The Aspirations, Challenges, and Influence of *The Program Evaluation Standards*, Third Edition

Paula E. Egelson, Ed.D

Divisional Director for Research and Accountability, Southern Regional Education Board (retired)

Background: The revisions to *The Program Evaluation Standards,* third edition (2011) were substantial. The authors of this revision were the most knowledgeable individuals to question about aspects of the PES.

Purpose: To better understand the historical roots and intent of *The Program Evaluation Standards*, to analyze impact, and to look toward the future, it was critical to examine the aspirations, challenges, and impact associated with the PES via interviews with the authors.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: This qualitative study included three basic interview questions, with supporting follow-up questions for each of the basic questions.

Data Collection and Analysis: The results of this study were determined by the authors' responses to interview questions, which were then coded by emergent patterns and themes.

Findings: Results of the interviews found that the authors aspired to provide relevant, up-to-date, and useful standards to guide evaluators, stakeholders, and students. They were able to successfully resolve challenges associated with the PES third edition revisions, and overall these resolved challenges made the edition stronger. Finally, the authors integrated the standards into their professional work, which positively influenced them, students, and other stakeholders.

Keywords: program evaluation standards; interviews; aspirations; impact; challenges.

Introduction

> There were two things that brought me to this project [the revision of the *The Program Evaluation Standards*]. It was the process of doing this work as scholar and a professional in the field. And I got a deeper understanding of the literature and scholarship in developing the standards.
>
> —PES third edition author, 2021

The *The Program Evaluation Standards*, third edition, is instructional and based on the shared vision of the authors and reviewers from the field of program evaluation. It adds value and deeper understanding of the standards for readers to learn the experiences and perspectives that the PES third edition authors used when creating this document.

Don Yarbrough (PES chair), Lyn Shulha, Rodney Hopson, and Flora Caruthers were the authors and members of the task force who revised the PES third edition. All but Flora Caruthers were able to participate in expert interviews for this journal article. The authors' interview responses represent their experiences and perspectives. The description below of the development of *The Program Evaluation Standards* third edition provides the foundation for this journal article and is critical in understanding the results of the author interviews.

The PES third edition authors began the revision process in 2005. The task force scoured the literature over a 10-year period and examined the literature that had produced the second edition. The authors realized that the standards themselves could not convey the complexity of using the standards in the field. As a result, they developed cases or vignettes for the standards. They noted that the creation of integrating cases across each set of standards was a significant addition to the third edition of the standards. Including cases meant implementing explicit values. The authors believed they needed to give evaluators an idea of what it meant to implement the standards, provide a context for implementation, and prepare evaluators for what they might face. Each of the four extended pedagogical cases fit together like a story, tying together all the standards in each dimension of quality. The new Evaluation Accountability standards expanded the case vignettes that were used in the Utility chapter.

Another critical inclusion for the 2011 edition was an emphasis on the extensive role of culture in evaluation. The authors understood that the issues of cultural relevance, cultural responsiveness, and cultural competence extend across all the standards. One author noted that they did a respectable job in this area and created a foundation for future work. All the authors agreed that this was an ongoing subject for discussion in the five years of developing the third edition. After much discussion about whether to have separate standards address culture, they decided to make the focus on culture an intentional and integral part of each set of standards.

During the development process the four authors divided into pairs, with each member of the pair taking the lead on one dimension of quality (accuracy, feasibility, propriety, utility, and evaluation accountability). The lead in each pair undertook the first in-depth reviews of the scholarship and led discussions of the needed revisions. The pairs worked together to keep the written notes current with the group discussions. The full task force reached consensus on all operational decisions about what to submit to the JCSEE.

Outreach resulted in a lot of field testing during the development and writing processes. The task force communicated with national organizations, local users, institutions of higher education, and nonprofits throughout Canada and the United States and in other selected countries around the world. The level of engagement was amazing. The authors visited the JSCEE member institutions (education research and education practitioner organizations) and traveled to various conferences speaking on behalf of the standards. There were opportunities for stakeholders to be completely candid and strongly express their likes and dislikes. Early field trials occurred with draft standards that were still in development, and the task force could review and revise to address concerns as they arose. The task force had many informal conversations during the first three years of the development, as well, which helped direct the initial revisions.

