SHARE

f · t · in · ✉

Principles and Practices for Federal Program Evaluation: Proceedings of a Workshop in Brief (2017)

DETAILS

GET THIS BOOK

FIND RELATED TITLES

CONTRIBUTORS

Committee on National Statistics; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

Principles and Practices for Federal Program Evaluation

Proceedings of a Workshop—in Brief

On October 27, 2016, the Committee on National Statistics (CNSTAT) held a 1-day public workshop on principles and practices for federal program evaluation. The workshop focused on reviews of existing policies of the Administration for Children and Families, the Institute for Education Sciences, the Chief Evaluation Office in the U.S. Department of Labor, and other federal agencies. The scope of the workshop included evaluations of interventions, programs, and practices intended to affect human behavior, carried out by the federal government or its contractual agents and leading to public reports intended to provide information on impacts, cost, and implementation.

A six-person steering committee facilitated the discussions and invited participants to comment on agency policies, which reference such principles as rigor, relevance, transparency, independence, and ethics, as well as objectivity, clarity, reproducibility, and usefulness. Workshop participants considered ways to strengthen existing practices and institutionalize the principles. The goal would be to bolster the integrity and protect the objectivity of the evaluation function in federal agencies—which is essential for evidence-based policy making.

FEDERAL EVALUATION, WITH THICK SKIN

Steering committee chair **Grover "Russ" Whitehurst** (Brookings Institution) opened the workshop by emphasizing how important it is for the federal government to have a strong evaluation effort, marked by rigor and with independence, enabling agencies to provide accurate and timely information to decision makers. He noted that the results of federal evaluation are often anticipated with trepidation and face opposition when the results raise questions about the effectiveness of a program. Whitehurst also stressed that evaluations often do not reach their most important audiences, citing a 2013 report[1] that found that more than half of senior government leaders had no knowledge of evaluations of the programs for which they were responsible. He encouraged participants to consider the history of federal program evaluation, its current status, and the possible development of more formal principles and practices.

---

[1]U.S. Government Accountability Office. (2013). *Program Evaluation: Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policy Making.* GAO-13-570. Available: http://www.gao.gov/products/GAO-13-570 [January 2017].

The National Academies of

SCIENCES · ENGINEERING · MEDICINE

HOW FAR WE HAVE COME: "THE INEVITABLE MARCH OF SCIENCE" OR AN ONGOING STRUGGLE?

Moderator **Howard Rolston** (Abt Associates; member, steering committee) began by noting the progress in the field over the past 50 years. Although observers may think these advances are simply a product of "the inevitable march of science," he cautioned not to take the progress for granted: the overall trend has been more and better evaluations, but there have been setbacks and threats to the field of evaluation. He stated that institutionalizing evaluation principles and practices would better protect their quality and integrity.

Larry Orr (Johns Hopkins University) first considered the history of evaluation: What have been the major challenges to the federal government in generating and using rigorous independent research? What circumstances over time have reduced or exacerbated vulnerabilities in evaluation work? He noted three challenges: resources for research and evaluation, resistance to rigorous evaluation, and convincing policy makers to use evaluation results. Looking at the question of resources, Orr has seen limited increase in resources and operating budgets for evaluation since his tenure as head of a federal evaluation agency in the 1970s; and he feels that on a whole evaluation is still grossly underfunded. Orr cited *Fighting for Reliable Evidence*[2] as an account of the early resistance to incorporating random assignment experimentation in social policy research. He discussed the effect of the 1966 "Coleman report"[3] on the education community's initial reluctance to support quantitative evaluation. Orr also noted that the field of international development initially resisted evaluation, but since 2000 there have been approximately 1,700 randomized controlled trials in developing countries. Convincing policy makers to act on research results is one of the biggest challenges in the field, Orr said. However, evaluation is only one of the many factors that influence policy, and a small one at that. He believes, however, that its role will continue to grow because of an increasing number of congressional mandates and efforts of the U.S. Office of Management and Budget (OMB) to push for more rigorous evidence.

