24. PROFESSIONAL STANDARDS AND PRINCIPLES FOR EVALUATIONS

DANIEL L. STUFFLEBEAM

Members of most professions and many other public service fields must comply with given standards or codes of performance and service. Such standards and codes aim to protect consumers and society from harmful practices, provide a basis for accountability by the service providers, provide an authoritative basis for assessing professional services, provide a basis for adjudicating claims of malpractice, help assure that service providers will employ their field's currently best available practices, identify needs for improved technologies, provide a conceptual framework and working definitions to guide research and development in the service area, provide general principles for addressing a variety of practical issues in the service area, present service providers and their constituents with a common language to facilitate communication and collaboration, and earn and maintain the public's confidence in the field of practice. Such standards and codes typically are defined by distinguished members of the service area, in some cases by government licensing bodies, and occasionally with full participation of users groups. Familiar examples are the standards of practice employed by the fields of law, medicine, clinical psychology, educational testing, auditing, and accounting. Other examples are the codes established for the construction, engineering, electrical, plumbing, and food service areas.

Historically, program evaluators did not have to be concerned about explicit professional standards for program evaluations, because until relatively recently there was not any semblance of an evaluation profession and there were not any standards for evaluations. However, such standards have come into prominence during the 1980s

D.L. Stufflebeam, C.F. Madaus and T. Kellaghan (eds.). EVALUATION MODELS. *Copyright © 2000. Kluwer Academic Publishers. Boston. All rights reserved.*

and 1990s. Their appearance signifies both the field's historic immaturity and its comparatively recent movement toward professionalization.

In the early 1980s two programs for setting evaluation standards emerged and have survived. The Joint Committee on Standards for Educational Evaluation was established in 1975. Through the years, this standing committee has been sponsored by 12 to 15 professional societies with a combined membership totaling over 2 million. The committee's charge is to perform ongoing development, review, and revision of standards for educational evaluations. This committee issued the *Standards for Evaluations of Educational Programs, Projects, and Materials* in 1981 and an updated version in 1994 called *The Program Evaluation Standards.* The Joint Committee also published standards for evaluating educational personnel in 1988, and in the late 1990s has been working on a set of standards for evaluations of students. The Joint Committee is accredited by the American National Standards Institute (ANSI) as the only body recognized to set standards for educational evaluations in the U.S.

The Evaluation Research Society was established in 1976 and was focused on professionalizing program evaluation as practiced across a wide range of disciplines and service areas. This society published a set of 55 standards labeled the *Evaluation Research Society Standards for Program Evaluations* (ERS Standards Committee, 1982). In 1986, ERS amalgamated with the Evaluation Network (E NET) to form the American Evaluation Association (AEA), which has a membership of about 2,000. AEA subsequently produced the 1995 *AEA Principles for Program Evaluations.*

The purpose of this chapter is to describe and discuss standards and principles that have been posited for program evaluation. The ERS/AEA standards and principles cut across many areas of program evaluation, while the Joint Committee standards concentrate on evaluations of education and training programs and services. Both provide authoritative direction for assessing program evaluation studies. However, the Joint Committee standards are considerably more detailed than the ERS/AEA standards and principles and address practical and technical concerns of importance to the general practice of professional evaluation. The chapter is organized to look first at each program of evaluation standards/principles and second to consider how they are interrelated and complementary.

THE ERS STANDARDS FOR PROGRAM EVALUATIONS

The original *ERS Standards for Program Evaluations* (ERS Standards Committee, 1982) were developed to address program evaluations across a broad spectrum, e.g., community development, control and treatment of substance abuse, education, health, labor, law enforcement, licensing and certification, museums, nutrition, public media, public policy, public safety, social welfare, and transportation. In July of 1977, the ERS president appointed a seven-member committee to develop the ERS standards. All committee members were evaluation specialists, with Scarvia B. Anderson serving as chair. This committee collected and studied pertinent materials, such as the draft standards then being developed by the Joint Committee on Standards for Educational Evaluation. Since the ERS's focus was considerably wider than educa-

tional evaluations, the ERS Standards Committee decided to prepare a set of general standards that the Committee deemed to be broader in applicability than those being devised by the Joint Committee on Standards for Educational Evaluation. The ERS Standards Committee then produced a draft set of standards and circulated it mainly to ERS evaluation specialists. Using the obtained reactions, the committee finalized and published the standards in September of 1982.