Teamwork was critical to the task force's writing success. The task force chair sought people who could work well together but who also brought different world views and perspectives. The task force members had deep respect for one another's points of view and needed those different points of view to be naturally open to all the information and opinions coming in from stakeholders (including national and international reviewers and field testers). The chair commented that the task force was selected by the JCSEE, and that the large number of volunteers was the aspect of the process he was most pleased about. Task force members had to have a certain openness to each other and to criticism. Another author mentioned that the chair was the glue for the task force cohesiveness. The group brought together a great deal of experience

from Canada and the midwestern, northeastern, and southeastern regions of the United States. Appendix C of *The Program Evaluation Standards* includes more than 400 outside contributors from North America and other countries around the world who contributed directly with reviews, field trials, or other feedback (Yarbrough et al, 2011).

Method

To better understand the historical roots and intent of *The Program Evaluation Standards*, analyze impact, and look toward the future, it was critical to examine the aspirations, challenges, and impact associated with the PES. The best method for this study was a qualitative approach, because it answered many "What?" questions. Interview questions were created by the author of this article, and each PES third edition author was contacted about participating in an interview about the development of the PES third edition.

The PES third edition expert author interview questions included:

- When writing or revising the standards, what did you hope to accomplish? Have your aspirations for the PES changed over time? How? Have your expectations been realized? Describe. What are your aspirations for the PES in the future?
- What challenges were a part of the PES development work? How were they resolved? Did addressing these challenges strengthen the PES? Explain.
- What has been the influence of the PES third edition? What are your reactions to these effects?

Participants

Three (Don Yarbrough, Lyn Shulha, Rodney Hopson) of the four task force members and authors of *The Program Evaluation Standards* (2011) took part in individual, recorded hour-long expert interviews via Zoom in fall 2021. During the writing of the third edition, Yarbrough and Shulha were full professors and veteran evaluators at public universities in the United States and Canada. Yarbrough was also the director of the Center for Evaluation and Assessment, which had an extensive portfolio of evaluation contracts and grants in education, social services, and public health throughout the United States. Shulha joined the Queen's University Faculty of Education in 1992 after receiving degrees at McMaster University, Western University, Queen's University, and the University of Virginia (with a Ph.D. in educational evaluation). She worked collaboratively in assessment and evaluation with Queen's Faculty of Health Sciences and with colleagues across Canada and in the United States, New Zealand, Scotland, and Australia. Shulha and Rodney Hopson had been students together at University of Virginia, where Shulha inspired Hopson as a junior colleague with projects in the Curry School of Education, under what was then called the Bureau of Educational Research. Hopson was an associate professor, former board member of the American Evaluation Association (AEA), and founding director of the AEA Graduate Education Diversity Internship program, and served in other key leadership roles of AEA and other professional associations. Flora Caruthers, who was unable to participate in an interview, worked for the Office of Program Policy Analysis and Government Accountability for the Florida legislature during the writing of the PES third edition.

Four professors who use the PES third edition standards and have a connection to the JCSEE were also each asked to write a paragraph about the influence of the standards on their work. They submitted their responses via email to the author of this article.

Procedures

Individual interviews for this journal article were conducted and recorded with three of the authors of the PES third edition. Individual interviews permitted the author of this article to obtain personalized and rich information (Mason, 2002). The semi-structured interviews of each of the authors were transcribed. Semi-structured interviews were preferable because they consisted of questions that were predetermined but were flexibly worded to permit respondents to answer freely and openly; the researcher was also able to ask follow-up questions (Hancock & Algozzine, 2006).

Responses were coded by patterns and themes using an inductive approach. The methodological guidance on inductive coding developed by Saldana (2012) was used. This is described as a "ground-up" way of doing qualitative research; the theory emerges from the raw data, and categories of findings emerge (Thomas, 2006).

The influence-of-the-standards paragraphs written by the four professors associated with the JCSEE were paraphrased by author of this journal article and included in the manuscript.

Instrument, Design, and Analysis

There were three basic interview questions (standards aspirations, development challenges, influence of the standards), with supporting follow-up questions for each of the basic questions. This was a qualitative research study. The results of this study were determined by the authors' responses that were coded by emergent patterns and themes.

