

---

# Concluding Thoughts and Reflections on the Special Issue on Program Evaluation Standards

*Journal of MultiDisciplinary Evaluation*  
Volume 19, Issue 43, 2023

**JMDE**  
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180  
<http://www.jmde.com>

Arthur E. Hernández

*Co-Editor, JCSEE Executive Committee, and Representative of the American Evaluation Association*

**Abstract:** This paper seeks to address the issue of the relevance of the current edition of *The Program Evaluation Standards*, published by the Joint Committee on Standards for Educational Evaluation. The paper addresses concerns related to the nature of the standards, their current applicability to practice, their comprehensiveness and completeness. The presented conclusions are that the standards are applicable, relevant, complete, and comprehensive.

**Author's Note:** Correspondence concerning this article should be addressed to Arthur Hernandez, Dreeben School of Education, University of the Incarnate Word, 4301 Broadway St. CPO 293, San Antonio, TX 78209, United States. Email: [aeherna8@uiwtx.edu](mailto:aeherna8@uiwtx.edu)

**Keywords:** *Program Evaluation, Standards, JCSEE, PES*

---

This special issue of *The Journal of MultiDisciplinary Evaluation (JMDE)* focuses on the program evaluation standards (PES) of the Joint Committee on Standards for Educational Evaluation (JCSEE). The JCSEE was established in 1975 to develop and promote the standards and highlight their use. This is important for at least two reasons. First, the standards are promulgated by the JCSEE, which may lead some to believe they pertain solely to evaluation in *educational settings or for educational purposes*, which is not the case. From the onset, with the first edition, the standards easily and obviously pertained to program evaluation efforts across the spectrum of applications, populations, and purposes. This intentionality is continued in *The Program Evaluation Standards*, third edition—as stated in the preface, it “provides an integrated guide for evaluating programs that foster human learning and performance across the lifespan” (Yarbrough et al., 2011, p. xii). The intentionality is further illustrated by *The Program Evaluation Standards*’s consideration of the object of evaluation. For example, it describes redefining and clarifying “project” and “program” and the many possible components of activity which may be subject to evaluation, and specifies particular “key elements” which are defining attributes of evaluable activity. While arguably not comprehensive in the sense of containing every possible permutation, nuance, or possible inclusion, the definition/description of object and action in the PES is conceptually sound, conveying key ideas for understanding and recognizing the purpose, focus, and activity of program evaluation and its grounding for evaluator and stakeholder alike. Finally, the PES provides a descriptive definition of program evaluation that clarifies the action and, among other things, provides a basis for the distinction between evaluation and other forms of inquiry:

- The systemic 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 improvements and/or accountability in the users’ programs and systems
- Ultimately contributing to organizational or social value. (p. xxv)

The breadth of this definition and description of scope and its usefulness is well illustrated by the application of the PES to the development and dissemination of an evaluation toolkit (see “The Integration of the Program Evaluation Standards into an Evaluation Toolkit for a Transformative Model of Care for Mental Health Service Delivery,” by Snow, Berger, Jaouich, Hood, and Salmon, in this special issue). The other contents of this special issue provide a basis for considering the PES in application to just a few of the many possible applications for different purposes and how professional evaluators can and do conceptualize and utilize the standards in regards to program evaluation broadly considered.

The second has to do with the fact that the current edition of the PES was published over 10 years ago, and some individuals may wonder whether they have become dated or even obsolete. The collected papers of this special issue all suggest that this is not the case (see “The Aspirations, Challenges, and Influence of *The Program Evaluation Standards*, Third Edition” by Egelson in this volume for one example). In fact, it is clear the entire collection of contributions included certainly support the applicability of the standards to current evaluation practice (even though, admittedly, some of the case studies / examples included in the current edition of the PES may be considered dated by some readers). The contents of this *JMDE* special issue suggest the ongoing applicability and meaningfulness of the standards despite their age and suggests the standards themselves are quite robust, appropriate, and durable. This should not be surprising, given that the standards were never intended to function as a “checklist” of practices which would “guarantee” sound, accurate, and meaningful evaluation. Instead, the standards were intended as a conceptual framework, requiring familiarity, adaptation, responsiveness, mindfulness, and consideration of purpose and use in their application. The capacity of the standards to describe criteria for evaluation quality results from the fact that the third edition of the standards was based on *strong conceptual and philosophical foundations* which were decided on the basis of input from theoreticians and practitioners, review of the literature related to the conceptual foundations of evaluation, and the judgment and interpretation of all these sources of information by the principal authors and members of the JCSEE, who represent a wide range of organizations and significant expertise, experience, and diversity related to the conceptualization of evaluation research, practice, and theory. This depth and breadth of gathered and considered information

and the mechanisms and representation of evaluators involved in that consideration and interpretation provides for strong confidence in the meaningfulness and durability of the result (see “‘Best Tradition’: CREATE, JCSEE and the Program Evaluation Standards,” by Klinger and Klinger, in this volume). In fact, the checks and balances provided by the consideration and proof of so many involved, and the ongoing demonstrations of the standards’ use and value (including the articles in this special issue) provide a strong basis for the present, ongoing, and current “validity” of the standards.

