of such scope and should be selected in such ways as to address pertinent questions about the object of the assessment and should be responsive to the needs and interests of specified audiences.

Comment: Not all assessment information needs to be collected "from scratch." Often the necessary data for an assessment activity already exist on a campus but in a variety of places. Assessment staff can better respond to the information needs of their audiences by being knowledgeable of all campus offices and the type of information collected and maintained by each.

A4. Valuational Interpretation

Description: The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

Comment: Individuals in different campus offices represent different campus perspectives. For example, student affairs personnel may conduct assessment activities with a student-consumer orientation while a faculty member may con-duct assessment with a focused, discipline-based approach. A particular per-spective followed in the collection and interpretation of data should be ac-knowledged by identifying the affiliation of the individuals conducting the

assessment in the final report and by providing a statement of their perspective in the report's introduction.

A5. Report Clarity

Description: The assessment report should describe the object being assessed and its context, and the purposes, procedures, and findings of the assessment so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

Comment. Beside including the components identified above (i.e., a clear description of the program, assessment procedures, and findings) in a report of findings, individuals responsible for assessment need to be aware of the "clarity" needs of the critical audiences. Faculty audiences may be more likely to understand and accept reports containing numerical tables and statistical tests than may be a parent association or a group of legislators.

A6. Report Dissemination

Description: Assessment findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

Comment: Often evaluators and individuals responsible for assessment do not have the authority to disseminate the results to people other than those commissioning the assessment. However, when given the opportunity to do so, it should be the responsibility of the assessment personnel to see that the assessment results are placed in the hands of people who can best respond to the information. It seems that too often assessment is conducted to satisfy external mandates for the information. The assessment is completed and the information is sent to the individual or group of individuals creating the mandate. What about the potential use of the information by internal or campus audiences? How often do the faculty see retention rates, senior exit interview comments, alumni survey results, or departmental grade distributions?

A7. Report Timeliness

Description: Release of reports should be timely, so that audiences can best use the reported information.

Comment: Assessments should be conducted so the findings will be available at times when the information has greatest value and utility. Program reviews should be available at the time budgets are established; student course selection patterns or ratings of course quality should be examined prior to curriculum reviews. Timely reports should also contain "timely" information. Assessments

should be completed in a reasonable amount of time such that the data examined are no more than a semester or year behind the current period. It may not be worth examining the assessment of course selection patterns if many of the courses no longer exist due to the great amount of time spent conducting the assessment.

A8. Evaluation Impact

Description: Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

Comment: The literature on evaluation utilization (Braskamp and Brown, 1980; Patton, 1986) is worthwhile reading for individuals who desire greater utilization of their assessment efforts. Some of the strategies for enhancing evaluation utilization or impact written about in the literature include identifying the information needs of the critical audiences, getting them involved in the design and planning of the assessment, keeping them informed about the assessment as it progresses, making assessment results easily attainable and clearly understood, and reporting assessment results in ways that suggest alternative actions for change or improvement.

B. FEASIBILITY STANDARDS

The feasibility standards are intended to ensure that an assessment will be realistic, prudent, diplomatic, and frugal. They are:

B1. Practical Procedures

Description: The assessment procedures should be practical, so that disruption is kept to a minimum, and that needed information can be obtained.

Comment: This standard addresses a critical issue in assessment. Many assessment activities require students to take a test or respond to an attitudinal survey outside of regular course requirements, that is, rising junior exams, senior surveys. As a consequence, assessment personnel struggle with many practical issues regarding the collection of data. How can a test be administered to several hundred or thousand students without taking up valuable class time or finding a time that all students can attend a large group meeting (e.g., evenings, weekends) or without having difficulty motivating the students to perform to the best of their ability or as honestly as possible.

B2. Political Viability

Description: The assessment should be planned and conducted with anticipation of different positions of various interest groups so that their cooperation

may be obtained, and so that possible attempts by any of these groups to curtail assessment operations or to bias or misapply the results can be averted or counteracted.

Comment: Again, the literature on evaluation utilization suggests ways to use one's knowledge of political viability or the different makeup and needs of the various audiences to maximize the utilization of assessment results.

B3. Cost-Effectiveness

Description: The assessment should produce information of sufficient value to justify the resources extended.

Comment: Should minimal effort and expense be invested in projects that are completed to satisfy state or campus mandates and little else? Or should we invest as much time and money as necessary to get the maximum use out of any assessment? How often do we hear faculty complain that the money spent on assessment could be better spent improving the programs being assessed? "We all know that the students can't write. Don't waste your money finding out something that we already know. Instead, buy more English teachers!" Assessment personnel need to consider the cost of an assessment activity in light of its benefits.