Jean Grossman (Princeton University) spoke about the challenges she faced both from within the federal government as an evaluation officer and in her role as a federal contractor. She noted three main issues: politics, money, and regulations. Grossman said that politics is "the elephant in the room;" evaluators are constantly fighting political pressure and a reluctance to accept evaluation results if they do not align with expectations. She said many see evaluation as a way of determining whether or not a program works, when really it is about determining whether it works better than something else. Grossman noted that political timing is also a factor, since the average 4-year time horizon for most policy makers often requires that programs be evaluated in that time frame, which may likely be too close to their inception for a meaningful evaluation. With regard to money, Grossman pointed out that only a small subset of federal funds goes directly to program evaluation—sometimes less than 0.5 percent for an agency. Seeking approval to use other administrative funding for evaluation can be difficult given the widely negative view of evaluation. Grossman said the biggest regulatory constraint to program evaluation she faced was the OMB Paperwork Reduction Act.[4] While the target turnaround for approval is 3 months, the average is 7-9 months. Since it might take up to 2 years to develop a sample and another year to analyze results and generate a report, the act makes it difficult to obtain the requisite baseline data within the approval time and essentially limits the work that can be done.

Ron Haskins (Brookings Institution) started his discussion with an anecdote about the 21st Century Community Learning Centers Program, which was evaluated and proven, despite strong community support and advocacy by then-Governor of California Arnold Schwarzenegger, to not affect student outcomes.[5] He noted that any sentence akin to saying "Everybody knows this program works" is an enemy of evidence-based policy. Haskins stressed the importance of having a statute that requires an evaluation when establishing or appropriating money for a program. He said that going a step further and adding language in the statute about random assignment can also prove very useful. Haskins gave examples of how adding evaluation language in the Welfare Reform Act and in the senate legislation for World Bank funding improved

---

[2]Gueron, J. M., and Rolston, H. (2013). *Fighting for Reliable Evidence*. New York: Russell Sage Foundation.

[3]Coleman, J. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

[4]"The purpose of the Paperwork Reduction Act (PRA), which governs information collections, is to minimize paperwork, ensure public benefit, improve Government programs, [and] improve the quality and use of Federal information to strengthen decision making, accountability, and openness in Government and society." Available: https://www.doleta.gov/ombcn/omb-controlnumber.cfm [January 2017].

[5]*The $1.2 Billion Afterschool Program that Doesn't Work*. Available: https://www.brookings.edu/research/the-1-2-billion-afterschool-program-that-doesnt-work/ [January 2017].

the utility of both programs. He also pointed to the early childhood home visiting programs under the Affordable Care Act (Obamacare–42 U.S.Code. 711), which specifically calls for evaluation through rigorous randomized controlled research.

THE STANDARD-BEARERS OF FEDERAL EVALUATION

Naomi Goldstein (Administration for Children and Families [ACF]) noted that the process to establish ACF's evaluation policy was fairly straightforward: the agency's leaders encouraged the evaluation office to develop it. The policy (published in 2012) confirms the agency's commitment, not only to conducting evaluations, but also to using evidence from evaluations to inform policy and practice. She offered the caveat, however, that evaluation is but one form of evidence necessary for learning and improvement. Goldstein discussed the five principles present in the ACF policy document: rigor, relevance, transparency, independence, and ethics.

- Rigor means getting as close as possible to the truth and being committed to using the most rigorous methods to do so, through the most appropriate evaluation method. She noted, however, that rigor does not automatically mean the use of randomized controlled trials.

- Relevance means considering the agency's needs, having strong internal and external partnerships, and disseminating findings in useful ways. She stressed that rigor without relevance could yield studies that are accurate but not useful.

- Transparency means operating in a way that supports the credibility of the data and allows for critique and replication of the methods used in the evaluation. It promotes accessibility and reinforces a commitment to release results regardless of the findings.

- Independence and objectivity are core principles of evaluation. While many parties should contribute to identifying evaluation questions and priorities, study methods and findings should be insulated from bias and undue influence.

- Ethics, she emphasized, means recognizing the importance of safeguarding the dignity, rights, safety, and privacy of participants in evaluations.

Goldstein closed by noting that disseminating policies have helped make the agency's principles a shared set of values both in the organization and with its program partners.

Demetra Nightingale (Urban Institute) discussed her previous experience as the chief evaluation officer of the U.S. Department of Labor (DOL). She explained that many of the dozen or so operating agencies within DOL have their own evaluation offices; and her role was not to centralize evaluation, but to raise the quality of evaluation and awareness of evaluation methodology. She noted that her office drew from the work of other prominent evaluation agencies when creating its policy, and she is proud that the policy is accepted and supported throughout the department. Nightingale reiterated Goldstein's point about the importance that policies contain a principle of rigor applicable to all types of evaluation and research. She said it is important to form decisions from a body of evidence and not just a single study, and she closed by reminding the group that ethics should apply to the protection of the participants and the integrity with which evaluations are conducted.