While a detailed description of the standards is not possible here, no consideration of the usefulness and relevance of the standards is possible without at least a brief review of the domains or attributes of quality presented in the PES (which are elaborated by related standards statements in the text). **Utility** pertains to whether the evaluation (including process use) meets the needs and rights of the full range of stakeholders in consideration of evaluation use. This attribute of evaluation quality is elaborated in terms of use, usefulness, influence, misuse, consequences, and judgments concerning evaluation worth. **Feasibility** addresses the effects of context, including culture, politics, power, and resource availability. The Feasibility domain is an important guide to decision making regarding whether to conduct an evaluation in a given situation, for a given purpose, or with a particular population based on, among other things, consideration of “cultural and political interests and needs of particular groups” (p. 71). **Propriety** deals with the ethical, moral, regulatory, and legal requirements and considerations involved in the conduct of an evaluation, including the rights of stakeholders and other persons impacted by the evaluation process or outcomes, and involves judgments concerning social justice and intended and unintended consequences for those involved and for society as a whole. Propriety involves evaluator judgment concerning the “rightness” of all aspects of the work—from deciding to do an evaluation, to how the various stakeholders (including all relevant communities) will be involved and affected—and emphasizes the fundamental rights of all. The **Accuracy** domain deals with minimizing error (including bias) and concerns the nature of information and processes for accurate analysis. While accuracy in evaluation (or any other form of inquiry) seems to be obvious, it is clear from the standards that the expectation is that evaluators will not simply apply “standard” methods of information gathering and analysis that are

formulaic, usual, or even a matter of consensus, without due consideration of the implications for accuracy and relation to the true experience of those involved every time. Finally, **Evaluation Accountability** results from “balancing” the other domains and seeks to provide insight into considerations and actions for demonstrating and documenting evaluation quality. This last standard suggests that evaluators have the responsibility always to consider the process and products of their work in terms of adherence to all the other standards to assure the highest possible quality of work. While these five domains/attributes of program evaluation may be differentially important (weighted) for a given evaluation exercise, they are all always applicable (see “Program Evaluation Standards for Utility Facilitate Stakeholder Internalization of Evaluative Thinking in the West Virginia Clinical Translational Science Institute,” by Curtis, Roy, Lewis, Dooty, and Mikkalik, in this special issue). These attributes/domains are elaborated in the PES to include important constituents and areas of consideration for understanding the five key attributes of evaluation quality and for mindful adaptation for use with given contexts, conditions, and communities. While the groupings are defined, elaborated, and delineated according to each designated attribute, the current edition of the standards also provides an alternative framework or structure for the elements of each for easier consideration and use, by also organizing each of the identified group components around key evaluation tasks: deciding whether to evaluate; negotiating and formalizing agreements, contracts, and budgets; determining who will evaluate; negotiating and developing evaluation purposes and questions; describing the program; designing the evaluation; managing the evaluation; collecting information; analyzing information; and communicating and reporting. While not meant as a road map to practice, this practical organization is quite useful for practitioners, commissioners, and consumers of evaluation alike. It is perhaps this organization which can best help users of the standards make sense of them in practice and apply them to evaluation activity decision-making from planning to execution to reporting. Taken together these concepts and their organized framework provide a strong, meaningful, and appropriate perspective from which to inform evaluation program planning and development and to judge program evaluation quality.

“Good” evaluation requires inclusion of stakeholder perspectives, priorities, and values in its design, which often requires that evaluators serve first as educators. The program evaluation

standards provide a useful framework and springboard for stakeholder consideration of the various aspects, requirements, uses, etc., of evaluation, and can serve as a mechanism for the development of stakeholder sophistication regarding evaluation, which can elevate and maximize the value of the process and product, including its meaningfulness, cost effectiveness, utility, etc. The standards can also serve to provide criteria against which evaluation processes and products can be held accountable. The use of the standards provides a perspective from which all stakeholders can judge. This accountability is first and foremost to those who will be impacted by the evaluation: funders, communities, policymakers, and so on, but included also is the community of the profession since the credibility and usefulness of the profession's efforts are influenced and impacted by the practice of each of its members. Evaluation is a matter of systematic appraisal of value. The determination/identification and demonstration of the essential elements, activity, and relationships within any system is essential to the "definition" of that system and thus to judgments concerning its ethics, rigor, accuracy, "correctness," credibility, "value added," and so forth. This is what the standards provide to evaluation practice broadly considered (see "How Should Program Evaluation Standards Inform the Use of Cost-Benefit Analysis in Evaluation?" by King and "What Do We Know about How the Program Evaluation Standards are Used in Public Health?" by MacDonald, Castelin, George, and Joseph in this special issue).