C. PROPRIETY STANDARDS

The propriety standards are intended to ensure that an assessment will be conducted legally, ethically, and with due regard for the welfare of those involved in the assessment as well as those affected by its results. These standards are:

C1. Formal Obligation

Description: Obligations of the formal parties to an assessment (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or to formally renegotiate it.

Comment: The formal negotiation of an assessment contract may not be necessary because most assessment activities are conducted by internal campus personnel. However, the rationale for developing a contract or formal agreement should not be overlooked. It is often tempting once into an assessment to stretch the original boundaries of the activity and to assess or evaluate other objects. For example, faculty are often threatened by curriculum reviews because they often turn into reviews of personnel. This is not to say that assessments should fail to recognize unintended outcomes or side effects, but rather to stress that straying from the original intent of an assessment may cause political problems that undermine the credibility and acceptance of the activity.

C2. Conflict of Interest

Description: Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the assessment processes and results.

Comment: Conflicts of interest can occur because most assessments are conducted by internal campus offices. The difficulties encountered by internal versus external evaluation units have been written about in the evaluation literature (House, 1986; Scriven, 1967). Internal units do not go away after an assessment is completed and, unfortunately, negative messages are often linked to their messengers. Assessment personnel must work closely with many of the people responsible for the programs being assessed. It may be in the best interest of a particular individual or unit to refrain from conducting an assessment that may jeopardize a necessary and cooperative working relationship.

C3. Full and Frank Disclosure

Description: Oral and written assessment reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the assessment.

Comment: Assessments resulting in negative findings, even when conducted for state legislators and administrators, need to be reported and dealt with in an open and honest manner. A sufficient amount of data should be provided in the assessment report to enable the reader to verify the results and conclusions.

C4. Public's Right to Know

Description: The formal parties to an assessment should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

Comment: Campus policy needs to be established by the campus administration and the individuals conducting assessments regarding public access to information. The policy needs to address questions like, Who has the right to know? Who has the right to refuse access to the information? or What are the proper channels for public requests for information? There is a sense of doing "public good" when conducting an assessment activity in higher education, especially in public institutions. It is difficult for me to think of assessment results that should not be shared with the various campus and public audiences. Obvious assessment results that should be shared with the general public include campus crime statistics, graduation rates of student athletes, or minority admissions.

C5. Rights of Human Subjects

Description: Assessments should be designed and conducted so that the rights and welfare of the human subjects are respected and protected.

Comment: Often assessments involve the testing or surveying of students to determine group rather than individual ability levels or attitudes. Promises to the student respondents of anonymity or confidentiality should be honored.

C6. Human Interactions

Description: Assessment personnel should respect human dignity and worth in their interactions with other people associated with an assessment.

Comment: Any information collected during an assessment that may be personally damaging to an individual should be carefully handled whether it is pertinent to the assessment or not. Such information may be orally reported to an appropriate audience and not printed in a written report. "Off-the-record" comments should be respected and gossip should not be repeated.

C7. Balanced Reporting

Description: The assessment should be complete and fair in its presentation of strengths and weaknesses of the object under investigation so that strengths can be built on and problem areas can be addressed.

Comment: Seldom does there exist a single truth or correct perspective when conducting an assessment. Instead, multiple truths about a program, project, or activity exist in the minds of the various constituencies (Stake, 1968). It is the responsibility of assessment personnel to see that different perspectives are detected and revealed in the final report.

C8. Fiscal Responsibility

Description: Assessment allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.

Comment: Given consistently inadequate assessment budgets, assessment personnel are used to being judicious in their spending, therefore, forcing easy compliance with this standard.

D. ACCURACY STANDARDS

The Accuracy Standards are intended to ensure that an assessment will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit. These standards are:

D1. Object Identification

Description: The object of the assessment (program, project, activity) should be sufficiently examined so that the form(s) of the object being considered in the assessment can be clearly identified.

Comment: The boundaries of a project or program need to be clearly defined prior to the conduct of an assessment. If, for example, an assessment activity is to develop new placement exams, then efforts should not be invested in evaluating the delivery of course material. Sometimes a valuable by-product of an assessment is the clarification or description of the object being studied. For example, a review of campus support services for minority students may provide the first complete listing of all campus services provided by the university, colleges, and departments.

D2. Context Analysis

Description: The context in which the program, project, or material exists should be examined in enough detail so that its likely influences on the object can be identified.

Comment: The context of the object being assessed needs to be identified to help audiences better interpret and understand the results of the study. For example, to understand data related to the effectiveness of a TA training program it is important to know certain contextual factors, such as program resources (staff, budget, and space), campus emphasis on undergraduate teaching, or willingness of faculty members and administrators to participate in the training.

