
❖ Introduction ❖

Each year, people spend trillions of dollars on programs designed to improve knowledge, skills, performance, and attitudes. We invest in these formal and informal programs in the hope that we and our communities will change for the better. With so much time, money, and other resources directed to educational and social programs, how can we be sure that our programs are efficient and effective? How can we improve them and hold them accountable for results?

This need to address program quality systematically has helped program evaluation develop into a formal specialty. By the late 1960s, program developers, sponsors, and evaluators recognized the need for rigorous standards to guide program evaluation. In 1974, representatives from three national professional organizations came together to form the Joint Committee on Standards for Educational Evaluation (JCSEE). By 1981, the JCSEE was a 501(c)(3) corporation supported by 12 North American professional organizations and had published its first set of evaluation standards, the *Standards for Evaluations of Educational Programs, Projects and Materials*. A full, detailed history of the JCSEE is beyond the scope of this introduction, but additional references are available on our website. To learn more about the history and organizational support of the JCSEE, visit <http://www.jcsee.org>.

Today, the JCSEE is supported by 17 sponsoring organizations and has been a member of the American National Standards Institute (ANSI) since 1989. During its 35-year history, the mission of the JCSEE has remained constant: to develop and implement inclusive processes producing widely used evaluation standards that serve educational and social improvement. In keeping with this mission, it has developed and disseminated three sets of evaluation standards: *The Program Evaluation Standards* (1981, 1994, 2011), *The Personnel Evaluation Standards* (1987, 2009) and *The Student Evaluation Standards* (2004, in development).

❖ EVALUATION STANDARDS DEVELOPMENT

The cornerstone of JCSEE standards development is the ongoing involvement of diverse stakeholders representing different perspectives. JCSEE procedures encourage stakeholders from inside and outside North America, including stakeholders who have not previously contributed, to participate in each new standards development. Moreover, regular renewal efforts also keep the standards informed by current scholarship and practice. Following JCSEE procedures, formal reviews must take place on a 5-year cycle. Every 10 years, the standards must go through a complete review process. These reviews must take place in order to maintain ANSI approval.

The development process for this third edition of *The Program Evaluation Standards* began in 2004 and followed the same procedures as previous editions. It included the following key components:

1. Formal initiation of standards review
2. Needs assessment
3. Development of draft standards
4. External validation panel monitoring and oversight
5. Regular reports, process monitoring, and approval by the JCSEE
6. National and international reviews
7. National hearings
8. Revision of multiple draft manuscripts responding to stakeholder and JCSEE suggestions and comments
9. Stakeholder field trials of revised drafts
10. ANSI process monitoring and comment periods
11. JCSEE vote to approve the final version of the standards
12. Dissemination and research on standards use and quality

In 2002, the JCSEE approved an addition to these procedures: a task force to lead each specific standards renewal process. The change was in response to the increased responsibilities for three different sets of standards and was first used with *The Personnel Evaluation Standards* (completed in 2009). Continuing this model, a JCSEE-approved Program Evaluation Standards task force led the revision process and produced the revised manuscripts. The JCSEE retained its

responsibility for regular review and final approval of the developing manuscript, but the manuscript was written, reviewed, and revised by the task force. Other procedures and components in the standards development process have remained the same as they were for previous editions.

In the development of the current edition, key considerations included the content and format of the standards statements and the clarifications and rationales for them. In the five years from initiation to final manuscript, the task force and JCSEE members conducted a North American needs assessment, a format survey about changes to the standard statements, numerous formal and informal presentations, discussions at national and international conferences, and extensive reviews of scholarship and practice literature on program evaluation, as well as all the steps in the process described above.

❖ TECHNICAL LANGUAGE IN *THE PROGRAM EVALUATION STANDARDS*

One important goal for standards in general is to clarify technical terms and make otherwise vague definitions fully operational and useful. Careful attention to terminology is not only typical for standards work but is also found in any craft or profession that has developed specialized tools and high-quality professional practices. For example, customers would worry about a plumber who could not read the local plumbing codes and skillfully implement standard practices effectively and efficiently.