The ERS standards are 55 admonitory, brief statements presented in about nine pages of text. An example is "1. The purposes and characteristics of the program or activity to be addressed in the evaluation should be specified as precisely as possible." The 55 standards are divided into the following six categories.

Formulation and Negotiation

The 12 standards in this group concretely advise evaluators that before proceeding with an evaluation they should clarify with their client as much as possible and in writing the evaluation work to be done, how it should be done, who will do it, who is to be served, protections against conflicts of interest, protections for participants and human subjects, the evaluation budget, and constraints on the evaluation. A general caveat for this subset of standards warns that initial evaluation planning decisions often must be revisited and revised as the evaluation evolves and circumstances change.

Structure and Design

The six standards concerned with structure and design note that evaluation plans must both prescribe a systematic, defensible inquiry process and take into account the relevant context. The key requirement here is to design the evaluation to produce defensible inferences about the value of the program being studied. The plan should clearly present and justify the basic study design, sampling procedures, data collection instruments, and arrangements for the needed cooperation of program personnel and other participants in the evaluation.

Data Collection and Preparation

The 12 standards here call for advance planning of the data collection process. The plan should provide for selecting and training data collectors; protecting the rights of data sources and human subjects; monitoring, controlling, and documenting data collection; controlling bias; assessing validity and reliability of procedures and instruments; minimizing interference and disruption to the program under study; and controlling access to data.

Data Analysis and Interpretation

Nine standards essentially call for tempering the data analysis and interpretation within the constraints of the evaluation design and data actually collected. These standards require evaluators to match the analysis procedures to the evaluation purposes; describe and justify use of the particular analysis procedures; employ appropriate units of analysis; investigate both practical and statistical significance of

quantitative findings; bolster cause-and-effect interpretations by reference to the design and by eliminating plausible rival explanations; and clearly distinguish among objective findings, opinions, judgments, and speculation.

Communication and Disclosure

Ten standards emphasize that evaluators must employ effective communication throughout the evaluation process. Particular requirements are to determine authority for releasing findings; organize data in accordance with the accessibility policies and procedures; present findings clearly, completely, fairly, and accurately; denote the relative importance of different findings; make clear the evaluation's underlying assumptions and limitations; be ready to explain the evaluation procedures; and disseminate pertinent findings to each right-to-know audience in accordance with appropriate, advance disclosure agreements.

Use of Results

The concluding "Use of Results" section includes six standards. These emphasize that evaluators should carefully attend to the information needs of potential users throughout all phases of the evaluation. Accordingly, evaluators should issue reports before pertinent decisions have to be made; anticipate and thwart, as much as possible, misunderstandings and misuses of findings; point up suspected side effects of the evaluation process; distinguish sharply between evaluation findings and recommendations; be cautious and circumspect in making recommendations; and carefully distinguish between their evaluative role and any advocacy role they might be playing.

The ERS standards are not the official standards of any group at this time. Their inclusion reflects their historical significance. Also, like the AEA guiding principles, they address a wide range of evaluations outside as well as inside education. Furthermore, the ERS standards are judged to be still valuable, since they apply to the full range of evaluation tasks, whereas the AEA guiding principles propose mainly a code of ethics for the behavior of evaluators.

