# What Do We Know about How the Program Evaluation Standards Are Used in Public Health?

Goldie MacDonald
*Centers for Disease Control and Prevention*

Kimberly A. Castelin
*Centers for Disease Control and Prevention*

Naje' M. George
*Tuskegee University*

Asmith Joseph
*Oak Ridge Institute for Science and Education*

**Background:** Released by the Centers for Disease Control and Prevention (CDC), Framework for Program Evaluation in Public Health prominently features the program evaluation standards (1999). The program evaluation standards (PES) include 30 statements in five domains: utility, feasibility, propriety, accuracy, and evaluation accountability. Despite decades of attention to the PES among framework users and others, how public health professionals apply these standards in their work is not well understood.

**Purpose:** The study sought to identify notable commonalities in how the PES are used in public health.

**Setting:** Application of the PES in evaluative work in public health and allied fields.

**Intervention:** Not applicable.

**Research Design:** The study included a search of subscription and nonsubscription sources to identify documents that included explicit content concerning use of standards in evaluative work in public health. Documents identified were screened using predetermined criteria to include or exclude each item in the study. Items included were reviewed and coded using codes developed before examining all documents. For each code, reviewers discussed data from all documents to identify commonalities and variations in application of standards.

**Findings:** The literature search returned 405 documents to be screened (179 from subscription and 226 from nonsubscription sources). Thirty-eight items were included in the study based on initial screening (11 from subscription and 27 from nonsubscription sources). The study revealed that authors discussed standards as a regular component of evaluation work, but precisely how standards were used was not always explained in detail. Also, authors did not always discuss standards statements but sometimes solely focused on general domains (e.g., feasibility or accuracy). When authors discussed specific statements, they were more descriptive in how they applied the PES (i.e., compared with articles that focused on general domains). Overall, authors placed far greater emphasis on Accuracy and Utility standards, compared with Propriety, Evaluation Accountability, or Feasibility. In many cases, authors used the PES in combination with other resources (e.g., checklists, guidelines, or other standards). Although program evaluation is crucial to public health practice, the mechanics of how professionals consider, integrate, or use evaluation standards is not fully understood.

## Introduction

The de Beaumont Foundation and Public Health National Center for Innovations (2020) engaged collaborators nationwide to update *10 Essential Public Health Services*. The document provides concise, shared concepts that define components of public health systems and practices and differentiate public health from health care (Castrucci, 2021). Originally published in 1994, *10 Essential Public Health Services* includes program evaluation as a necessary activity. However, there is still much to understand about how individuals and organizations conceptualize, implement, or use program evaluation in public health. Specifically, although the program evaluation standards (PES) are featured in *Framework for Program Evaluation in Public Health* (CDC, 1999) and other resources, how public health professionals use these standards in evaluative work is not well understood. The Joint Committee on Standards for Educational Evaluation (JCSEE) developed the current edition of the PES through formal and informal needs assessments; reviews of relevant literature and other materials; and consultation with diverse collaborators, domestic and international (JCSEE, 2021). They organized 30 standards into five domains that correspond to important dimensions of evaluation quality, including utility, feasibility, propriety, accuracy, and evaluation accountability (Yarbrough et al., 2011). Despite the importance of program evaluation in public health and the visibility of these standards for decades, there is much to be learned about commonalities and variations in use across practitioners and public health activities. This study sought to (1) identify and understand commonalities in PES use in public health, and (2) pinpoint shared challenges in using these standards. Ideally, findings can be used to promote discussion and enhance use of PES among evaluators and nonevaluators who participate in program evaluation in public health and beyond.

## Data Collection and Analysis

A literature search of subscription and other sources identified documents that addressed PES use in public health and allied fields (e.g., healthcare or nursing). The search spanned a 20-year period (2001 to 2021) following release of *Framework for Program Evaluation in Public Health* (CDC, 1999). This framework is widely used in public health and features PES domains and statements. A professional librarian conducted a Boolean search of APA PsycINFO, the Cumulative Index of Nursing and Allied Health Literature (CINAHL) database, Embase, MEDLINE, and Scopus. The search also included items from Google Scholar and from CDC Stacks, a digital archive of documents for public health professionals, researchers, and the public. The search yielded 179 subscription and 226 nonsubscription materials to be screened.