Similarly, clarifying *The Program Evaluation Standards* requires concept and language specificity. *The Program Evaluation Standards* will use some technical terms that may not be familiar to all stakeholders; for example, *randomized field trial*, *quasiexperiment*, *ethnography*, *process use*, *instrumental use*, *program theory*, *service delivery*, and *institutional review boards*. In addition, they will rely on important concepts that have more nuanced meanings than in common parlance, for example, *programs* and *projects*, *standards*, *utility*, *feasibility*, *propriety*, *accuracy* and *accountability*. Key terms will be described fully in the text or in the appended glossary.

Clarifying these terms serves one primary purpose—to improve communication about value and quality. For example, not all stakeholders will be adept at implementing randomized field trials, ethnographies, or other evaluation methodologies and procedures. However, familiarity with such terms and concepts helps with communication about the nature and value of information to assist decision making. When program and evaluation stakeholders have knowledge of the concepts and terms used to ground the standards, they can engage effectively in conversations that guide programs and their evaluations.

❖ IMPORTANT ROLES FOR PROGRAM AND EVALUATION STAKEHOLDERS

Anyone whose life is affected by a program or its evaluation is a *stakeholder* in that program and its evaluation. The standards have been developed with consideration of the many different roles that stakeholders and groups of stakeholders play in program development, implementation, and evaluation.

With regard to program development and implementation, some roles are more administrative, such as reviewing programs for funding, guiding program designs and implementation, and deciding whether and how to implement programs. Other roles, especially those of the program staff and participants, support the detailed work of daily program implementation. Some roles, such as those of staff members or participants, usually engage numerous individuals. In nearly all programs, specific individuals may engage in multiple roles.

Similarly, many different stakeholders and groups of stakeholders play important evaluation roles. Some of the most important are

- proposal reviewers—those who scrutinize, evaluate, and decide which proposed programs receive funding based on considerations of needs and program quality, including the quality of the proposed program evaluation;
- sponsors—those who fund the program evaluation;
- clients—those who commission the evaluation;
- evaluators—anyone with professional experience and training in systematic program evaluation serving in a professional capacity;
- designers—evaluators and other stakeholders who work together to plan the evaluation and its purposes, goals, and objectives, including technical and domain specialists, selected program staff, participant group representatives, community or organizational representatives, government representatives, and consultants;
- implementers—evaluators and other stakeholders who work together to manage, administer, orchestrate, oversee, implement, and assure the quality of the evaluation, including evaluation staff, other technical specialists, selected program staff, volunteers, and consultants;
- evaluation participants—those who take part in the programs and provide information and perspectives for evaluation purposes;
- other respondents—anyone else who provides important information about the program;
- intended users—those whose needs are identified and addressed by evaluation processes and products, resulting in improved knowledge, skills, decisions, or other actions;

- other users—those whose needs are not specifically identified during evaluation planning but who have access to the evaluation processes and products, resulting in changes in knowledge and understanding of the program, decisions, and other actions;
- evaluation educators—those who help other learners, including program and evaluation stakeholders and beginning evaluators, develop more knowledge and skill in evaluation; and
- metaevaluators—those who evaluate program evaluations using these and other standards.

Just as with programs, individual evaluation stakeholders can play single or multiple roles in the evaluation.

❖ WHAT ARE *STANDARDS*?

The word *standard* as used in *The Program Evaluation Standards* has two key features. First, the standards identify and define evaluation quality and guide evaluators and evaluation users in the pursuit of evaluation quality. Second, these standards are not “laws” but are voluntary, consensus statements developed with extensive stakeholder input and then discussed, revised, and approved by the members of the JCSEE following ANSI procedures.

These standards are both similar to and unlike other types of standards in important ways. Like technical standards that specify exact dimensions and tolerances for manufactured items or content standards that outline learning processes and outcomes, evaluation standards identify ways to improve quality. However, unlike many technical and content standards, these evaluation standards do not specify exact procedures to be followed in any specific setting. They require responsiveness and judgment in each evaluation setting. In problem-solving terminology, these standards provide *heuristics* to be used in defining and addressing the novel problems that evaluation settings present.