THE AEA EVALUATION PRINCIPLES

Following the 1986 merger of E Net and ERS to create AEA, the amalgamated organization revisited the issue of professional standards for evaluators. After considerable discussion at both board and membership levels, the AEA leaders decided to supplement the ERS standards summarized above with an updated statement of evaluation principles. In November 1992, AEA created a task force and charged it to develop general guiding principles rather than standards for evaluation practice. The task force, chaired by William R. Shadish, subsequently drafted the *Guiding Principles for Evaluators.* Following a review process made available to the entire AEA membership, the task force finalized the principles document. After an affirmative vote by the AEA membership, the AEA board adopted the task force's recommended principles as the official AEA evaluation principles. AEA then published the principles in a special issue of AEA's *New Directions for Program Evaluation* periodical (Task

Force on Guiding Principles for Evaluation, 1995). The "Guiding Principles" are presented as a 6-page chapter in this special issue. The AEA guiding principles are consistent with the prior ERS Standards but shorter in the number of presented statements. Essentially, the AEA principles comprise 5 principles and 23 underlying normative statements to guide evaluation practice. The principles, with a summary of the associated normative statements, are as follows.

"A. *Systematic Inquiry:* Evaluators conduct systematic, data-based inquiries about whatever is being evaluated." This principle is supported by three normative statements. These charge evaluators to meet the highest available technical standards pertaining to both quantitative and qualitative inquiry. Evaluators are also charged to work with their clients to ensure that the evaluation employs appropriate procedures to address clear, important questions. The evaluators are charged further to communicate effectively, candidly, and in sufficient detail throughout the evaluation process, so that audiences will understand and be able to critique the evaluation's procedures, strengths, weaknesses, limitations, and underlying value and theoretical assumptions and also make defensible interpretations of findings.

"B. *Competence:* Evaluators provide competent performance to stakeholders." Three normative statements charge evaluators to develop and appropriately apply their expertise. Evaluator(s) must be qualified by education, abilities, skills, and experience to competently carry out proposed evaluations, or they should decline to do them. They should practice within the limits of their capabilities. Throughout their careers, evaluators should constantly use pertinent opportunities to upgrade their evaluation capabilities, including professional development and subjecting their evaluations to metaevaluations.

"C. *Integrity/Honesty:* Evaluators ensure the honesty and integrity of the entire evaluation process." Five normative statements are provided to assure that evaluations are ethical. Evaluators are charged to be honest and candid with their clients and other users in negotiating all aspects of an evaluation. These include costs, tasks, limitations of methodology, scope of likely results, and uses of data. Modifications in the planned evaluation activities should be recorded, and clients should be consulted as appropriate. Possible conflicts of interest should be forthrightly reported and appropriately addressed. Any misrepresentation of findings is strictly forbidden, and evaluators are charged to do what they can to prevent or even redress misuses of findings by others.

"D. *Respect for People:* Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact." The five normative statements associated with this standard require evaluators to show proper consideration to all parties to the evaluation. In focusing the evaluation, collecting information, and reporting findings, the evaluator should identify and respect differences among participants, e.g., age, disability, ethnicity, gender, religion, and sexual orientation. Pertinent codes of ethics and standards are to be observed in all aspects of the evaluation. The evaluator should maximize benefits to stakeholders and avoid unnecessary harms; observe informed consent poli-

cies; deal proactively, consistently, and fairly with issues of anonymity and confidentiality; and do whatever is appropriate and possible to help stakeholders benefit from the evaluation.

"E. *Responsibilities for General and Public Welfare*: Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare." Five normative statements are given to support this principle. Evaluators are charged not to be myopic but to show broad concern for the evaluation's social relevance. Evaluators have professional obligations to serve the public interest and good as well as the local need for evaluative feedback. They should consider the program's long-range as well as short-term effects, should search out side effects, and should present and assess the program's broad assumptions about social significance. They should balance their obligation to serve the client with services to the broader group of stakeholders. They should involve and inform the full range of right-to-know audiences and, within the confines of contractual agreements, give them access to the information that may serve their needs. In interpreting findings evaluators should take into account all relevant value perspectives or explain why one or some of these were excluded. Keeping in mind the interests and technical capabilities of their audiences, evaluators should report findings clearly and accurately.

THE JOINT COMMITTEE PROGRAM EVALUATION STANDARDS

The Joint Committee on Standards for Educational Evaluation developed the *Standards for Evaluations of Educational Programs, Projects, and Materials* between 1975 and 1980. This is a 161-page book that essentially includes detailed presentations of each of 30 standards. Each standard includes a statement of the standard, an explanation of its requirements, a rationale, guidelines for carrying it out, pitfalls to be anticipated and avoided, warnings against overzealous application, and an illustrative case.