Ruth Neild (Institute for Education Statistics [IES]) strongly endorsed the principles presented by Goldstein and Nightingale. She said that the evaluation offices being discussed are more alike than they are different and that IES uses many of the same strategies. What might separate IES from other agencies is the way it incorporates formal peer review of its evaluation reports—required by its authorizing legislation—to ensure scientific merit and promote rigor and integrity. This process mitigates external and internal threats to IES reports, both by reducing the potential for suppression or alteration of their findings and ensuring report quality, thoroughness, and relevance.

Jack Molyneaux (Millennium Challenge Corporation [MCC]) explained that MCC is a small, independent federal agency that provides international development assistance. It was created to aid in global economic growth and to ensure that international investments are being used for the right purpose and yielding the expected results—both of which are contingent on having a credible evaluation strategy. MCC's policy mirrors those of prominent evaluation agencies in many ways but has key differences: one is the requirement that every project, no matter how small, be subjected to independent evaluation. To manage the cost and scope of evaluations, MCC promotes developing evaluation design in tandem with the program design. Moreover, Molyneaux explained, evaluation management is kept separate from operations, and evaluators have editorial independence in reporting their results, meaning they can accept or

reject any feedback from the sponsor. Because of the nature of MCC's projects (building roads, for example), measuring attributable impact cannot always be done using randomized controlled trials. Molyneaux reiterated Goldstein's point about the need to use methods that are appropriate to each program and emphasized that rigor is still at the forefront. He said that MCC's candor in releasing favorable and less-than-favorable results has spawned dialogue that has helped them improve their program design.

REFINING THE TRICKS OF THE TRADE

Mark Shroder (U.S. Department of Housing and Urban Development [HUD]) said that HUD's policy statement on evaluation was created after seeing the success of other agencies' documents and, as such, closely resembles several of them, noting especially the importance of transparency. However, although he agreed that every methodologically valid report should be published, he does not believe that reports of evaluations that were not methodologically sound should have to be released. Molyneaux agreed that some evaluations are indeed stronger than others; but he said that MCC prefers not to be the censor, instead encouraging the peer reviewers to weigh in on methodological quality.

Clinton Brass (Congressional Research Service)[6] commented on the pros and cons of incorporating evaluation policies into a statute. Although some practitioners believe it is useful to include them to ensure they remain part of the discussion, others believe the inclusion yields too narrow a focus. Brass gave an example of a tiered-evidence initiative that narrowly defined "evidence" (for internal and external validity) as primarily coming from impact evaluations—a point of controversy in the evaluation field.

Judy Gueron (president emerita, MDRC; member, steering committee) noted two gaps in current principles and practices: the way policies for evaluation are written in a "one-off" way that does not promote replication and the need for sharing knowledge and educating the public on the importance of evaluation.

Tom Feucht (National Institute of Justice) raised the issue of independence in terms of funding: When an agency places a requirement for evaluation in its policy, does that consequently yield the type of one-off evaluations to which Gueron referred? Conversely, how are other programs' evaluations funded when they do not have the same written

---

[6]Clinton Brass stated that his comments reflect only his views and not those of the Congressional Research Service.

provisions?

Bethanne Barnes (Office of Management and Budget [OMB]) noted a paper OMB wrote for the Commission on Evidence-Based Policy that discussed how funding structures can affect the development of a portfolio of evidence. Nightingale briefly described how DOL organizes prioritizes its funding streams. She said weighing the importance of evaluation against other mission-critical activities is a "professional balancing act."

PUTTING PRINCIPLES INTO PRACTICE: A BALANCING ACT

Moderator Gueron opened the session by noting that the key to protecting scientific quality is having a strong evaluation design and a strong team.

Rebecca Maynard (University of Pennsylvania; member, steering committee) stressed the importance of agencies "taking ownership" of an evaluation and fully understanding its purpose in order to define the appropriate strategy. She also noted the importance of estimating the net costs associated with the measured impacts of programs, policies, and practices.

Goldstein pointed out the difference between IES's peer-review methods (which screen studies prior to release) and MCC's methods (which encourage extensive peer review but release all studies): she can see value in each approach. Barnes agreed that there is a place for post release reviews in the discussion of scientific quality and said that OMB's clearinghouses conduct similar reviews of federal and nonfederal documents. She further noted that if an individual study is incorporated into a larger portfolio of work, a post study review can consider where the study fits into the bigger picture.