An important characteristic of the individual program evaluation standards is that they exist in dynamic tension with each other. Few if any evaluations provide the opportunity to maximize quality in the application of each standard. Because of these limitations, a balanced application of individual standards depends on human values and choices in specific situations. In implementing the program evaluation standards, stakeholders must decide how to create the best quality evaluations based on prioritized needs. The standards can be applied to all evaluations, but the exact ways they are applied will differ.

Taken as a whole, the program evaluation standards address the possible dimensions of quality in program evaluations. To help clarify the standards and make them manageable, they are organized according to the five general attributes of quality: evaluation utility, feasibility, propriety, accuracy, and accountability. Each of these attributes of evaluation quality and their supporting standards are discussed thoroughly in separate parts of the book. Each part presents case scenarios and applications to address how the standards can help evaluators and evaluation users respond to complex evaluation challenges.

Another distinguishing characteristic of these standards is their lack of regulatory status. In contrast to standards defined by law or regulatory code, the program evaluation standards are *open standards*. *Open* in this sense means that they are voluntary and consensual even though they are developed with due process.

Evaluators and evaluation users can, however, agree contractually to be guided by *The Program Evaluation Standards* in specific evaluation work. They can agree to investigate whether the standards are well-implemented and how well they supported evaluation quality. The JCSEE strongly recommends that evaluators and evaluation users commit themselves to evaluation practices based on these standards.

❖ WHAT ARE *PROGRAMS* AND *PROJECTS*?

No discussion of *The Program Evaluation Standards* would be complete without describing and defining *programs* and *projects*. Earlier editions of *The Program Evaluation Standards* defined educational programs as “activities that are provided on a continuing basis” and defined educational projects as “activities that are provided for defined periods of time” (JCSEE, 1994, p. 3).

In this third edition, the definitions have expanded. Today, the general consensus among program and project managers and administrators is that programs and projects are not distinguished primarily by duration, since some projects last longer than some programs; nor do projects and programs consist of qualitatively different types of structures and functions. What distinguishes the two is their relationship to one another in specific situations. For example, it is quite common for federal programs to provide the framework for specific state-level and local projects, but rarely do projects provide the framework for programs. With regard to applications of these standards, it is not necessary to distinguish between programs and projects except to identify how they are related in specific evaluations.

This third edition also emphasizes that programs are much more than just activities. They consist of multiple components. In addition, all important components of programs can be the objects of evaluation in their own

right or can be part of a componential evaluation of the whole program. Specific components that can be evaluated include

- contexts and how they interact with programs and program components;
- participants and other beneficiaries as well as those who encounter costs or loss of benefits;
- needs, problems, and policy spaces in programs and their contexts;
- goals and objectives;
- resources and costs of all kinds, including staff, facilities, materials, and opportunity costs;
- activities, procedures, plans, policies, and products;
- logic models, beliefs, assumptions, and implicit and explicit program theories explaining why and how programs should work; and
- outputs, results, benefits, outcomes, and impacts.

Our full description of programs includes the key elements discussed above. We describe programs as the systematic application of resources guided by logic, beliefs, and assumptions identifying human needs and factors related to them. Defined completely, a program is

- a set of planned systematic activities
- using managed resources
- to achieve specified goals
- related to specific needs
- of specific, identified, participating human individuals or groups
- in specific contexts
- resulting in documentable outputs, outcomes, and impacts
- following assumed (explicit or implicit) systems of beliefs (diagnostic, causal, intervention, and implementation theories about how the program works)
- with specific, investigable costs and benefits.

These standards are not designed to be equally applicable to all programs. Rather, they are tailored for educational, human services, human resource development, health, wellness, and other types of programs in which the goals include changes in human motivation, attitudes, knowledge, skills, and performance.

❖ **WHAT IS *EVALUATION*?**

Earlier editions of *The Program Evaluation Standards* defined *evaluation* as the “systematic investigation of the worth or merit of an object”

(JCSEE, 1994, p. 3). The “object” in this case is the program under review. While maintaining the core focus on systematic investigation and quality, more recent definitions of *evaluation* have expanded the number of terms that denote value to include *merit*, *worth*, *importance*, and *significance*. Other recent definitions focus more on active purposes, such as judgment, decision making, improvement, and accountability assessment. Sometimes evaluation is described by its position in the program development chronology; for example, as the judgment of quality made when a program is completed (a *summative* purpose) or while the program is still developing (a *formative* purpose).

