The 2011 Program Evaluation Standards: a framework for quality in medical education programme evaluations

Valerie Ruhe PhD1 and J. Donald Boudreau MD FRCPC2

1Consultant, Educational Program Evaluation, Sechelt, BC, Canada
2Director, Office of Physicianship Curriculum Development, Faculty of Medicine, McGill University, Montreal, QC, Canada

Keywords
medical education, programme evaluation, quality, standards

Correspondence
Dr Valerie Ruhe
Educational Program Evaluation
201-5470 Inlet Avenue
Sechelt, BC
Canada V0N 3A3
E-mail: ruhex001@hotmail.com

Accepted for publication: 3 May 2012
doi:10.1111/j.1365-2753.2012.01879.x

Abstract

Rationale, aims and objectives Based on input from 400 stakeholders over 6 years, the 2011 Program Evaluation Standards represents an in-depth analysis of values, meaning and measurement and their relationships in programme evaluation. Evaluation quality is achieved by balancing five attributes: utility, feasibility, propriety, accuracy and evaluation accountability. These attributes are used to organize 30 standards, 200 strategies and 197 hazards.

Method In response to a call from the authors of the standards, we have used them to guide our meta-evaluation of McGill’s undergraduate physicianship programme.

Results Our findings show how the standards illuminate the tensions, dilemmas and hazards inherent in all stages of programme evaluation studies and offer helpful strategies for designing and conducting high-quality evaluation studies.

Conclusions Based on our experience, the third edition needs to be used as a reference document in all stages of evaluations of medical education programmes.

Introduction

Rationale, aims and objectives

Evaluating the quality of medical education is of substantive importance. To maintain its status as a self-regulating profession, medicine relies on public trust [1]. Programme evaluation is a mechanism for public accountability. ‘Evaluation’ is the systematic investigation of programme merit and worth, but evaluation purposes can also be targeted to specific information needs. Evaluations should be legal, ethical, effective and in the public interest [2]. High-quality evaluation studies enhance the credibility of evaluation findings and the likelihood that they will be used to benefit society. But unlike research, evaluation studies are shaped by context and stakeholders’ values, perspectives and needs, as well as by the purposes and intended uses of evaluation findings.

The Joint Committee on Standards for Educational Evaluation (JCSEE) was founded in 1975 as a coalition of major professional associations concerned with the quality of evaluation studies of US government-funded health care and social services programmes. The only body accredited by the American National Standards Institute (ANSI) to set standards for educational evaluation in the United States, the JCSEE developed two sets of ANSI-based standards in 1981 and 1994. These two editions were important steps in the professionalization of programme evaluation [2].

The Program Evaluation Standards (third edition) was published in the fall of 2011 and is the first revision in 18 years [3]. It provides an authoritative guide to designing, implementing and conducting high-quality evaluation studies of educational and training programmes in the United States and Canada. The third edition reflects the perspectives of 400 evaluators and clients involved in needs assessments, national and international reviews, field trials, national hearings and literature reviews over 6 years.

The standards are applicable to all educational programme evaluations from faculty development workshops to undergraduate medical programmes. The JCSEE recommends that clients and evaluators try using the standards and share their reflections. In response to this call, we will (1) discuss the role of values in evaluation, (2) summarize the 2011 standards, (3) show how we used them to evaluate the evaluation of an innovative undergraduate medical programme and (4) present our findings and recommendations to improve evaluation quality. In this paper, the term ‘standards’ refers to the third edition. Unless otherwise indicated, all page numbers are from the third edition.

The role of values in evaluation

Values are judgments about what is good or bad. Values are at the heart of evaluation, and etymologically, the word ‘evaluation’ is derived from ‘value’. Evaluation processes stem from and return to
the formation of values [4]. Values underlie decisions and judgments at all stages of evaluation studies, for example, identifying programme and evaluation goals, selecting methods, wording survey items, coding and interpreting data, setting standards of acceptability and making recommendations [2]. In evaluation, the central question is ‘What is success and from whose point of view?’