Considering that practitioners and politicians may be interested in more than just the bottom line, Gueron asked to what extent evaluations should include interpretations in order to produce more relevant and useful results. Grossman asserted that it is not fair to taxpayers if a study costs a significant amount of money and only provides limited answers. Furthermore, evaluators would likely appreciate being given the leeway to explore the mechanisms behind results—as long as they can delineate between which aspects of an evaluation are confirmatory and which are explanatory. Whitehurst agreed about the usefulness of supplemental analyses and interpretations but countered that evaluations were frequently funded to answer specific questions, sometimes specified in statute, so that including supplemental and exploratory results along with interpretations of the meaning of the

findings exposes the agency to unintended political backlash. He argued for separating the primary products of an evaluation, which a federal evaluation agency has to own, from supplementary and interpretative derivatives, which can be carried out and published independently by third parties.

Gueron asked how one might handle a report in which the findings vary across outcomes, subgroups, time periods, or settings (perhaps not to a point of statistical significance) or deviate greatly from the expected results. Goldstein responded that it is important to highlight these differences, give the necessary caveats, and move forward to new research questions. **Miron Straf** (Virginia Polytechnic Institute and State University) said that the key is often in program implementation. He believes that agencies should encourage exploration and not constrain evaluators to stick too closely to an initial protocol, asserting that it is in this manner that one can really determine what works. Gueron said that transparency is key to convincing audiences of the credibility of an evaluation and ensuring neutrality in presentation. She asked the group to weigh in on the pressures of timing when balanced against a desire to release complete results. Whitehurst emphasized the need for schedules that are appropriate to the context—with neither too tight nor too distant a deadline—and also said it is important to make evaluation data available for secondary analysis. Rolston said that the practice of registering studies helps to enhance transparency. Maynard agreed that registering studies and laying out standards and expectations about evaluation methods and reporting can help things flow more smoothly.

Gueron next asked the group how federal agencies can reinforce independence in evaluations and protect against pressure to bias the selection of contractors or the reporting of results. Shroder noted that for HUD, the threat of "bias" is sometimes introduced by a requirement that the agency pick a small business contractor, which rules out several qualified evaluators. He added that there may be a tradeoff between an evaluator's independence and an agency's capacity to evaluate and learn. It may be important for a mission-oriented unit to evaluate itself and own its learning agenda.

Gueron then raised the issue of how far the concept of independence extends: Is a contractor seen as an extension of the agency? Does it undercut contractors' credibility if they work for an agency seen as partisan? With that in mind, how does one attract the best people to do the evaluation work? Whitehurst suggested that design competitions—in which the focus is not what work will be done, but how it will be done—can address this issue. Barnes, Rolston, and Neild all agreed that independence between federal agencies and contractors should not be viewed as an either-or situation but rather as a relationship that has to be managed throughout each project. Gueron said she has learned over time that there does not need to be a tradeoff between ethics and rigor, either. She noted the critique that rigorous random assignment is too demanding in certain contexts. Maynard said that if a method other than a randomized controlled trial is proposed to answer questions about impact or effectiveness, there needs to be a compelling justification. **Christina Yancey** (U.S. Department of Labor) pointed out that rigorous methods may overlook small or hard-to-sample populations. She believes the ethical approach is to study these groups, even if the data obtained do not meet a certain scientific standard.

Turning to funding, Gueron asked if there are ways to implement evaluation policies and practices that guard against political pressures tied to funding. **Constance Citro** (CNSTAT) stressed the need for qualified staff, noting that financial pressures often lead to hiring caps for staff. Goldstein said that getting high-quality staff within the constraints of the federal hiring system can be difficult, but that mobility of federal employees between government and contractors can help. Neild and Nightingale added that keeping staff engaged and encouraging them to continue their own research can help agencies to compete with contractors or academia.