In the third edition, we expand the descriptive definition of program *evaluation* to include

- the systematic investigation of the quality of programs, projects, subprograms, subprojects, and/or any of their components or elements, together or singly
- for purposes of decision making, judgments, conclusions, findings, new knowledge, organizational development, and capacity building in response to the needs of identified stakeholders
- leading to improvement and/or accountability in the users’ programs and systems
- ultimately contributing to organizational or social value.

The first component in the definition above, systematic investigation, requires some additional discussion because it so often leads to confusion about how systematic evaluation compares and contrasts with research.

Unlike most program evaluation, most educational and social science research focuses primarily on the need to produce credible, generalizable knowledge about the nature of the world around us. Typically, research questions do not address the quality of a specific program but rather gaps and uncertainties in important theories or areas of knowledge. A researcher typically designs a research project to study these uncertainties and gaps and strives for complete control of how the research, including any interventions or treatments, is performed. The primary audience for most research is the researchers’ peers. A secondary audience consists of those who might use this generalized knowledge or translate it into practice.

In contrast, program evaluations help stakeholders answer specific questions or make decisions about specific programs and their components. Evaluators use their professional skills to investigate such things as a program’s development, processes, theory, viability, outcomes, and impact. They use their skills in the service of stakeholders’ needs and to investigate questions about value to specific stakeholders.

Distinguishing research from evaluation projects based on their methodologies is difficult. For example, sometimes evaluations of standardized interventions are conducted using controlled experimental or quasiexperimental designs with randomized control groups or naturally occurring comparison groups. In these approaches, the evaluation manager strives for considerable control over the intervention and manages both the quality of the intervention and its evaluation. Many evaluators with a background in experimental research prefer these approaches because they help isolate the net impact of the program as an intervention. However, not all programs are simple enough or sufficiently controlled and standardized, especially in complex naturally occurring contexts. Even when programs are well-developed and amenable to standardization, implementing an evaluation component can alter the programs, their contexts, and their outcomes in significant and unpredictable ways. This direct evaluation impact on the program may be positive and result in program improvements. However, direct evaluation impact interferes with generalization that research aspires to. In particular, findings about programs directly affected by evaluation processes do not necessarily generalize to replications of the program without similar evaluation components.

Similarly to experimentalists, researchers trained in naturalism, ethnography, and qualitative designs may set about to understand a program in order to increase knowledge about how and why it functions or does not function in its larger context. When such projects focus more on generating dependable knowledge and less on the quality of the program and its value to its stakeholders, they are more akin to research studies than program evaluation.

Sometimes, those who commission evaluations believe that evaluative inquiry and research inquiry are one and the same. With regard to the available tools, instruments, and methodologies, evaluation and research projects do share many similarities. However, they are very different in how they are planned and managed, and in how they generate value. In many program evaluation situations, serious difficulties arise because those in charge do not adequately address these differences. The remaining chapters in this book provide extensive discussion about how to plan, implement, and improve program evaluations. We believe this book is required reading for program or evaluation stakeholders who confuse social science research with program evaluation. For example, Part I, describing evaluation utility, details how evaluation creates value and how orientation to stakeholders' needs is an important defining feature of evaluation as opposed to research. Part II presents the factors affecting evaluation feasibility and also provides important detail about the relationships between program evaluations and the programs they evaluate. Part III, on evaluation propriety, discusses the intense

social and human relationships that are the hallmark of programs and their evaluations and the need for attention to and respect for social and human factors, especially customs, cultures, laws, ethics, morality, and human rights. Part IV, on evaluation accuracy, clarifies how social science methods are used in evaluation projects to create value. It describes how evaluation projects rely on methods from a wide variety of sources to accurately address the needs of evaluation users.