The 30 standards are grouped according to four essential attributes of a sound evaluation: utility, feasibility, propriety, and accuracy. The Joint Committee advises both evaluators and clients to apply the 30 standards so that their evaluations satisfy all four essential attributes of a sound evaluation.

1. An evaluation should be **useful.** It should be addressed to those persons and groups that are involved in or responsible for implementing the program being evaluated. The evaluation should ascertain the users' information needs and report to them the relevant evaluative feedback clearly, concisely, and on time. It should help them to identify and attend to the program's problems and be aware of important strengths. It should address the users' most important questions while also obtaining the full range of information needed to assess the program's merit and worth. The evaluation should not only report feedback about strengths and weaknesses, but also should assist users to study and apply the findings.

The utility standards reflect the general consensus found in the evaluation literature that program evaluations should effectively address the information needs of clients and other right-to-know audiences and should inform program improvement processes.

2. An evaluation should be **feasible.** It should employ evaluation procedures that are parsimonious and operable in the program's environment. It should avoid disrupting or otherwise impairing the program. It should control as much as possible the political forces that might otherwise impede and/or corrupt the evaluation. And it should be conducted as efficiently and cost-effectively as possible. This set of standards emphasize that evaluation procedures must be workable in real world settings, not only in experimental laboratories. Overall, the feasibility standards require evaluations to be realistic, prudent, diplomatic, politically viable, frugal, and cost-effective.

3. An evaluation should meet conditions of **propriety.** It should be grounded in clear, written agreements defining the obligations of the evaluator and client for supporting and executing the evaluation. The evaluation should protect all involved parties' rights and dignity. Findings must be honest and not distorted in any way. Reports must be released in accordance with advance disclosure agreements. Moreover, reports should convey balanced accounts of strengths and weaknesses. These standards reflect the fact that evaluations can affect many people in negative as well as positive ways. The propriety standards are designed to protect the rights of all parties to an evaluation. In general, the propriety standards require that evaluations be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation as well as those affected by the results.

4. An evaluation should be **accurate.** It should clearly describe the program as it was planned and as it was actually executed. It should describe the program's background and setting. It should report valid and reliable findings. It should identify the evaluation's information sources, measurement methods and devices, analytic procedures, and provisions for bias control. It should present the strengths, weaknesses, and limitations of the evaluation's plan, procedures, information, and conclusions. It should describe and assess the extent to which the evaluation provides an independent assessment rather than a self-assessment. In general, this final group of standards require evaluators to obtain technically sound information, analyze it correctly, and report justifiable conclusions. The overall rating of an evaluation against the 12 accuracy standards is an index of the evaluation's overall validity.

The 17 members of the original Joint Committee were appointed by 12 professional organizations. The organizations and their appointed members represented a wide range of specialties—school accreditation, counseling and guidance, curriculum, educational administration, educational measurement, educational research, educational governance, program evaluation, psychology, statistics, and teaching. A fundamental requirement of the Committee is that it include about equal numbers

of members who represent evaluation users groups and evaluation methodologists. Over the years the Joint Committee's sponsoring organizations have slightly increased. (At the publication of the 1994 *The Program Evaluation Standards,* the committee was sponsored by 15 organizations, including AEA.[1]) Daniel L. Stufflebeam chaired the Joint Committee during its first 13 years, James R. Sanders served as chair during the next 10 years, and Arlen Gullickson has been the chair since the end of 1998, All three are members of the Western Michigan University Evaluation Center, which has housed and supported the Joint Committee's work since its inception in 1975.

In each of its standards-setting projects, the Joint Committee engaged about 200 persons concerned with the professional practice of evaluation in a systematic process of generating, testing, and clarifying widely shared principles by which to guide, assess, and govern evaluation work in education. In each project, the Committee sought widely divergent views on what standards should be adopted. The Committee subsequently worked through consensus development processes to converge on the final set of standards.