MAKING THEM STICK: INSTITUTIONALIZING THE PRINCIPLES

William Sabol (Westat; member, steering committee) discussed his experience leading a federal statistical agency (Bureau of Justice Statistics), which relies on CNSTAT's *Principles and Practices for a Federal Statistical Agency*.[7] He asserted that those principles—relevance, credibility, trust, and a strong position of independence—are very similar to what is being discussed for federal evaluation. Sabol gave examples of how that volume addressed independence, which included (1) separation of the statistical agency from the parts of the department that are responsible for policy making and for law enforcement activities; (2) control over professional actions, especially the selection and appointment of qualified and professional staff; (3) authority to

---

[7]National Research Council. (2013). *Principles and Practices for a Federal Statistical Agency, Fifth Edition*. Committee on National Statistics, C.F. Citro and M.L. Straf, eds. Washington, DC: The National Academies Press.

release information without prior clearance and ad- herence to predetermined schedule of release; and (4) the ability to control information technology systems, tied largely to protection of data. Sabol disagreed with Grossman's point that the Paper- work Reduction Act had been a hindrance—it gives OMB the authority to coordinate and develop poli- cies for the 13 primary federal statistical agencies. In addition, OMB's creation of the Interagency Coun- cil on Statistical Policy and the 2002 Confidential Information Protection and Statistical Efficiency Act were important developments in promoting gov- ernment-wide statistical standards.

Barnes, speaking in her role as head of the OMB evidence team, noted that OMB's Office of Informa- tion and Regulatory Affairs (OIRA) recently issued Statistical Policy Directive Number 1 (also referred to as the trust directive), which essentially codifies the information in *Principles and Practices for a Fed- eral Statistical Agency.* She acknowledged, however, that evaluation functions do not have a similar type of overarching structure, in part because evaluation has developed more slowly and the nature of the structures in individual agencies has been so var- ied. She mentioned a recent report[8] that showed that agencies with centralized evaluation offices had broader evaluation coverage and greater use of evaluation data. OMB has informally created an Interagency Council on Evaluation Policy (of which Goldstein is cochair), which could be the basis for a formalized structure.

Sabol asked participants what external entities could do to help institutionalize evaluation principles and what evaluation agencies themselves can do. White- hurst commented that Congress could help by pro- viding OMB with the authority to oversee the pro- cess. Brass said that evaluation activities seem to be Balkanized both within and among agencies—evalu- ation versus performance management, applied re- search versus methods, etc.—and this situation could be addressed when considering how to institution- alize principles for evaluation. Rolston reminded the group how Congress has thwarted evaluation efforts and severely limited the use of randomized controlled trials at least twice in the past. Securing broad legislation on the use of evaluation may help institutionalize practices. Feucht said that the idea of using CNSTAT's *Principles and Practices for a Federal Statistical Agency* and OMB guidelines for statistical agencies as markers when considering a parallel structure for evaluation agencies could be a good

---

[8]U.S. Government Accountability Office. (2014). *Some Agencies Reported that Networking, Hiring, and Involving Pro- gram Staff Help Build Capacity.* GAO-15-25. Available: http:// www.gao.gov/products/GAO-15-25 [January 2017].

move towards institutionalization.

Nightingale reiterated Whitehurst's point about OMB's role, noting that it has encouraged confor- mity among offices by requiring evidence-based justifications for budget increases and clarifying that the term "statistical purposes" includes evalu- ation. DOL also includes a chapter on evidence in its strategic plan. **Christopher Walsh** (Department of Housing and Urban Development) suggested that the 24 agencies subject to the Chief Finan- cial Officers Act (P.L. 101–576) should use the re- quirement to include program evaluation in their strategic plans as an opportunity to institutional- ize evaluation principles. **Lauren Supplee** (Child Trends) asked if there could be any potential down- side to institutionalizing evaluation principles. Sabol thought not. **Daryl Kade** (Substance Abuse and Mental Health Services Administration) asked if the transition to a new presidential administration presents an opportunity for institutionalizing ma- jor evaluation principles. Sabol concurred that ev- ery administration presents a new opportunity. He noted that *Principles and Practices for a Federal Sta- tistical Agency* has been updated every 4 years: the 6th edition will be published in 2017. **Sandy Davis** (Bipartisan Policy Center) said there appears to be congressional interest in improving evaluation, on both sides of the aisle, noting the 2016 Evidence- Based Policymaking Commission Act.

Regarding ethics, Sabol asked about whether there should be any constraints on staff who do both evaluation and other scientific work and about potential conflicts of interest because of external partnerships. Feucht said that the nature of grants results in a wide range of relationships between programs and external entities. Shroder noted that some agencies do not permit professional staff to publish their research without approval, but he be- lieves they should be free to publish so long as they clarify that their opinions are their own.

CHEERLEADERS, NAYSAYERS, LARGE AND SMALL EVALUATORS: FOSTERING SUPPORT AND INCLUSION

Maynard opened the session by noting that while there are several stakeholder groups who have a vested interest in developing a principles and prac- tices document for evaluation, there are others who would not consider it such a good idea. Regardless, she believes evaluations should be designed with the expectation that results may be positive, neu- tral, inconclusive, or negative.