One last consideration about how to describe and define program evaluation quality is presented in Part V. Evaluation projects and subprojects can themselves be documented and evaluated so that they can be improved and held accountable. Such “evaluations of evaluations” are called *metaevaluations*. Just as program evaluation is often confused with social science research, *metaevaluation* is often confused with *meta-analysis*. *Meta-analysis* refers to the practice of reviewing multiple research studies addressing the same phenomena to draw the most supportable generalizable conclusion based on the quality of the research studies. As a primary way of investigating evaluation quality, *metaevaluation* is discussed more completely in Part V, *Evaluation Accountability*. In contrast, meta-analysis, as a specific research technique, is not addressed in this book.

❖ CONTENTS OF THIS BOOK

The chapters in this book are designed to respond to a wide variety of users and their roles. *Applying the Standards*, the section following this introduction, continues with some of the themes introduced here and describes in greater detail factors that will contribute to optimal use of the standards. Subsequent chapters present the attributes of high-quality evaluations and the standards that support achieving evaluation quality. Sections at the end of the book present a glossary, references that serve to document the specific attributes of quality, and an index. Standard statements from previous editions are included in Appendixes A and B.

❖ THE CORE CHAPTERS ON EVALUATION QUALITY: UTILITY, FEASIBILITY, PROPRIETY, ACCURACY, AND ACCOUNTABILITY

The goal of this book is to help users recognize and improve evaluation quality. Each of Parts I through V presents one attribute of quality and its accompanying standards. After an introductory overview to the attribute and the introduction of a case scenario, chapters on individual standards provide

clarification and rationales, specific considerations for implementation, and some hazards to try to avoid. To help illustrate how standards are used, the chapter on each standard concludes with an application of the case scenario. Taken all together, the applications tell the story of how evaluation quality is constructed in specific situations.

A major attribute of evaluation quality is utility, addressed extensively in Part I. *Utility* discusses use, usefulness, influence, and misuse. It describes when and how evaluation worth is created, for example, when evaluations contribute to stakeholders' learning, inform decisions, improve understanding, lead to improvements, or provide information for accountability judgments. Utility is supported by eight standards.

Part II presents the factors affecting evaluation feasibility. *Feasibility* discusses the effects of contexts, cultures, costs, politics, power, available resources, and other factors on evaluations. It details which feasibility factors to consider before implementing an evaluation and how to increase or maintain feasibility in different contexts. Feasibility is supported by four standards.

Part III describes propriety: the moral, ethical, and legal concerns related to evaluation quality. *Propriety* considers the rights of stakeholders and other persons and details the responsibilities of all stakeholders, especially evaluation professionals, in an evaluation. Setting thresholds for adequate propriety can be difficult and requires balancing different stakeholders' needs and situations. Social justice considerations may play a significant role, but not all stakeholders have the same values or conceptions of social justice. Propriety is supported by seven standards.

Part IV discusses how to increase the accuracy of findings and conclusions. *Accuracy* discusses reliability, validity, and reduction of error and bias. The eight accuracy standards address quality in data collection, analysis, logic, conclusions, and communication. The chapter is intended for all stakeholders and does not require prior technical skills in statistics, measurement, ethnography, methodology, or research. Accuracy is supported by eight standards.

Part V is called *Evaluation Accountability*. This encompassing attribute of evaluation quality results from balancing utility, feasibility, propriety, and accuracy. Internal and external metaevaluation, discussed extensively in the accountability standards, provide the methodology used to increase and document evaluation quality. Evaluation Accountability introduces three new standards but is also supported by the 27 specific utility, feasibility, propriety, and accuracy standards.

❖ HOW TO CONTRIBUTE TO CURRENT DISCUSSIONS AND FUTURE REVISIONS

Stakeholder contributions are critical to the continued development of the *The Program Evaluation Standards*. The JCSEE is committed to collecting and responding to feedback from those who use or are affected by the use of these standards. We encourage all stakeholders to try out this edition in any situations that seem appropriate. We need to know whether and how the features and content work for you in your evaluation roles. Interested stakeholders can find out more about how to get involved in standards research and development by contacting us at <http://www.jcsee.org>. We are already planning for the next edition and invite you to be an important part of that process as this current edition is used, evaluated, and revised. Updates will be posted regularly at [jcsee.org](http://www.jcsee.org).

❖ REFERENCES

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.