D3. Described Purposes and Procedures

Description: The purposes and procedures of the assessment should be monitored and described in enough detail so that they can be identified and evaluated.

Comment: Aside from meeting an external mandate for information, the purpose of an assessment activity is not always clear. Questions such as, Why bother? or For what reason are we doing this? need to be addressed in the planning stages of an assessment and in the final report. Using a utilization of information perspective, I would argue that a purpose statement should indicate potential uses of the information. Rather than having the purpose for a junior exam in writing skills be to test junior writing ability, I would suggest that the purpose is to determine competency levels in writing and diagnose common errors that may need to be better addressed in the curriculum.

D4. Defensible Information Sources

Description: The sources of information should be described in enough detail so that the adequacy of the information can be assessed.

Comment: Depending on who you ask on a typical campus you may get different answers to the same question. As bothersome as it is, departmental, college, and campus statistics often do not match. Different formulas are used or different policies are followed in determining various campus statistics, including enrollments, progress toward a degree, grade point average, retention rate, probationary status, or FTE (full-time equivalent) staff. In addition to trying to obtain the most accurate information, assessment personnel should indicate in the final report the source of the statistics used in the study.

D5. Valid Measurement

Description: The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

AND . . .

D6. Reliable Measurement

Description: The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information arrived at is valid for the given use.

Comment: Assessments make generous use of both standardized and locally developed instrumentation. Efforts should be made to determine the validity and reliability of an instrument either by reading available manuals and reviews of existing tests and scales or by conducting pilot studies for locally developed instruments.

D7. Systematic Data Control

Description: The data collected, processed, and reported in an assessment should be reviewed and corrected so that the results of the assessment will not be flawed.

Comment: A benefit of many campus assessments is the development of an information network. As previously stated, assessment data, or at least some portion thereof, often exist on campus prior to the conduct of an assessment. It is usually a matter of finding its location and learning how to access it. By documenting the data retrieval process there will be more systematic control of data for future assessments.

D8. Analysis of Quantitative Information

Description: Quantitative information in an assessment should be appropriately and systematically analyzed to ensure supportable interpretations.

AND . . .

D9. Analysis of Qualitative Information

Description: Qualitative information in an assessment should be appropriately and systematically analyzed to ensure supportable interpretations.

Comment: Care should be taken that the procedures used to analyze data are appropriate for the questions being asked in the assessment and the type of information being collected. Statistical procedures also should be selected on their potential for being understood and perceived as credible by the intended audiences. Graphs are more readily understood by less statistically trained audiences than are tables of analyses of variance and multiple regression results. Statistical tests should be appropriate for the type of data collected (i.e., parametric versus nonparametric statistics) and sample size should provide sufficient inferential power. Statistical significance should always be interpreted with respect to the practical significance of the finding.

In recent years much has been written in the area of evaluation and the analysis of qualitative data (Guba and Lincoln, 1981; Miles and Huberman, 1984; Patton, 1987). Qualitative analyses are used to look for patterns and categories in the data that enable the reader to draw defensible conclusions about the object being studied. Or, as Brinkerhoff and others (1983) state, "to seek confirmation and consistency." However, these same authors warn us not to force consensus when dealing with contradicting and conflicting evidence.

D10. Justified Conclusions

Description: The conclusions reached in an assessment should be explicitly justified so that the audience can assess them.

Comment: A final assessment report should include a sufficient amount of valid, reliable, and credible information that not only allows readers to judge the interpretations and conclusions of the author but allows them to draw their own conclusions. Each assessment audience has a unique perspective, history, and training that may cause them to interpret the results differently from one another. However, to draw their own conclusions audiences need to have a clear understanding of how the individual conducting the assessment developed his or her conclusions.

D11. Objective Reporting

Description: The assessment procedures should provide safeguards to protect the assessment findings and reports against distortion by the personal feelings and biases of any party to the assessment.

Comment: After an assessment is completed and the final report is submitted

to the client, usually a campus administrator, the author of the report has little control over the client's handling of the information. There is very little the author can do if the client wishes to change or delete findings prior to passing the report on to funding agents, legislators, or administrative boards. Individuals responsible for assessment must make it clear to their clients that it is in the best interest of the institution to accept less-than-positive findings and attempt to address their shortcomings.

CONCLUSIONS

Worthen and Sanders (1987) believe that the purpose of meta-evaluation is to "help evaluation live up to its potential." Could we say the same about the reason for conducting a meta-assessment? Many assessments are not living up to their potential. Many assessments are completed to satisfy mandates from external sources such as state legislators or accrediting agencies. Often institutional compliance drives the assessment more than does the potential usefulness of the activity. Unfortunately, the quality of the assessment is often of less importance than its mere completion.