Jon Baron (Arnold Foundation) believes that any guidance document should take on a "less-is-

more" approach, highlighting a few key principles and making a persuasive case for each. He does not think it beneficial to try and cover the whole landscape of evaluation in a single document, nor does he think that a long document would be read. He mentioned IES and National Science Foundation's *Common Guidelines*,[9] which could potentially serve as a starting point for an evaluation policy document. He said that a central goal of evaluation efforts should be to grow the body of interventions that are backed by credible evidence. Shroder raised the concern that evaluation funding and review of information requests are still closely controlled by Congress and OIRA, respectively. Whitehurst believes that, because of that control, a document of this nature should be produced by a foundation or similar organization (as opposed to a direct stakeholder), be addressed directly to Congress, and take the form of proposed legislation. Gueron agreed that foundations frequently fund activities that the federal government is less likely to fund—communication and dissemination, for example—but they are less concerned with exploring new learning, often taking a position that "we already know enough." Rolston added that support from a key player like OMB could bolster the acceptance of the principles.

Sherry Glied (New York University) asked about the issue of magnitude and power for some of the smaller evaluation agencies and what to do when a program or budget is not big enough to support a desired study. Would smaller experiments be acceptable? Should quasi-experimental analysis be used routinely and be supplemented by randomized controlled trials once evidence accumulates? Maynard said, yes, there is benefit in conducting smaller experiments in these cases. Nightingale cautioned the participants that any evaluation policy document needs to be applicable to the variety of agencies trying to build evaluation offices. In response to a query from Maynard, Nightingale explained that smaller agencies are represented in cross-agency evaluation groups that OMB convenes. **Jeff Dowd** (U.S. Department of Energy) echoed Maynard's concern, cautioning the group not to forget about smaller agencies with decentralized evaluation offices.

Feucht identified three groups that might be opposed to a principles document: program managers with the "I tried it and it works" philosophy; practitioners who may see an investment in evalua-

tion as detracting from direct services; and smaller agencies that may be underrepresented in a push towards randomized controlled trials. Baron replied that a response to this would be to focus on evaluating components of a program rather than the entire program—for example, looking at preschool interventions rather than the entire Head Start Program.

WRAPPING UP

Whitehurst drew the group's attention back to the scope of the workshop: evaluation of federal programs intended to affect human behavior. He added that U.S. taxpayers make a decision to fund those programs with the goals of improving opportunities and reducing identified problems and that failure to use their money in a way that can contribute to those goals is a disservice to them. While the evaluation principles that are currently in place are very sound, he argued that legislation is needed for permanence and stability. He sees the legislation taking one of two forms: (1) an agency-by-agency approach that supports the creation of independent research and evaluation offices and affords them protections and statutory guidelines, or (2) aligning with the Paperwork Reduction Act, creating separate legislation and giving OMB some general authority over this function, similar to what is in place for the statistical agencies. He acknowledged that funding had been addressed several times in the workshop and said that budgets for evaluation would need to be included in the legislation. Whitehurst reiterated the importance of peer review for holding the producers of the work responsible for its quality, and reminded participants of OMB's prior practice of rating the quality of evaluation efforts as another accountability measure. Baron added that the proper use of peer review and such techniques as specifying confirmatory versus explanatory hypotheses are important.

Gueron noted the earlier discussion on the tension between focusing on rigor and making evaluations useful. Whitehurst reinforced Baron's point about the value of evaluating program components as a way to mitigate that tension. **Miron Straf** (Virginia Tech) disagreed, stressing what he sees as a need to move away from a myopic approach of focusing on an effect size of a single intervention to look at social programs as part of a complex system. Goldstein concluded by noting that while peer review can be valuable, it is a practice, rather than a principle, and falls under the larger umbrella of quality control—one of several very important principles to be considered.

---

[9]Institute of Education Sciences and National Science Foundation. (2013). *Common Guidelines for Education Research and Development.* Available: https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf?WT.mc_id=USNSF_124 [January 2017].

For additional information regarding the meeting, visit nas.edu/Principles-for-Federal-Program-Evaluation.

Division of Behavioral and Social Sciences and Education

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

The nation turns to the National Academies
of Sciences, Engineering, and Medicine for
independent, objective advice on issues that
affect people's lives worldwide.
www.national-academies.org