Evaluations can inform sound decisions about the future of programmes or programme improvements. When exposing desirable or undesirable consequences, evaluators must consider social values and political contexts. The uses of evaluation findings are informed by the values of stakeholders, that is, those whose lives are affected by programmes. Stakeholder values are central to evaluation, and the optimal use of findings depends on how well they identify with and respect the value perspectives embedded in each stage of the evaluation [3]. Evaluation studies should not be perceived as reflecting primarily the views of those with vested interests [3]. Moreover, the use of findings is optimized when stakeholders’ values, which may be nebulous or implicit, are brought into the open, and when stakeholders have some involvement in evaluation studies [3].

The 2011 Program Evaluation Standards

The third edition constitutes ‘a defensible set of guiding principles’ (p. 9) and a code of professional evaluation practice. Evaluation quality is comprised of five attributes: utility, feasibility, propriety, accuracy and evaluation accountability. These attributes provide a framework for organizing the 30 standards (Appendix).

Utility

Utility is the usefulness of evaluations to clients and other ‘right-to-know’ audiences. The many uses of programme evaluation include justifying resource expenditures; documenting change; determining the programme merits; assisting leaders in policy decisions; making contributions to the scholarly literature; revealing unintended consequences and improving programmes. The eight utility standards are intended to increase the extent to which programme stakeholders find evaluation processes and products valuable in meeting their needs (Appendix).

Feasibility

Feasibility is the extent to which evaluations are efficient and cost-effective. The four feasibility standards deal with the sufficiency of resources, and the suitability of evaluation plans, systems and databases for programme contexts, goals and activities.

Propriety

Propriety is ‘adherence to the highest principles and ideals (including professional ethics, civil law, moral code and contractual agreements)’ (p. 291). The seven standards that support propriety address what is ‘proper, fair, legal, right, acceptable and just in evaluations’ (p. 106). Propriety also encompasses the social and human relationships, for example, stakeholder needs, roles and responsibilities.

Accuracy

Accuracy refers to the truthfulness, validity and dependability of evaluation findings, interpretations and judgments about quality. The eight accuracy standards encompass research designs and methods, data collection, analysis, reliability, validity, error and bias and the logic of conclusions.

Evaluation accountability

A new attribute in the third edition, evaluation accountability refers to the adequate documentation of processes, agreements, data files and reports and an internal and external meta-evaluative perspective focused on improvement and accountability of processes and products. Meta-evaluations are evaluations of evaluation studies. Meta-evaluations can be either ongoing or retrospective. Whether based on systematic assessments or reflections, they involve self-awareness, pausing, openness to monitoring, scrutiny and changing course. They should not be confused with meta-analysis, a method for synthesizing findings from research studies. The importance of meta-evaluation is indicated not only by the evaluation accountability standards, but also by Hanssen, Lawrenz and Duneet’s (2008) study in the American Journal of Evaluation [5]. In sum, meta-evaluations are encouraged not only by the authors of the third edition but also by the evaluation profession.

How can the 2011 standards be used?

The 2011 standards map out the dimensions of evaluation quality and provide comprehensive criteria to judge the quality evaluation studies. The standards offer 200 strategies and warn of 197 hazards that may emerge in evaluation processes and products. Although the term ‘hazard’ is not defined, the examples given include inappropriate beliefs, poor decisions and risky actions, which could stall or derail the process or weaken the findings.

The standards can be used as evaluations are in progress or after they are competed. As an example of the former, one powerful use is for stakeholders and evaluators to meet to discuss the standards, their own values about programme evaluations, and the evaluation plan, process and products [3]. As for retrospective use, the authors provide three illustrative meta-evaluative case studies: (1) Centerville Health for Eco-Educated Residents, a community-based health promotion programme, (2) HOPE, a community integration project and (3) Teacher Quality Enhancement, a State Department of Education programme. These cases present emergent dilemmas, problems and challenges, and successful decisions, strategies and actions that enhanced evaluation quality.

A few points need to be made. First, the use of the standards is voluntary and consensual; they are not legal or mandated requirements. Second, they should be applied at all stages of evaluation studies: planning, design, implementation, reporting and use of findings. Third, the standards are ‘generic’. Because evaluations take place in unique contexts, there are many different ways to apply the standards. These applications need to be ‘adaptive and responsive to local contexts [and] not applied literally and superficially following a simple recipe’ (p. xxxii). Finally, the JCSEE intends for evaluators to integrate several standards in all quality dimensions, not focus in depth on one or two at the expense of the others.
Method
We used the third edition to guide a systematic meta-evaluation, that is, an evaluation of our evaluation studies of McGill University’s physicianship programme, a set of innovative teaching modules rolled out sequentially throughout the undergraduate medical programme from 2005 to 2009. Because the third edition was not published until after our evaluation study was almost completed, our reflections are necessarily retrospective. By showing what we did, probing which 2011 standards were most applicable, and what we might have done differently, we hope to contribute to improved evaluation practice in a particularly complex pedagogic environment, that is, preclinical and clinical medical education.

Context

Physicianship programme, Faculty of Medicine, McGill University
McGill’s physicianship programme is an integrated and mandatory component present in all 4 years of the medical school’s undergraduate curriculum. The programme goals are to teach the healer and professional roles. Physicianship includes a specific cognitive base and is operationalized as a set of patient-centred values and skills such as role identification, clinical observation, attentive listening, narrative understanding and professionalism. A longitudinal mentorship, called physician apprenticeship, is a flagship course within the programme. It aims to promote reflection on the students’ transition from laymen to physicians and provide a safe environment for them to explore and attend to enculturation in medical school. The apprenticeship is structured to provide sentinel opportunities for guided reflection in various settings: small group meetings with student peers and near peers; visits to palliative care and chronic care units; one-on-one meetings; online platforms and learning portfolios. Details of the physicianship programme and the physician apprenticeship course have been published elsewhere [6,7].

Programme stakeholders
The physicianship stakeholder groups are: (1) students enrolled in the programme, (2) candidates applying to the medical programme, (3) residents and alumni, (4) faculty and clinician educators, (5) McGill administrators such as dean and associate dean, (6) the Admissions Office, (7) the community at large (the island of Montreal), (8) patients and (9) the evaluation team. The faculty members who participated in this course, as mentors and (hoped-for) role models, were all clinicians experienced in teaching medical students and with affiliations to university-based clinical sites. The evaluation team consisted of three faculty leaders in curricular development and three educational evaluation specialists.

Physicianship programme evaluation plan
We conducted a series of physicianship evaluation studies based on a comprehensive mixed methods design. The primary purpose was to develop an in-depth understanding of the unfolding new curriculum process and to determine how well students internalized physicianship values, beliefs and concepts. A secondary objective was to demonstrate accountability to clinical teachers, thereby sustaining enthusiasm and ‘buy-in’ for curricular renewal over 4 years. Other purposes were to justify resource expenditures; document change; determine the merit of a programme; assist leaders in policy decisions; inform research and reveal unintended consequences. To achieve these goals, we designed an evaluation planning chart to lay out our hypotheses, indicators and methods and facilitate communication and oversight. An excerpt of this chart is in Table 1.

For reporting purposes, the team considered using different evaluation models to organize the findings: Kirkpatrick’s (1994) four-level model [9], logic models and participatory evaluation. We adopted Stufflebeam’s (2003) Context-Input-Process-Product (CIPP) [10] model because of its prior successful application in the Office of Faculty Development [11]. CIPP is a comprehensive ‘roadmap’ that helps identify programme components and the values underpinning evaluations [2]. CIPP channelled our discussions on the programme’s conceptual foundation, clarified the curricular process and measurement of impacts, and organized the findings and recommendations. By emphasizing clarification over justification, CIPP helped us better understand the programme.

Results
The results of our meta-evaluation of physicianship are in Table 2.

Conclusions

Lessons learned
Table 2 shows how we used the 2011 standards to guide our retrospective evaluation of the Physicianship preclinical and clinical educational programme evaluations. This meta-evaluation enabled us to see the five dimensions of evaluation quality ‘in action’, that is, how they played themselves out in practice at various stages. As the authors of the third edition said, we found that meta-evaluation guided by the standards can illuminate various values, issues, dilemmas, tensions and hazards commonly encountered in evaluation studies (columns 2 and 3). During our evaluations, some of these difficulties were hidden from, or in the background of, our conscious awareness. However, our meta-evaluation brought them into the foreground where they were clearly visible. The ‘recommendations’ in column 4 show different decisions and actions we might have taken if we had used the standards as a reference document. We hope that sharing our insights will help build capability for future evaluation studies at McGill and convince other medical educators and evaluators of the advantages of using the standards to guide their own evaluations.

Accuracy
The accuracy standards shone a light on our evaluation design, which was initially experimental but evolved with information needs towards qualitative and mixed methods. In year 3, qualitative data captured student and faculty experiences and programme strengths and weaknesses. In hindsight, our initial design could have been cyclical, with findings used to inform programme
changes, which would also be evaluated. At times, our design was cyclical, for example, we eliminated student portfolios in response to student satisfaction scores. But we could have gone further, for example, interviewing francophone students to see how terminology not easily translated into French, for example, ‘healer’, impacted their understanding and acceptance of Physicianship values and concepts and whether any changes were needed to meet their needs [13].

Utility

There were several ways in which the physicianship evaluation findings were used. First, we developed a much richer understanding of the meaning of physicianship, of the programme [6,14] and of the evaluation process. Our findings informed research [6] and instrumentation for enhancing faculty professionalism [15]. This latter use had not been foreseen in our planning chart but will improve the quality of the clinical experience for students in future.

Trade-offs

Table 2 show that trade-offs across the standards need to be made at each stage. Although involving a broader base of stakeholders, for example, medical faculty, administrators or patients, might have strengthened the accuracy and utility of our findings, we realized that their full participation was not always feasible or practical. Similarly, McGill’s ethics review requirements posed understandable challenges in obtaining the consent and participation of patients, although patient outcomes are important for evaluating the impact of clinical training [16]. Both examples illustrate necessary trade-offs between propriety and feasibility.

Summary

As shown in Table 2, evaluation quality is achieved by balancing utility, feasibility, propriety, accuracy and evaluation accountability. The standards illuminate the interplay between facts and values, knowledge and practice and between epistemology, ethics and politics. They shine a light on quality at every stage of evaluation studies from planning to implementation and use. Meta-evaluation is an iterative process of pausing, taking stock, restarting, reviewing and revising, a process called ‘conscientization’ [17]. By knowing reality, we can better transform it. If we had used the standards throughout our evaluations, we would have been conscious of, and alerted to, these tensions, hazards and trade-offs and adopted some of the helpful strategies offered as the evaluations were in process.

In the end, our series of studies enabled us to use mixed methods to build an argument based on multiple sources of evidence about programme merit and worth. Our argument rested on both evidence (facts) and values, that is, our claim that the relevance of the programme in teaching 21st century patient-centred skills is an important dimension of programme merit and worth [12]. The lessons we have learned through conducting this retrospective review – a meta-evaluation – have provided some new understandings that will hopefully be meaningful to our colleagues, both in the local context (at McGill) and in the broader community (of medical educators and evaluators). Based on our
<table>
<thead>
<tr>
<th>Stages of our programme evaluation (2005 to 2010)</th>
<th>Applicable standards*</th>
<th>Tensions, dilemmas and hazards</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To evaluate physicianship, we designed a series of evaluation studies.</td>
<td>Utility, Feasibility, Propriety, Accuracy, Evaluation Accountability</td>
<td>Our evaluations were guided by the CIPP model, McGill’s ethics review requirements and our Evaluation Planning Chart.</td>
<td>Use the third edition to inform the planning chart and activities in all evaluation stages.</td>
</tr>
<tr>
<td>A team of two programme administrators, including JD Boudreau, Director, Office of Curriculum Development (author) and three programme evaluators, including Valerie Ruhe (author), designed and conducted the evaluations.</td>
<td>Propriety P6 Conflict of Interest?</td>
<td>A broad base of stakeholders elicits diverse perspectives and generates buy-in for the study and commitment to using the findings for programme improvement. Given their workloads and professional demands, how could more key stakeholders have been involved?</td>
<td>Encourage influential stakeholders (e.g. directors of physicianship modules) to read a summary of the third edition and meet to discuss applications.</td>
</tr>
<tr>
<td>Our planning chart did not include a formal agreement about the roles of evaluators and stakeholders.</td>
<td>Propriety P2 Formal Agreements</td>
<td>Stakeholder roles, e.g. data collection, analysis, interpretation and/or reporting can range from minimal to considerable. Early in the planning stage, negotiate, clarify and make explicit the roles of stakeholders and evaluators. Revise later if needed.</td>
<td>Use mixed methods to build an argument about the strengths and weaknesses of the programme. In the absence of quantitative differences, qualitative data can provide valuable insights into programme strengths and weaknesses. Be cautious in adopting tools validated in other circumstances and contexts.</td>
</tr>
<tr>
<td>We conducted a quasi-experimental, pre/post-design to investigate changes in students’ values, beliefs and attitudes.</td>
<td>Accuracy A6 Sound Design and Analysis</td>
<td>Given the difficulty in controlling for intervening variables, experimental studies of educational interventions often do not yield statistically significant differences [2].</td>
<td></td>
</tr>
<tr>
<td>Implementation stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We did not always meet our project management timelines.</td>
<td>F1 Project Management U7 Timely and Appropriate Communicating and Reporting</td>
<td>Pressuring faculty or students to participate in a timely manner would have discouraged buy-in and support for evaluations.</td>
<td>Revise the timelines as needed. Hire a project manager or provide training in project management.</td>
</tr>
<tr>
<td>Our evaluation plan had two studies of undergraduate students’ perspectives, but no study of patient perspectives.</td>
<td>P1 Responsive and Inclusive Orientation F3 Contextual Viability</td>
<td>Because physicianship is enacted in clinical encounters, patients were deemed to be critical stakeholders. Patient satisfaction data could have supported an argument for the social significance of the programme. We had some baseline metrics, but ethics requirements and time constraints were barriers to more extensive patient involvement.</td>
<td>We might have been more diligent in making inquiries about avenues of collaboration with the public relations offices of local academic health centres. We found out serendipitously that some of the hospitals had strategies like the ‘Guardian Angel Initiative’ to invite patients to identify health care providers with exemplary interpersonal skills.</td>
</tr>
<tr>
<td>Our evaluation plan was mostly linear, not iterative.</td>
<td>Utility U5 Relevant Information</td>
<td>An iterative evaluation cycle may not be feasible because of limited resources and faculty workloads.</td>
<td>Be responsive in adapting the evaluation plan to local changes. When feasible, use findings formatively and document actions taken.</td>
</tr>
<tr>
<td>Stages of our programme evaluation (2005 to 2010)</td>
<td>Applicable standards*</td>
<td>Tensions, dilemmas and hazards</td>
<td>Recommendations</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>------------------------</td>
<td>--------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>We did not sufficiently articulate or emphasize the programme’s social significance.</td>
<td>Utility U4 Explicit Values</td>
<td>Patient-centred communication skills are important 21st century skills that help medicine to serve the public and maintain public trust [12].</td>
<td>Early on, articulate the programme’s theoretical rationale and social significance as important dimensions of the programme’s merit and worth.</td>
</tr>
<tr>
<td>Reporting stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We used the CIPP model, a comprehensive approach to organizing findings for reporting purposes.</td>
<td>Accuracy A6 Sound Design and Analysis</td>
<td>In medical education, the CIPP (Context, Input, Process, Product) distinctions are not always clear-cut.</td>
<td>Use caution in selecting and applying programme evaluation or logic models so that they fit the evaluation purpose and information needs. Appropriate use of carefully considered models should make evaluation findings easier to interpret, organize and report.</td>
</tr>
<tr>
<td>Students were not regularly informed of evaluation findings.</td>
<td>Propriety P5 Transparency U8 Concern for Consequences and Influence</td>
<td>We felt that transparency needed to be balanced against any misuse of findings and unintended consequences.</td>
<td>Periodically distribute a one-page summary of evaluation findings to students.</td>
</tr>
<tr>
<td>Use of findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The evaluation design was comprehensive and reflected multiple purposes and uses.</td>
<td>Utility U6 Meaningful Processes and Products</td>
<td>Over time, it became clear that our series of studies was ambitious.</td>
<td>Negotiate and write clear statements of resource availability in the evaluation plan. Track and report all uses made of the evaluation findings.</td>
</tr>
<tr>
<td>Meta-evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We meta-evaluated near the end of our study.</td>
<td>Evaluation Accountability E2 Internal Meta-evaluation E3 External Meta-evaluation</td>
<td>We did not meta-evaluate until after our evaluations were almost over because the 2011 Standards had been under revision.</td>
<td>Use the standards to conduct periodic, formative meta-evaluations while evaluations are in progress. Use evaluator journaling, professional development workshops, stakeholder meetings and/or surveys that ask stakeholders to rate the evaluation process by indicating whether particular standards were ‘fully addressed’, ‘partially addressed’ or ‘not addressed’ at different stages of the evaluation.</td>
</tr>
</tbody>
</table>

*The standards in column 2 are examples for illustrative purposes. For a complete list at each stage, see the Functional Table of Standards (p. xli).
experience, we encourage medical educators to study and use the third edition as a critical reference document in all stages of evaluation studies.

Acknowledgements

Funding was provided by the Faculty of Medicine, its benefactors, alumni (class of 1955) and the Max Bell Foundation. The Faculty is grateful to the Max Bell Foundation for its generous financial support. Dr. Boudreau is especially beholden to the Arnold P. Gold Foundation.

Author bios

Valerie Ruhe, PhD, is a consultant in educational test validity and programme evaluation. She has 13 years of programme evaluation experience in distance education, K-12 and higher education in Canada and the United States. She has published several articles on test validity and programme evaluation, and a book entitled Evaluation in Distance Education and E-Learning: The Unfolding Model.

J. Donald Boudreau, MD, is the Director of the Office of Physician Curriculum Development, Faculty of Medicine, McGill University and an Arnold P. Gold Foundation Associate Professor of Medicine. He has published extensively in medical education.

References


Appendix

The 2011 Program Evaluation Standards (third edn)

Utility standards

U1 Evaluator Credibility Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.

U2 Attention to Stakeholders Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.

U3 Negotiated Purposes Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.

U4 Explicit Values Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.

U5 Relevant Information Evaluation information should serve the identified and emergent needs of stakeholders.

U6 Meaningful Processes and Products Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.

U7 Timely and Appropriate Communicating and Reporting Evaluations should attend to the continuing information needs of their multiple audiences.

U8 Concern for Consequences and Influence Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Feasibility standards

F1 Project Management Evaluations should use effective project management strategies.

F2 Practical Procedures Evaluation procedures should be practical and responsive to the way the program operates.

F3 Contextual Viability Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.

F4 Resource Use Evaluations should use resources effectively and efficiently.
Propriety standards

P1 Responsive and Inclusive Orientation Evaluations should be responsive to stakeholders and their communities.

P2 Formal Agreements Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.

P3 Human Rights and Respect Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.

P4 Clarity and Fairness Evaluations should be understandable and fair in addressing stakeholder needs and purposes.

P5 Transparency and Disclosure Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.

P6 Conflicts of Interests Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.

P7 Fiscal Responsibility Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

Accuracy standards

A1 Justified Conclusions and Decisions Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.

A2 Valid Information Evaluation information should serve the intended purposes and support valid interpretations.

A3 Reliable Information Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.

A4 Explicit Program and Context Descriptions Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.

A5 Information Management Evaluations should employ systematic information collection, review, verification, and storage methods.

A6 Sound Designs and Analyses Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.

A7 Explicit Evaluation Reasoning Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.

A8 Communication and Reporting Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

Evaluation accountability standards

E1 Evaluation Documentation Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

E2 Internal Meta-evaluation Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.

E3 External Meta-evaluation Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards.

Source: JCSEE Website: http://www.eval.org/evaluationdocuments/progeval.html
Note: The standard names and statements are reproduced with the written permission of Dr. Donald B. Yarbrough, Director, University of Iowa Center for Evaluation and Assessment.