Assessment has tremendous potential for improving the quality of instruction in higher education. Among other contributions, assessments can identify institutional strengths and weaknesses and indicate areas needing improvement. The *Standards* discussed in this paper can be used to help assessment reach its potential.

Individuals responsible for commissioning and conducting assessments can use the *Standards* to review assessment plans, monitor assessment activities, and evaluate completed assessments. The *Standards* can be followed internally by the person(s) responsible for the assessment or externally by individuals who are independent of the assessment. Awareness and knowledge of the *Standards* can also make the different assessment audiences better consumers. (Brinkerhoff and colleagues, 1983, pp. 205–207 provide an excellent summary of ways to use meta-evaluation/assessment.)

I was brought up with the notion that if something is worth doing then it is worth doing right. However, when I think about the many assessment activities being conducted today I am compelled to reverse the statement and say, "If assessment is done right, then it is worth doing." Doing assessment right means using assessment to encourage self-study and to suggest ways for institutional improvement. It means addressing the questions, "Now that the state board wants this information how can we learn about ourselves through planning the assessment? through studying the assessment process? or through analyzing the information collected?" I believe adherence to the professional standards throughout an assessment enhances our ability to answer these questions and to maximize our use of assessment efforts.

REFERENCES

Banta, T. W. (1988). *Implementing Outcomes Assessment: Promise and Perils*. New Directions for Institutional Research, no. 59. San Francisco: Jossey-Bass.

Blumenstyk, G. (1988). Diversity is keynote of states' efforts to assess students' learning. *Chronicle of Higher Education*, July 20, pp. A17, A25–A26.

Braskamp, L. S., and Brown, R. D. (eds.) (1980). *Utilization of Evaluative Information*. New Directions for Program Evaluation, no. 5. San Francisco: Jossey-Bass.

Brinkerhoff, R. O., Brethower, D. M., Hluchyj, T., and Nowakowski, J. R. (1983). *Program Evaluation: A Practitioner's Guide for Trainers and Educators*. Boston: Klower-Nijhoff.

Davis, B. G. (1989). Demystifying assessment: Learning from the field of evaluation. In Peter Gray (ed.), *Achieving Assessment Goals Using Evaluation Techniques*. New Directions for Higher Education, no. 67, pp. 5–20. San Francisco: Jossey-Bass.

Ewell, P. T. (ed.) (1985). *Assessing Educational Outcomes*. New Directions for Institutional Research, no. 47. San Francisco: Jossey-Bass.

Gray, P. J. (ed.) (1989). *Achieving Assessment Goals Using Evaluation Techniques*. New Directions for Higher Education, no. 67. San Francisco: Jossey-Bass.

Guba, E. G., and Lincoln, Y. S. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. San Francisco: Jossey-Bass.

Halpern, D. F. (ed.) (1987). *Student Outcomes Assessment: What Institutions Stand to Gain*. New Directions for Higher Education, no. 59. San Francisco: Jossey-Bass.

House, E. (1986). Internal evaluation. *Evaluation Practice* 7: 63–64.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects, and Materials*. New York: McGraw-Hill.

Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage.

Miles, M. B., and Huberman, A. M. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods*. Beverly Hills, CA: Sage.

Ory, J. C., and Parker, S. (1989). A survey of assessment activities at large research universities. *Research in Higher Education* 30: 373–383.

Patton, M. Q. (1987). *Creative Evaluation*, 2nd ed. Newbury Park, CA: Sage.

Patton, M. Q. (1986). *Utilization-Focused Evaluation*, 2nd ed. Newbury Park, CA: Sage.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (ed.), *Curriculum Evaluation*. American Educational Research Association Monograph Series on Evaluation, no. 1. Chicago: Rand McNally.

Stake, R. E. (1970). Objectives, priorities, and other judgement data. *Review of Educational Research* 40(2): 181–212.

Stake, R. E. (1968). The countenance of educational evaluation. *Teachers College Record* 68: 523–540.

Stufflebeam, D. L. (1968). *Evaluation as Enlightenment for Decision Making*. Columbus, OH: Ohio State University, Evaluation Center.

Westling, J. (1988). The Assessment Movement is based on a misdiagnosis of the malaise afflicting American higher education. *Chronicle of Higher Education*, October 19, pp. B1–B2.

Worthen, B. R., and Sanders, J. R. (1987). *Educational Evaluation: Alternative Approaches and Practical Guidelines*. New York: Longman, Inc.