Results

The authors' responses included not only PES third edition aspirations, challenges, and influence, but also remarks about PES operational decisions, PES dissemination, contents of a future PES edition, PES dissemination, and cultural responsiveness. It was decided to include these additional topics in the Results section because they were critical to and at the heart of the PES third edition.

Aspirations

The authors had a variety of responses to the question about their aspirations for the PES third edition standards. One author expected that when the standards were put out there, people would use them and see them as a tool. The standards were to be viewed as a two-prong tool for evaluators (to think about their work) and for evaluator educators (to think about their instruction). Instructors would shape a course around the standards and conduct workshops on the standards. In addition, the standards would be used as a problem-solving tool. What was different about these standards was that evaluation clients could use them as well, to see if they were getting good service from evaluators. They could expect accountability about the work being done for them. In short, the aspirations for the standards were that it would be a guide for evaluation practitioners and for stakeholders, and would shape instruction for those entering the field.

Another author didn't distinguish what *he* wanted to accomplish from what he wanted the Joint Committee on Standards for Educational Evaluation to achieve. Overall, it was for the standards to lead to improved program evaluation. The difference was that he saw himself as trying to provide leadership to the group. Specifically, he wanted the Joint Committee to bring the standards up to date and address issues that were important. This author noted that enormous transformations in evaluation had occurred over five years. He was focused on making the standards culturally responsive and bringing them up to date in a variety of fields. The PES third edition task force was always attuned to the ethical and legal frameworks.

There was new and vibrant literature about pedagogical cases and what made the most effective pedagogical cases. There were subcomponents under each standard that helped people to become better program evaluators. This author wanted to see standards used appropriately. His hopes were realized, and he credited the full committee (JCSEE)—their ability to work together and learn from one another—for that.

Challenges

One author really didn't see challenges in completing the revisions to *The Program Evaluation Standards*. The team members knew they needed to be face-to-face to do the work, so the chair and task force members worked to find the money to do that. Each time they met, it would be for two or three days, with no interruptions, as to maximize their time together through coordinated schedules to incorporate meeting, eating, and writing continuously. The task force discussed and accepted specific tasks and agendas, completed them mostly on time, and reported back to one another on them, sometimes working individually and sometimes collectively, while being guided by questions at the outset, such as:

- What format do we think would be the most relevant and appropriate for this edition? If we are changing the format, why are we changing it and how should we anticipate feedback?
- Why do you want to include cases? What should be the content of the cases?
- What is the logic of our thinking in the revision process, and potential additions to the standards?

Other authors recounted challenges that the task force overcame. One remarked about how political the standards development became, related to the traditions of the first and second editions. The previous authors wanted to include second-edition grammatical forms and content in the main text of the third edition, but the third-edition task force didn't want to be boxed in by those constraints. In short, what to do with the second-edition text was a challenge.

Loss of focus of the task force was a real possibility. The sheer amount of reading was a challenge. The authors had to go over the scholarship as well as extensive comments and feedback, and they didn't want to misstate people's points of view. The task force wanted these standards to be attuned to people who disagreed with one another. It was hard to walk the line,

remaining inclusive and coherent. The task force had to wrap their heads around everything they needed to know to pull the standards together and conceptualize them, and that was a barrier. New evaluation work was always coming from the field, and it was overwhelming to keep track of it. The number of colleagues and experts who reviewed the task force's work and gave then a hearing was great, but the task force then needed to organize the feedback. Of course, money was needed to fund this enterprise, and time was needed in order to operate. But, the team never had any disagreements that derailed them. They had no problem stating their opinions. They did not personalize disagreements, but made consensus decisions and moved on.

Influence

The three authors responded to the question about the influence of the standards in personal but also professional ways. One author was immersed in standard development for five years, working on the standards, teaching the standards, and doing credibility tests on them. If these standards were going to be developed, they had to have relevance. All the team members felt the same way. This wasn't just an intellectual process; it was a practical process too.

The authors used the standards every day. They did evaluations and shared the standards with clients. In fact, the standards were required primary or secondary texts for education evaluation, doctors, and natural science research courses at some universities. Also, the National Institutes of Health (NIH) and the Centers for Disease Control adopted them—with the NIH using the standards in solicitations, too.