Two reviewers screened each abstract for explicit attention to standards. In certain cases, entire papers were reviewed to know whether to include an item. Reviewers identified whether attention to standards was present and active (i.e., not implied or inferred). To be selected, authors did not have to specifically reference PES, but items selected prominently featured one or more standards. In many articles, authors mentioned quality or utility in passing, but attention to these constructs was not a substantial component. In cases where reviewers did not agree on whether to include a document, a third reviewer screened the item and all reviewers discussed discrepancies to reach consensus.

The screening process produced 12 subscription and 32 nonsubscription (i.e., open access) items to be coded. While coding these items, 6 documents that should not have been included were excluded. In total, we reviewed and coded 38 items (11 subscription and 27 nonsubscription) dated from 2003 to 2021, using 15 codes that we developed before examining all documents. For example, we identified the public health activity or program discussed in each document; the standards domains or statements explicitly mentioned; the ways standards were used in each document; and which other standards or resources were used. Initially, three reviewers coded the same three documents and discussed the results to ensure that all had a shared understanding of the concepts or constructs to be located and coded. For each code, reviewers extracted and recorded information in a shared spreadsheet. Data analysis included collaborative review of information extracted from all articles to explore how content for each code was similar or different and identify emergent patterns or themes.

## Findings

Documents reviewed primarily included journal articles, dissertations, and evaluation reports. Authors addressed a range of practice areas: chronic disease prevention and health promotion, cultural awareness in nursing, disaster or emergency response, maternal and child health, occupational health and safety, tobacco use

prevention and control, and workplace health promotion, for example. From these documents, distinct findings emerged with regard to PES application: use of standards domains and statements considerably varied among authors; the Evaluation Accountability domain was solely represented through metaevaluation and was generally underrepresented; and authors regularly applied the PES in combination with other resources relevant to their practice area or program. These findings are described in detail below.

## *Differences in Application of Standards Domains and Statements*

Data analysis revealed differences in how authors discussed standards domains and statements within domains. In certain cases, standards were discussed at the domain level only (i.e., authors did not explicitly reference statements within a domain). Authors also referenced constructs that overlap with PES content (e.g., credibility, ethicality, reliability, or validity) to explain the quality of evaluation activities, without directly matching these constructs to corresponding PES domains or statements. In these cases, it is not clear whether authors were drawing on content in the PES without noting a connection or using related constructs as a complement or supplement to PES domains or statements.

For all articles reviewed, focus on Accuracy and Utility standards exceeded Evaluation Accountability, Feasibility, and Propriety standards, at both domain and statement levels. Although certain articles discussed Propriety standards, this domain was least documented. Within domains, authors used some standards statements more than others. In the Utility domain, Standard U2 was most common (i.e., evaluations should attend to the full range of individuals and groups invested in a program or evaluation). In the Feasibility domain, Standards F2 (i.e., procedures should be practical and responsive to how a program operates) and F4 (i.e., resources should be used effectively and efficiently) were used most often. In the Propriety domain, Standards P4 (i.e., evaluations should be understandable and fair to those with an interest in the program or evaluation) and P7 (i.e., evaluations should be fiscally responsible) were most common. In the Evaluation Accountability domain, Standard E2 (i.e., implement internal metaevaluation) was documented more than others.

Attention to reliability and validity are cornerstones of epidemiology, the basic science of public health. Yet, a narrow focus on reliability and validity can limit attention to other, relevant dimensions of accuracy discussed in the PES (e.g., description of program context, conclusions and decisions explicitly justified in reference to cultures and contexts, and appropriate communication and reporting). Sixteen of 38 articles primarily addressed accuracy. Not unexpectedly, in the Accuracy domain, Standards A2 (i.e., information from an evaluation should serve intended purposes and support valid interpretations) and A3 (i.e., procedures should yield reliable, dependable, and consistent information) were used most frequently. However, in articles that emphasized data quality, authors addressed various aspects of accuracy (e.g., appropriate data collection methods, sample type and procedures, data completeness, data timeliness, and data comparability).