Each set of Joint Committee *Standards* is a living document. The Joint Committee is a standing committee. The Committee encourages users of each set of standards to provide feedback on applications of the standards along with criticisms and suggestions. From the outset of its work, the Joint Committee has provided for periodic reviews and improvement of the standards. This feature of its work is consistent with requirements for maintaining the Committee's accreditation by the American National Standards Institute (ANSI).

The Committee's review of its 1981 program evaluation standards led to the development of a second edition, *The Program Evaluation Standards* published in 1994. Like the first edition, 30 standards are presented within the 4 categories of utility, feasibility, propriety, and accuracy. The Committee merged some of the original standards and added some new ones. New illustrative cases were included that pertain to more diverse areas of application than did the illustrations in the 1981 version. The 1994 version covers education and training in such settings as business, government, law, medicine, the military, nursing, professional development, schools, social service agencies, and universities.

The Program Evaluation Standards (Joint Committee, 1994) are summarized in Table 1[2]. ANSI approved these *Standards* as an American National Standard on March 15, 1994. Readers are advised to study the full text of *The Program Evaluation Standards,* so that they can internalize them and apply them judiciously at each stage of an evaluation. The summary presented in Table 1 is only a starting point and convenient memory aid.

The Joint Committee offered advice on which of the above 30 standards are most applicable to each of 10 tasks in the evaluation process: deciding whether to evaluate, defining the evaluation problem, designing the evaluation, collecting information, analyzing information, reporting the evaluation, budgeting the evaluation, contracting for evaluation, managing the evaluation, and staffing the evaluation. The Committee's judgments of the different standards' applicability to each evaluation

Table 1. Summary of the Program Evaluation Standards

UTILITY

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1 Stakeholder Identification. Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

U2 Evaluator Credibility. The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3 Information Scope and Selection. Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

U4 Values Identification. The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

U5 Report Clarity. The evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

U6 Report Timeliness and Dissemination. Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7 Evaluation Impact. Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

FEASIBILITY

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

F1 Practical Procedures. The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

F2 Political Viability. The evaluation should be planned and conducted with anticipation of the 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 evaluation operations or to bias or misapply the results can be averted or counteracted.

F3 Cost Effectiveness. The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

PROPRIETY

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

P1 Service Orientation. Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

P2 Formal Obligations. Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obliged to adhere to all conditions of the agreement or formally to renegotiate it.

P3 Rights of Human Subjects. Evaluations should be designed and conducted to respect and to protect the rights and welfare of human subjects.

P4 Human Interactions. Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

P5 Complete and Fair Assessment. The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

Table 1 (continued)

PROPRIETY

P6 Disclosure of Findings. The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.

P7 Conflict of Interest. Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

P8 Fiscal Responsibility. The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

ACCURACY

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1 Program Documentation. The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

A2 Context Analysis. The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

A3 Described Purposes and Procedures. The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4 Defensible Information Sources. The sources of information used in program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5 Valid Information. The information-gathering procedures should be chosen or developed and then implemented, so that they will ensure that the interpretation arrived at is valid for the intended use.

A6 Reliable Information. The information-gathering procedures should be chosen or developed and then implemented, so that they will ensure that the information obtained is sufficiently reliable for the intended use.

A7 Systematic Information. The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.

A8 Analysis of Quantitative Information. Quantitative information in an evaluation should be appropriately and systematically analyzed, so that evaluation questions are effectively answered.

A9 Analysis of Qualitative Information. Qualitative information in an evaluation should be appropriately and systematically analyzed, so that evaluation questions are effectively answered.

A10 Justified Conclusions. The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

A11 Impartial Reporting. Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

A12 Metaevaluation. The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Source: Joint Committee on Standards for Educational Evaluation (1994).

task are summarized in Table 2. The 30 standards are listed down the side of the matrix, while the 10 evaluation tasks are presented across the top. The Xs in the various cells indicate that the Committee judged the standard was particularly applicable to the given task. While the Joint Committee concluded that all of the standards are applicable in all educational program evaluations, the functional analysis is intended to help evaluators quickly identify those standards that are likely to be most relevant to given tasks.

The Committee also presented and illustrated five general steps for applying the standards. These are (1) become acquainted with *The Program Evaluation Standards,* (2) clarify the purposes of the program evaluation, (3) clarify the context of the program evaluation, (4) apply each standard in light of the purposes and context, and (5) decide what to do with the results. The Committee also suggested ways to employ the standards in designing an evaluation training program.

The Program Evaluation Standards are particularly applicable in evaluations of evaluations, i.e., metaevaluations. In such studies, the metaevaluator collects information and judgments about the extent to which a program evaluation complied with the requirements for meeting each standard. Then the evaluator judges whether each standard was "addressed," "partially addressed," "not addressed," or "not applicable." A profile of these judgments provides bases for judging the evaluation against the considerations of utility, feasibility, propriety, and accuracy, and in relation to each standard. When such metaevaluations are carried out early in an evaluation, they provide diagnostic feedback of use in strengthening the evaluation. When completed after a program evaluation, the metaevaluation helps users to assess and make prudent use of the evaluation's findings and recommendations.

PERSONNEL EVALUATION STANDARDS

As mentioned earlier, the Joint Committee also developed *The Personnel Evaluation Standards* (1988). This document includes 21 standards organized according to the four basic concepts of propriety, utility, feasibility, and accuracy. These standards reflect the fact that personnel qualifications and performance are critically important concerns for evaluating programs and that personnel evaluation is important in its own right for helping to assure the delivery of sound, ethical professional services. *The Personnel Evaluation Standards* are designed to give educators and board members a widely shared view of general principles for developing and assessing sound, respectable, and acceptable personnel evaluation systems, plus practical advice for fulfilling the principles.

Institutions need effective personnel evaluation systems to help select, retain, and develop qualified personnel and to supervise and facilitate their work and development. Individual professionals need valid assessments of their performance to provide direction for improvement and be accountable for the responsiveness and quality of their services. The state of personnel evaluation in educational institutions has been poor (Joint Committee, 1988). The Joint Committee's sponsoring organizations charged the Committee to devise personnel evaluation standards that institutions

Table 2. Analysis of the Relative Importance of 30 Standards in Performing 10 Tasks in an Evaluation

	1. Deciding Whether to Evaluate	2. Defining the Evaluation Problem	3. Designing the Evaluation	4. Collecting Information	5. Analyzing Information	6. Reporting the Evaluation	7. Budgeting the Evaluation	8. Contracting for the Evaluation	9. Managing the Evaluation	10. Staffing the Evaluation
U1 Stakeholder Identification	×	×	×			×		×	×	
U2 Evaluator Credibility	×			×				×	×	×
U3 Information Scope & Selection			×	×		×	×	×		
U4 Values Identification			×	×	×	×				
U5 Report Clarity						×				
U6 Report Timeliness & Dissemination						×		×	×	
U7 Evaluation Impact	×					×				
F1 Practical Procedures			×	×					×	
F2 Political Viability	×			×				×	×	×
F3 Cost Effectiveness	×				×		×		×	
P1 Service Orientation	×	×				×		×	×	
P2 Formal Agreements	×		×	×			×	×	×	
P3 Rights of Human Subjects				×		×		×	×	

P4 Human Interactions				×						
P5 Complete & Fair Assessment			×	×		×				
P6 Disclosure of Findings						×		×	×	
P7 Conflict of Interest	×							×	×	×
P8 Fiscal Responsibility							×	×		
A1 Program Documentation	×	×	×	×	×	×	×	×		
A2 Context Analysis	×	×		×	×	×			×	
A3 Described Purposes & Procedures		×	×	×		×	×	×		
A4 Defensible Information Sources			×	×		×				
A5 Valid Information			×	×						
A6 Reliable Information			×	×					×	
A7 Systematic Information				×						
A8 Analysis of Quantitative Information			×		×					
A9 Analysis of Qualitative Information			×		×					
A10 Justified Conclusions			×		×	×				
A11 Impartial Reporting			×			×				×
A12 Metaevaluation	×	×	×	×	×	×	×	×	×	×

could use to correct weaknesses in their personnel evaluation practices and/or develop new, sound personnel evaluation systems.