One author noted that the standards were rooted in scholarship. They were an organizing principle for evaluators doing scholarly work. They were a framework in which to consider the role of scholarship. Evaluators could create consensual truth to make policy decisions.

The higher the stakes in the evaluation, the more stakeholders would want the evaluation to turn out a particular way. If the standards were used authentically, they safeguarded evaluators from being impeached. One author stated in an interview:

> I was the lead evaluator on a grant evaluation and the awardees were simply not implementing [the program as designed]. I explained to the awardees that the first year was just a process evaluation. The awardees

who were from different organizations got angrier and angrier [with one another and with me]. Each year's evaluation report following the PES third edition, documented the deficits and ways to improve. When the program officer responded to the lack of progress on outcomes and called for an audit, had the annual evaluations not adhered to the PES third edition in their rigor and also pointed out ways to improve the interventions, they would have lost their funding rather than being told to follow the evaluation's roadmap for improvement.

The impact of the standards on the authors was extensive. The PES work helped one author sharpen his cultural, responsive, and equitable work lenses and collaborations of how work was organized to ensure that standards were met. This author continued to feel responsible for the Propriety standards, the ones that had been revised by him. Overall, the authors believed people wanted to know that the evaluators were doing the work with honesty and fidelity.

Finally, an author talked in the interview about the influence of the standards from a global perspective. He remarked:

> I am working on a global project where there is a "do no harm initiative." I try to think about harm as it connects to the standards. My global work is an extension of the work I have done with the standards.

To get another perspective, it was valuable to query others in the fields of evaluation and research with connections to the JCSEE about the influence of the standards in their work. These evaluators and researchers were chosen because the know and use the PES regularly. Two professors mentioned sharing the standards and assigning projects using the standards to their graduate students. One researcher described the PES as offering a different direction in providing an equity-based vision for authentic results and solutions. Another researcher remarked that the serious structural and content changes made to the PES third edition permitted his task force on *Classroom Assessment Standards* to strive for relevance and usefulness. Here are their summaries:

Steve Bingham, a retired professor at High Point University, NC, summarized that his greatest and most affirming lesson learned is that, if taught as part of a doctoral program, particularly in dissertation research, program evaluation has a better-than-even chance of being used and useful for public school districts and their students. In the

pursuit of education as evidence-based practice, this was good news.

John Fischetti, pro vice-chancellor at the University of Newcastle, Australia, noted that in a worldwide climate of high-stakes assessment supported by questionable metrics that serve political aims rather than educational aspirations, *The Program Evaluation Standards* offered a different direction. Using a knowledge-based framework, the standards converged theory and practice to support an equity-minded vision for authentic results and solutions. These standards drive teaching, learning, and assessment practices to help guide success not for some, but for all.

Barbara Howard, a retired associate professor at Appalachian State University, NC, used *The Program Evaluation Standards* primarily for teaching graduate school courses on program evaluation and school improvement. Her students conducted program evaluations, aligned with the appropriate standards, of actual programs operating in their schools or districts. The standards were meant to be used as guidelines. Some standards did not apply well for use by school administrators not using outside evaluators. However, the standards provided excellent guidance to principals and superintendents intent on determining whether the programs within their schools were providing desired results.

Don Klinger, pro vice-chancellor at the University of Waikato, New Zealand, was involved with the Joint Committee on Standards for Educational Evaluation over the years. He was one of the authors of *The Personnel Evaluation Standards* and the co-chair of the task force that developed *Classroom Assessment Standards. The Program Evaluation Standards* was a critical resource for that task force's development work. The structure of the second edition of *The Program Evaluation Standards*, including the four broad categories of accuracy, feasibility, propriety, and utility, provided the foundation for *The Personnel Evaluation Standards*. While the standards within each of these four categories varied, his team valued the underlying structure of *The Program Evaluation Standards*, including the term "should," although it was perhaps the greatest source of debate. Later, he was part of the task force that created the classroom assessment standards (CAS). Admittedly, the CAS appear far removed from *The Program Evaluation Standards*. Nevertheless, the substantial structural changes made to the third edition of *The Program Evaluation Standards* gave his task force permission to rethink the original student evaluation standards in terms of their intended purposes and audience, and to create a new set of

standards that would be more approachable and usable by their intended audience.