In articles that applied standards statements from all five domains, data analysis revealed distinct commonalities in how statements were used. Compared with authors who primarily focused on standards domains, authors who focused on standards statements discussed application in far more detail. For example, when authors used domain-level descriptions as general checkpoints regarding quality, they did not explain how each domain was operationalized or satisfied to determine quality. In contrast, we noted greater variety in how authors used standards statements. In one case, authors translated statements to a rubric to assess and improve a program monitoring system in an Australian rural health program. These authors mapped elements of a program monitoring system to content in the PES to "unpack the quality and potential of monitoring activities" (Kelly & Reid, 2021, p. 43). In some cases, authors applied one or more statements to assess the quality of a specific element of an evaluation or to support decisions about evaluation design or implementation. In other examples, standards statements were used to explain actions to encourage and support participation in an evaluation for a perinatal nutrition program (Bodenhamer, 2016) and served as reference points to assess strengths and weaknesses in evaluation of European disaster and risk management programs (Beerens, 2021). Authors also used standards statements to adjudicate, explore, or understand sticky points, tensions, or uncertainties in evaluative work. For example, authors used concurrent metaevaluation to balance usability and methodological rigor in an assessment tool and process to identify promising workplace health promotion programs (Dunet et al., 2008). In each case, authors used content in the PES that directly aligned to their context, interests, or information

needs (e.g., to assess, explain, or make decisions regarding evaluative work).

## *Evaluation Accountability Domain Solely Represented as Metaevaluation*

Published in 2011, the third edition of PES is the first to include the Evaluation Accountability (EA) domain; two of three statements in this domain focus on metaevaluation (i.e., internal and external metaevaluation). Like Propriety standards, EA standards were underrepresented in the items we reviewed. In 26 articles published since 2011, only 8 included content from five PES domains, and 9 addressed four domains. But, among all documents reviewed, those that applied metaevaluation checklists (with concise checkpoints for standards statements) consistently demonstrated more parity in use of standards across the five domains. In other words, attention to metaevaluation went hand in hand with attention to standards in domains other than EA. Four articles published in 2011 or later that primarily addressed EA focused on metaevaluation, and paid no direct attention to Standard E1, Evaluation Documentation. Four articles published before evaluation accountability was added to the PES also included content on metaevaluation, and all used checklists based on the PES.

Because metaevaluation as an evaluative activity predates the EA domain and standards, not all authors (even after 2011) explicitly describe it as an exercise in applying PES. For example, a metaevaluation of public funds evaluations in Thailand did not explicitly reference EA when discussing the PES, but the study was deeply rooted in literature on metaevaluation (Tongchiw, 2013). Among documents published in 2011 or later, authors conducted metaevaluation in a variety of ways and varied in terms of their reliance on content from one or more PES domains. To illustrate, in a metaevaluation of evaluations of health care programs focused on chronic care, the author employs Stufflebeam's *Program Evaluations Metaevaluation Checklist* (1999) but solely focuses on the Accuracy domain and statements (Fields, 2014). The Joint Committee on Standards for Educational Evaluation (2018) checklist of PES domains and statements was not discussed among tools to conduct metaevaluation in articles reviewed.

## *Program Evaluation Standards Used in Combination with Other Resources*

In more than two-thirds of documents reviewed, authors used other resources in combination with PES content. In many cases, authors combined PES domains or statements with resources specific to a program or topic area. Authors drew on content from other resources to match their context or information needs. In documents that introduced other resources, reviewers identified more than 30 distinct items that were combined, to varying degrees, with PES content.

In Table 1, examples of these resources are grouped into three categories: checklists, frameworks or models, and guidelines or standards. Although not all classifications were straightforward and category labels can be differently interpreted (e.g., definition of "framework" versus "guideline"), documents in each category share similar characteristics (e.g., format, intended uses, type of content). For example, authors used *Framework for Program Evaluation in Public Health* (CDC, 1999) to organize or illustrate application of standards domains. Authors presented content in this framework as documentation of attention to evaluation standards (i.e., rather than discussing specific standards statements in one or more domains). In several documents that applied standards statements, authors weaved in other resources to expand or strengthen discussion of PES. For example, alongside Accuracy standards, authors noted additional criteria regarding data quality. Other authors incorporated content from specific participatory approaches to program evaluation as they applied Propriety and Utility standards statements. Use of other resources in combination with content in PES was one of the most pronounced commonalities across documents reviewed. In many cases, evaluation quality was not considered in reference to PES alone.