The 1988 *Personnel Evaluation Standards* are focused on assessing and improving the systems that educational organizations use to evaluate instructors, administrators, support staff, and other educational personnel. This book is intended to be used by board members and educators in school districts, community colleges, four-year colleges, universities, professional development organizations, and other educational institutions.

The utility standards were placed first in *The Program Evaluation Standards,* because program evaluations often are ad hoc. A program evaluation would be done not as a matter of course, but because it is needed and could make an important difference in delivering and improving services. Evaluators and their clients should first make sure that findings from a program evaluation under consideration would be used before taking the trouble to address concerns for feasibility, propriety, and accuracy. For example, it makes no sense to develop a sound data collection and analysis plan, a contract, and a budget if no one is likely to read and act on the projected report. In such a case it is better to abort the evaluation as soon as it is known that carrying it out would make no difference. For these reasons, evaluators should first apply the utility standards to assure that an evaluation could impact on program quality and delivery. If there is no prospect for use, then the evaluator and client should stop the process. In that event they need not look at the standards of feasibility, propriety, and accuracy. But if there is a good prospect for utility, the evaluator should systematically turn to consideration of the full set of standards.

The situation in personnel evaluation is different. Mainly, personnel evaluations are not ad hoc. They are basically inevitable, no matter how badly they will be done. Thus, the Joint Committee said the personnel evaluator should deal first with the contemplated evaluation's propriety. A key reason for this decision is that the first propriety standard addresses the issue of service orientation. This standard emphasizes that the fundamental purpose of personnel evaluation must be to provide effective, safe, and ethical services to students and society. Personnel evaluations especially must help protect the interests of students by uncovering harmful practices of teachers, administrators, etc., as well as providing feedback to help such persons improve their services to the students. The bottom line thrust of *The Personnel Evaluation Standards* is to help assure that students are served well, that services constantly improve, and that harmful practices are quickly uncovered and promptly addressed.

To balance this emphasis on service orientation, *The Personnel Evaluation Standards* also stress that personnel evaluation practices should be constructive and free of unnecessarily threatening or demoralizing characteristics. In this positive vein, personnel evaluations can and should be employed to help plan sound professional development experiences and help each professional assess and strengthen her or his performance. Such evaluations should identify the educator's deficiencies and strengths.

COMPARISON OF THE JOINT COMMITTEE STANDARDS AND THE ERS/AEA STANDARDS AND PRINCIPLES

Comparisons of the substance of the ERS/AEA and Joint Committee standards and principles documents reveal key differences and similarities in the standards and principles (Cordray, 1982; Covert, 1995; Sanders, 1995; Stufflebeam, 1982). While the Joint Committee's standards focused on evaluations in education, the ERS standards and principles addressed evaluations across a variety of government and social service sectors. Essentially everything covered by the ERS standards is also covered by the Joint Committee's standards, but the latter's coverage is much more detailed and goes deeper into evaluation issues. The Joint Committee's presentations of standards have averaged more than 100 pages, while the ERS/AEA presentations of standards and principles each numbered less than 10 pages. Further, the Joint Committee standards were developed by a joint committee whose 17 members were appointed by 12 professional organizations with a total membership of over 2 million. The ERS standards and the AEA principles were developed by single organizations with memberships of about 1,000 and 2,000, respectively. The standards and principles-development task forces of these organizations respectively had 6 and 4 evaluation specialists respectively, whereas the Joint Committee had 17 members. Another key difference is that the Joint Committee standards were developed by a combination of evaluation users and evaluation specialists, while the ERS standards and AEA principles were developed almost exclusively by evaluation specialists. Finally, the AEA principles were formally adopted by AEA, whereas the Joint Committee's 1994 *Program Evaluation Standards* were accredited by ANSI, but have not been formally adopted by any of the Committee's sponsoring organizations.

The differences in lengths of the documents reflect perhaps somewhat different purposes. The ERS/AEA efforts have focused almost exclusively at the level of general principles to be observed by evaluators. The Joint Committee also stresses general principles—as seen in its requirements for utility, feasibility, propriety, and accuracy—but also attempts to provide specific and detailed standards of good practice along with guidelines for meeting the standards. In this sense, the Joint Committee's standards include both general requirements of sound evaluations and rather specific advice for meeting these requirements. Nevertheless, both standards/principles-setting programs emphasize that the standards and principles must be seen as general guides and that evaluators and their clients must consult and employ much more specific material when dealing with the details of design, measurement, case studies, statistics, reporting, etc.

Both sets of documents are in substantial agreement as to what constitutes sound evaluation practices. Evaluators should seek out and involve their intended audiences in clarifying evaluation questions and in reporting evaluation findings. Evaluations should be beyond reproach, with evaluators adhering to all relevant ethical codes. Moreover, evaluators should strive to produce valid findings and should be careful not to present unsupportable conclusions and recommendations. In addition, evaluators should carefully sort out their roles as independent inquirers from their social advocacy roles and make sure that their evaluations are not corrupted by con-

flicts of interest. Also, the Joint Committee standards, the ERS standards, and the AEA principles concur that evaluations occur in politically charged, dynamic social settings and call on evaluators to be realistic, diplomatic, and socially sensitive, while maintaining their integrity as evaluators. Both standards/principles-setting movements stress that sound evaluation is vital to the functioning of a healthy society. Service providers must regularly subject their services to evaluation, and evaluators must deliver responsive, dependable evaluation services. Professional standards are a powerful force for bringing about the needed sound evaluation services.

CLOSING COMMENTS

This chapter has provided an overview of the state of standards setting in the field of evaluation as practiced in the U.S. Professional standards and principles are seen as important for assessing and strengthening evaluation practices. It is a mark of American evaluators' move toward professionalism that two separate but complementary standards/principles-development movements are now more than two decades old and continuing. It is fortunate that two sets of standards/principles have been developed. They provide cross-checks on each other, even though they are appropriately aimed at different constituencies. The two sets of presentations have proved to be complementary rather than competitive. The ERS/AEA standards and principles address evaluations across a wide range of disciplines and service areas, while the Joint Committee standards have honed in on education. It should be reassuring to educational evaluators that all the important points in the ERS/AEA standards and principles are also covered in the Joint Committee standards. There seem to be no conflicts about what principles evaluators should follow in the two sets of materials. Moreover, evaluators outside education can find that the details in the Joint Committee standards can help to buttress the general principles in the ERS/AEA standards and principles. (For example, see Patton's chapter in this book.)

For the future the two groups should continue to work at reviewing and updating the standards and principles as needed. They should also promote effective use of the standards and principles. Especially, they should encourage evaluation educators to build the standards and principles into every evaluation degree program and into special training sessions for evaluation users as well as specialists. Evaluators should also employ the evaluation standards and principles to conduct and report metaevaluations. If the standards and principles become well established and if they are regularly applied, then both evaluation consumers and producers will benefit. Adherence to the evaluation standards and principles will improve the quality of evaluations and should increase their value for improving programs and services. These points seem to provide an ample rationale for evaluators to obtain, study, apply, and help to improve the ERS/AEA standards and principles and the Joint Committee standards for program and personnel evaluations. The fact that these efforts developed independently gives added credibility to the consensus reflected in their reports about what constitutes good and acceptable evaluation practice. Now it is time for collaboration as the Joint Committee and AEA move ahead to advance

the professional practice of evaluation through adherence to high standards and principles of practice.

NOTES

1. The membership of the Joint Committee on Standards for Educational Evaluation, as of publication of the 1994 *The Program Evaluation Standards,* included the American Association of School Administrators, American Educational Research Association, American Evaluation Association, American Federation of Teachers, American Psychological Association, Association for Assessment in Counseling, Association for Supervision and Curriculum Development, Canadian Society for the Study of Education, Council of Chief State School Officers, Council on Postsecondary Accreditation, National Association of Elementary School Principals, National Association of Secondary School Principals, National Council on Measurement in Education, National Education Association, and National School Boards Association.

2. The summary statements of the 30 program evaluation standards are printed here with the permission of the Joint Committee for Educational Evaluation.