PES Operational Decisions

The full task force reached consensus on all operational decisions about what to submit to the JCSEE. Various authors noted during interviews:

- "Propriety standards are probably what I knew less about but was the most interested in. I was trying to put my head around propriety and then I got it."
- "I struggled with the notion of propriety, which is hardly referenced. What ethics do values have? I had conversations with Ernie House about this."
- "It got to be all-encompassing. I was the only one of my evaluator colleagues going around introducing myself using aspects of the standards… such as P1 [Propriety – Responsive and Inclusive Orientation] or U4 [Utility – Explicit Values]."
- "We were attuned to feasibility as well. There had been quite a bit of work done in project management and how to design and implement programs. That is an enormous field."
- "People thought without accuracy, feasibility, and propriety you don't have utility. There was a tension there. We tried to be open to everyone."
- "I felt like I was a graduate student, going back and having to figure out all the scholarship from the second edition to the third edition. This was a collaborative and independent activity based on our assignment of the standards."

The developers of the third edition knew there would be pushback because the third edition looked so different, with the inclusion of cultural contexts and responsiveness in evaluation, metaevaluation, extended pedagogical case vignettes, and a new set of standards that were not in the first and second editions. Said one of the authors in an interview:

We knew absolutely there was going to be pushback about the new edition, no matter what. We wanted pushback and considered it a good thing. It was frustrating in part because we needed to be encouraging and accepting of all suggestions, if the revised standards were to be the best they could be. Some opposition had to do with just keeping the standards the same, and that didn't make them better. Most of the feedback did make them [the standards] better,

once we were able to reconcile and integrate different opinions, which were often opposed to one another, not just to what the task force had put forward.

A PES-third-edition task force author in an interview spoke specifically about standards development and outreach:

For example, we paid a lot of attention to latest version of the AEA's guiding principles. We wanted to complement them. Development is a challenging enterprise and needs to be done in a way stakeholders can own them. That meant a lot of outreach. Outreach to member organizations, but practitioners too. The standards needed to belong to them, the intended beneficiaries, and users.

The Next Editions of the PES

So, what about a revision of the PES in the future? The authors have not heard a clamor for new standards or a rewrite. They believe that there are a number of ways the standards could be improved without an overhaul. The standards need to be tweaked, not totally revised. One never wants to be in a hurry to revise standards. They can be looked at every five years, and an addendum could be included.

One of the authors said if there was a rewrite of PES today, the Accuracy standards would certainly get lots of attention. Accuracy issues have to do with truth and trust as well as shared representations of consensual reality. The author shared that the structure and major topics in accuracy can stay the same, but that they need to be updated to integrate insights and scholarship about how to achieve accuracy in program descriptions, process descriptions, and outcomes and impact assessments. Recent scholarship also deepens insights into inferring causality from naturalistic frameworks, such as case study, quasi-experimental, and semi-experimental field studies. Without being critical of either second- or third-edition Accuracy standards, it can still be admitted that recent advances in scholarship can guide problem-solving with and application of the Accuracy standards. Recent scholarship contributes to a better understanding of validity and validity arguments as they apply to interpretations, conclusions, and consequences. This scholarship has significant importance for the Accuracy standards and their connection to utility. Current understanding of validity argumentation bridges the divide between practitioners with a sole

focus on either quantitative or qualitative methods. The author believes that the Accuracy standards were well aimed but need to be supplemented by deep understanding of current scholarship for their best use in improving evaluations.

Funding PES Revisions

Funding of evaluations is another area that needs to be addressed in future editions, noted another author. This author didn't understand it until working as an evaluator with Indigenous groups in Canada. Many funders—especially governments—invested in programs for marginalized groups with the intention of promoting inclusiveness, redress, or equity of opportunity and outcomes. In doing so, they typically established specific "deliverables" that they expected those applying for the funding to address and, hopefully, produce.

When collaborating with Indigenous peoples, this author finally realized that funders may not have fully understood how requiring programs to work toward such deliverables to qualify for funding may have actually helped to maintain the very system that marginalized Indigenous people in the first place. Specifically, funders may not have fully considered what the path to these deliverables might look like for the beneficiary group. They may not have considered (a) different ways of knowing, (b) the system of individual and communal values that shapes decision-making, and (c) how power is shared or distributed within a community.