Table 1. Example Resources Used in Combination with Content in the Program Evaluation Standards

| Resource type | Resource name |
|---|---|
| Checklists | • *Key Evaluation Checklist* (Scriven, 2007)<br>• *Program Evaluations Metaevaluation Checklist* (Stufflebeam, 1999)<br>• RIMES (reporting recommendations intended for pharmaceutical risk minimization evaluation studies) statement (Smith et al., 2018) |
| Frameworks or models | • *The CIHI Data Quality Framework* (Canadian Institute for Health Information, 2009)<br>• "Framework for Program Evaluation in Public Health" (Centers for Disease Control and Prevention, 1999)<br>• *IAP2 Spectrum of Public Participation* (International Association for Public Participation, 2007)<br>• Purnell model for cultural competence (Purnell, 2002)<br>• RE-AIM (reach, efficacy, adoption, implementation, and maintenance) framework (Glasgow et al., 1999) |
| Guidelines or standards | • AEA *Guiding Principles for Evaluators* (American Evaluation Association, 2018)<br>• Evaluation Research Society standards for program evaluation (Anderson et al., 1982)<br>• Government auditing standards (United States Government Accountability Office, 2011)<br>• HEAL (hierarchy of evidence and appraisal of limitations) grading system (Gugiu, 2015)<br>• *Norms and Standards for Evaluation* (United Nations Evaluation Group, 2016)<br>• *Public Health Code of Ethics* (American Public Health Association, 2019)<br>• *Quality Standards for Development Evaluation* (Organisation for Economic Co-operation and Development, 2010)<br>• Society for Prevention research standards of evidence (Flay et al., 2005) |

## Conclusions

The study examines the use of PES in different organizations, programs, and settings in public health and allied fields. Data analysis revealed commonalities in how practitioners conceptualize and operationalize evaluation quality to examine or make determinations about evaluative work. We interpreted authors' use of PES content and complementary resources as visibly purposeful in nearly all documents reviewed. The findings provide a glimpse into real-world flexibilities in using standards to meet specific needs in a variety of evaluation and program contexts.

There are lessons from these documents for individuals and organizations in public health and beyond. Those who develop tools or training on the PES should balance descriptions of the standards and why they are important with more content on how and when standards can be applied in evaluative work. For example, in *Framework for Program Evaluation in Public Health* (CDC, 1999), PES domains were embedded in a graphic used to

represent the document for more than two decades. Although standards statements are presented in the framework, how they can be used is not discussed in detail.

There is also much opportunity to emphasize and encourage creativity in application of the PES in everyday practice or professional development activities focused on program evaluation. *The Program Evaluation Standards* (Yarbrough et al., 2011) includes rich case examples in each standards domain, but a book may not be an ideal vehicle to reach those tasked with program evaluation in diverse organizations or workplaces at the front lines of public health. Wider dissemination and promotion of open-access, plain-language case examples and job aids on how to apply standards domains and statements can be helpful.

In the articles we reviewed we found less emphasis on the Propriety standards than on standards in other domains. Statements in the Propriety domain address crucial components of evaluation practice that should not be overlooked (e.g., fairness, human rights, inclusion, and respect for others). Documents included in this review did

not likely capture the current landscape of activities or discourse relevant to content in the Propriety domain. Similarly, editorial choices can make items harder to identify in literature searches or abstract reviews.

Reviewing a small number of documents focused on standards use confirms there is much more to know about how practitioners understand and apply the PES and other resources to reflect on, judge, and improve evaluation processes and products. Data analysis revealed patterns in PES use in public health, but this study is limited to content in searchable literature. It does not include evaluative work not captured in journal articles, dissertations, or evaluation reports archived in accessible databases. Future studies should include primary data collection (quantitative or qualitative) with evaluators and others who use the PES to fully understand this dimension of evaluation practice.

## Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## References

American Evaluation Association. (2018). *Guiding Principles* [Brochure]. https://www.eval.org/Portals/0/Docs/AEA_289398-18_GuidingPrinciples_Brochure_2.pdf

American Public Health Association. (2019). *Public health code of ethics*. https://www.apha.org/-/media/files/pdf/membergroups/ethics/code_of_ethics.ashx

Anderson, S. B., Braskamp, L. A., Cohen, W. M., Evans, J. W., Gilmore, A., Marvin, K. E., Shipman, V. C., Vanecko, J. J., & Wooldridge, R. J. (1982). Evaluation Research Society standards for program evaluation. *New Directions for Evaluation*, *1982*(15), 7–19. https://doi.org/10.1002/ev.1306

Beerens, R. J. (2021). *Improving disaster response evaluations: Supporting advances in disaster risk management through the enhancement of response evaluation usefulness*. [Doctoral dissertation, Lund University]. Research Portal.

Bodenhamer, S. A. (2016). *Using participatory methods to create a logic model for Healthiest Babies Possible*. [Master's thesis, Simon Fraser University]. Summit Institutional Repository.

Canadian Institute for Health Information. (2009). *The CIHI data quality framework*. https://secure.cihi.ca/free_products/dq-data_quality_framework_2009_en.pdf

Castrucci, B. C. (2021). The "10 Essential Public Health Services" is the common framework needed to communicate about public health. *American Journal of Public Health*, *111,* 598–599. https://doi.org/10.2105/AJPH.2021.306189

Centers for Disease Control and Prevention (CDC). (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, *48*(RR-11), 1–40. https://www.cdc.gov/mmwr/PDF/rr/rr4811.pdf

Dunet, D. O., Sparling, P. B., Hersey, J., Williams-Piehota, P., Hill, M. D., Hanssen, C., Lawrenz, F., & Reyes, M. (2008). A new evaluation tool to obtain practice-based evidence of worksite health promotion programs. *Preventing Chronic Disease*, *5*(4). https://www.cdc.gov/pcd/issues/2008/oct/07_0173.htm

Fields, J. (2014). *A metaevaluation of evaluations of health care programs that employ the chronic care model* (Publication No. 241) [Doctoral dissertation, Western Michigan University]. ScholarWorks. https://scholarworks.wmich.edu/dissertations/241

Flay, B. R., Biglan, A., Boruch, R. F., González Castro, F., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151–75. https://doi.org/10.1007/s11121-005-5553-y

Glasgow, R. E., Vogt, M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. American Journal of Public Health, 89(9), 1322–7. https://ajph.aphapublications.org/doi/epdf/10.2105/AJPH.89.9.1322

Gugiu, P. C. (2015). Hierarchy of evidence and appraisal of limitations (HEAL) grading system. *Evaluation and Program Planning, 48,* 149–59. https://doi.org/10.1016/j.evalprogplan.2014.08.003

International Association for Public Participation. (2007). *IAP2 spectrum of public participation*. https://www.iap2canada.ca/Resources/Documents/IAP2%20Spectrum_vertical.pdf

Joint Committee on Standards for Educational Evaluation (JCSEE). (n.d.). *Program Evaluation Standards.* https://evaluationstandards.org/program/

Joint Committee on Standards for Educational Evaluation. (2018). Checklist of the program evaluation standards statements. https://wmich.edu/evaluation/checklists

Kelly, L. M., & Reid, C. (2021). Baselines and monitoring: More than a means to measure the end. *Evaluation Journal of Australasia*, *21*(1), 40–53. https://doi.org/10.1177/1035719X20977522

Organisation for Economic Co-operation and Development (OECD). (2010). *Quality standards for development evaluation.* https://doi.org/10.1787/9789264083905-en

Public Health National Center for Innovations. (2020). *The 10 essential public health services.* https://phnci.org/uploads/resource-files/EPHS-English.pdf

Purnell, L. (2002). The Purnell model for cultural competence. *Journal of Transcultural Nursing*, *13*(3), 193–196. https://doi.org/10.1177/10459602013003006

Scriven, M. (2007). *Key evaluation checklist.* https://wmich.edu/sites/default/files/attachments/u350/2014/key%20evaluation%20checklist.pdf

Smith, M. Y., Russell, A., Bahri, P., Mol, P. G., Frise, S., Freeman, E., & Morrato, E. H. (2018). The RIMES statement: A checklist to assess the quality of studies evaluating risk minimization programs for medicinal products. *Drug Safety*, *41*(4), 389–401. https://doi.org/10.1007/s40264-017-0619-x

Stufflebeam, D. L. (1999). *Program evaluations metaevaluation checklist.* https://wmich.edu/sites/default/files/attachments/u350/2014/program_metaeval_long.pdf

Tongchiw, T. (2013). *The meta-evaluation of revolving funds evaluation in Thailand.* [Doctoral dissertation, National Institute of Development Administration]. NIDA Wisdom Repository.

United Nations Evaluation Group. (2016). *Norms and standards for evaluation.* http://www.unevaluation.org/2016-Norms-and-Standards

United States Government Accountability Office. (2011, December). *GAO-12-331G government auditing standards: 2011 revision.* https://www.gao.gov/assets/files.gao.gov/assets/gao-12-331g.pdf

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.