The "quality" of evaluation may be judged as a matter of the degree to which the activity conforms to expectations based on standards or guidelines of practice. The PES provide these criteria, addressing important and essential questions of the evaluation exercise, including decision-making regarding whether to evaluate; who should conduct the evaluation; what will be evaluated and how; what resources will be required and whether that investment is worthwhile; how the evaluation results will be used and shared; how the evaluation is to be conducted and managed; and how meaning, implications, and conclusions will be determined/derived, among other things; and whether these decisions reflect and respect the perspectives, needs, and expectations—and conform to the requirements of the philosophical foundations—of all involved.

Perhaps some of the criticism directed at the PES stems from a lack of familiarity with the standards. Research presented in this volume (see "Using Dissemination Research Approaches to Understand the Awareness, Adoption, and Use of

the Program Evaluation Standards," by Morrison and Cunningham) suggests a general lack of sophisticated understanding of the standards' history and development, along with limited references in the academic literature to the standards (see "The Program Evaluation Standards in Evaluation Scholarship and Practice," by Watts, Castillo, Akwetey, and Pham), the consequences of which may be exacerbated by the standards' date of publication and the constrained and waning efforts at their dissemination. However, the same research suggests that practitioners are using the standards in some way in the consideration, conduct, and communication of their evaluation efforts. This latter finding is certainly heartening and suggests that while the standards are very infrequently the focus of scholarship and equally unlikely to be referenced in academic publications related to evaluation, *the practice of evaluation* is guided by—and judgments of quality are based on—the tenets of *The Program Evaluation Standards*.

Finally, this discussion and special issue would be lacking if they didn't address at least two of the more recent concerns raised related to the applicability of the standards, those being (1) the consideration of so-called "virtual" evaluations—that is, evaluative efforts which depend on remote (i.e., videoconferencing) contact and interactions; and (2) the lack of sufficient emphasis on what is commonly referred to as culturally responsive and equitable evaluation.

During the COVID-19 pandemic in 2020 and the associated public health imperatives, including social distancing and work-from-home mandates, videoconferencing greatly facilitated the continuation of work across multiple sectors. Almost overnight, in response to these pandemic prevention measures, gatherings of the public of any size were severely constrained, and thus the usual activity associated with evaluation (as for so many other activities and work) was doubtless affected. Notwithstanding these challenges, evaluation activity continued, and thus questions arose concerning the applicability of the standards in a situation/context that their developers, authors, and contributors could hardly have anticipated. However, even a cursory examination of the standards makes clear that they provide appropriate and applicable guidance in this condition of work. For example, standards related to accuracy, propriety, feasibility, and utility come to mind: Valid and Reliable Information (quality of information), Transparency, Project Management, Practicality, Contextual Viability, Resource Use, and Attention to Stakeholders, to name but a few. Clearly, mindful and reflective consideration of the

existing standards can provide a conceptual basis and framework to guide and judge evaluation quality, even in the context of significant social disruption and upheaval such as that caused by the COVID-19 pandemic.

The second concern deals with the raised social consciousness of historical and ongoing racism, which owed its recent genesis to the tragic murder of George Floyd, a Black man killed by police in Minneapolis, Minnesota, during an arrest. The social upheaval that resulted brought to the forefront, for many, the importance of recognizing and responding to the inherent societal and cultural implications of racism, its embeddedness in our social systems, and the consequences of failing to address that racism. These important considerations are recognized and included in the statements of standards which intend to ensure that good, sound, and authentic evaluation practice guard against inherent, implicit, unconscious, and socially pervasive injurious prejudice and racism. It is clear that a thoughtful reflection on the meaning and implications of the standards from the perspective of current knowledge, sensitivity, and community/social self-awareness can serve as a means of avoiding injurious actions, outcomes, and implications. Among some of the standards that may inform this attention and concern are those related to evaluation accountability, accuracy, utility, propriety, and feasibility, including but not limited to Explicit Values, Relevant Information, Contextual Validity, Responsive and Inclusive Orientation, Human Rights and Respect, Clarity and Fairness, Justified Conclusions and Decisions, and Internal and External Metaevaluations.

In the end, no set of standards is meaningful, useful, or interpretable outside the perspective of experience. The standards promulgated by the third edition of the PES seem all the more appropriate, applicable, and persuasive given the context of what has been learned and experienced since they were first published. While there exist now a number of sets of standards for evaluation practice separately or in the context of professional practice published by various academic and professional organizations, the framework and comprehensiveness of the standards presented in *The Program Evaluation Standards* provide an important perspective from which to consider, interpret, and apply those other standards. "One way to learn how to use the standards effectively is to apply them to different evaluation situations and to solicit feedback from knowledgeable others. Discussing applications with peers and other evaluation stakeholders will make the learning more valuable" (p. xxxi). It is from this perspective and orientation that the standards described in the

*The Program Evaluation Standards*, third edition, are recommended as appropriate, applicable, and relevant.

## References

- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *Joint committee on standards: The program evaluation standards*. Sage.