Logic models are often used as tools to chart the path from current conditions to expected outcomes, but the logic required by these models is typically grounded in thinking that minimizes systemic influences and responsibilities. These models helped to target places where program leaders and intended beneficiaries could be required to bear responsibility when funding expectations were not met. Evaluators who truly have a deep understanding of how diversity strengthens the fabric of a society might be well positioned to help funders collaborate with potential beneficiaries to redefine program accountability as a two-way street.

Evaluators asking funders questions such as:

- How can we support you as individuals to experience more inclusivity, opportunity, and equity in relation to our current goals or initiatives?
- How can we support you in shaping a program that is contextually relevant, sensitive to

community values, and has a strong potential for reach and impact?

Evaluation that supports this type of effort has more connection to the demands of developmental evaluation than it does to a classic "needs assessment," in which an evaluator would enter the program community using the cultural and social lenses of the funding agency to assess the kind of "intervention" that is required to meet the funder/government mandate. The current program evaluation standards are different in that they were designed to support both evaluators and program users.

Dissemination

The dissemination of the standards was a topic that kept cropping up throughout the three author interviews. Author perspectives varied. Explained one author in an interview, "If the standards weren't serving a purpose, nobody cared, or they are serving a purpose and still serving it effectively. Maybe everyone who wants the standards has them."

Along the same lines, another author remarked in an interview, "I used to think the standards could speak for themselves. The best we can hope for is that the standards will inspire people to learn more about this extremely hard work."

One author saw dissemination of the standards quite differently. This author noted in an interview, "We did not have a visioning session about how we were going to think about dissemination in the future." Continuing on the same thread in the interview, this author believed that if the task force was serious about standards use, a dissemination process would have been developed. Yet, there was no money for dissemination. "We would only deliver them [the PES], but we needed to get them out to the world." The quotes above showed the differences in how the authors viewed dissemination of the PES third edition; some saw the standards as speaking for themselves, another author believed a formal dissemination plan should have been included.

One author suggested in an interview that in the dissemination of the PES third edition in the future, print will no longer be the primary learning tool. Instead, the author suggested using a YouTube format to convey how to use the standards and what they mean.

Discussion

The goals of this study were to better understand the PES's historical roots and intent and to analyze impact. To look toward the future it is critical to examine the aspirations, challenges, and impact associated with the PES. In short, the authors' aspirations were for the standards to guide evaluation practitioners, guide stakeholders, and shape instruction for those entering the field. There is the realization, at least by one author, that dissemination of the PES continues to be a weak component.

There were challenges—pushback about the new edition from others, volumes of reading material, lack of time—but the task force members were highly collaborative and functional and were able to successfully address the issues. In most cases, the challenges made the standards stronger. These task force members were chosen carefully, and that helped mitigate any of the challenges they encountered. In some cases, these challenges pointed out weaknesses in the development process, which was a good thing, because it enabled those weaknesses to be addressed and improved.

Each author addressed the question of the influence of standards from a personal yet professional perspective. In addition, the revision of the PES was a professional highlight for each of the authors. The use of the PES is now a part of their daily professional lives. External evaluators and researchers were more extensive regarding influence of the PES, citing their impact on graduate students, their representation of an equity-minded vision, and their providing a model for more relevant and useful standards.

Regarding future studies, a dissemination study on the PES is in order. Dissemination strategies that are research-based with positive outcomes should be implemented with PES and the results documented. Second, developing cases studies centered around the PES third edition authors' integration of the standards professionally would be useful. Finally, creating guidelines for when to tweak or revise standards would be valuable.

Author's Note

References

Fournier, D. M. (1994). [Review of the book *The program evaluation standards: How to assess* evaluations, by The Joint Committee on Standards for Educational Evaluation]. *Journal of Educational Measurement*, *4*(31), 363–368.

Hancock, D., & Algozzine, B. (2006). *Doing case study research*. Teachers College Press.

Mason, J. (2002). *Qualitative researching* (2nd ed.). Sage.

Saldana, J., (2012). *The coding manual for qualitative researchers*. Sage.

Thomas, D., (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage.