



Comments on Slavin

Becoming Critical Readers: Issues in Transparency, Representation, and Warranting of Claims

Judith L. Green and Audra Skukauskaitė

This response to Slavin's article (2008) explores the issues of transparency, representation, and warranting of claims in Slavin's descriptions of the work of others and his suggestions for program evaluation syntheses. Through contrastive analyses between Slavin's representations of the program evaluation synthesis efforts of five organizations and their own representations of their synthesis work, the authors uncover the ways in which the gaps in Slavin's text become consequential for readers, limiting readers' opportunities for developing understandings of synthesis efforts. Through contrastive analyses, the authors foreground the need for transparency in reporting, raise questions about the use of narrowly focused reviews, and demonstrate how a contrastive analysis of texts may constitute a productive approach to critical reading and an opportunity for learning.

Keywords: contrastive analysis; critical reading; ethnographic perspective; logic of inquiry; transparency

Robert E. Slavin's article "What Works? Issues in Synthesizing Educational Program Evaluations" (this issue of *Educational Researcher*, pp. 5–14) outlines four goals to reach a broad audience for reading, understanding, and doing program evaluation syntheses. Slavin sets out to reach this audience by discussing what he argues are key issues in synthesizing research on educational programs, by bringing forward and briefly contrasting the methods used in the major synthesis efforts, and by proposing solutions to methodological problems inherent to syntheses of program evaluation research. In articulating his goals in this way, Slavin defines his audience as people both within and outside of his tradition of synthesis work in program evaluation research. He views his audience, therefore, as readers of *Educational Researcher (ER)* who may or may not be familiar with the work on which he draws or with the research traditions he references.

As members of this broad audience in education—teachers of research methods and researchers of teaching and learning processes in diverse classrooms—we approached this article with the goal of expanding our understanding of the growing area of program evaluation research in general and of learning

more about his approach specifically. Our goal, therefore, was to explore the conceptual arguments and information Slavin provides so that we could represent his arguments to our students in ways that are consistent with his tradition of research and research synthesis, given that his perspective differs from our theoretical, philosophical, and research traditions.

The approach we used to uncover his arguments in a systematic way was guided conceptually by ethnographic and discourse analysis perspectives (e.g., Bloome, Carter, Christian, Otto, & Shuart-Faris, 2005; Green & Bloome, 1997). As ethnographers who focus on discourse and the social construction of knowledge, we seek to understand how different groups engage in their work and to make visible what members of the social group need to know, understand, and do in order to participate in a local community in socially appropriate ways (Agar, 1994; Castanheira, Crawford, Dixon, & Green, 2000; Heath, 1982). Guided by this perspective, an additional goal in reading this article was to gain understandings of how to engage in program evaluation research and research synthesis from the point of view of the organizations identified by Slavin as well as from Slavin's perspective in his article.

In making visible our stance to reading Slavin's article, we seek to contribute to a discussion of how to meet the reporting standards of the American Educational Research Association (AERA, 2006) and Slavin's goals for transparency in conducting and publishing empirical social science research. Building on Slavin's goal for creating critical readers, we selected the issue of transparency in reporting research as the focus of our response in order to explore what information is needed to read his article critically. Given that Slavin provides a contrast of the work in program evaluation research syntheses across five organizations, we welcomed this opportunity to explore his article and learn from the information he provides what he counts as a critical reading of synthesis work as well as solutions to "methodological problems inherent in syntheses of program evaluations" (p. 6).

If we were to accept Slavin's arguments and suggestions as a basis for critically reading the work of synthesis across organizations, then as readers we must accept his characterization of the work of others. As our analysis will demonstrate, this was not possible, given that as we read we identified gaps in information in the text about the organizations. By uncovering the kinds of knowledge that we needed to become critical readers of Slavin's text, we make visible gaps in the text that led us beyond

Educational Researcher, Vol. 37, No. 1, pp. 30–40
DOI: 10.3102/0013189X08314828
© 2008 AERA. <http://er.aera.net>

Slavin's article to texts of the organizations whose approaches he critiques. Through his process, we identify a range of information necessary to understand the broader context he sought to represent. Specifically, we focus on how he represents the work of others in this rapidly developing body of reviews. In this response, therefore, we make visible how the brevity of descriptions Slavin provides in his contrasts of the five organizations he identifies as leading sites for evaluation syntheses constitutes a lack of transparency that created challenges for us and thus potentially for other readers who do not share his theoretical or methodological framework.

In our response, we present an analysis of our reading of the perspectives on program evaluation research and synthesis across organizations and examine issues of transparency, or limits to transparency, in Slavin's contrasts of the work of others. We view this response as a turn in the hermeneutical conversation on epistemology (Kelly, 2006) with both Slavin and readers of *ER*.

Why Transparency? Creating Transparency for a Critical Reading of Slavin's Contrasts

To explore issues of transparency, we build on AERA's goals of promoting transparency of research methods and of making visible the logics of inquiry that guide reporting of empirical research studies. The Standards for Reporting on Empirical Social Science Research in AERA Publications (AERA, 2006)¹ states that transparency is achieved when authors clearly present a statement of the problem formulation, the sources of evidence reported, and a rationale for issues of measurement and classification. Given these guidelines, we focus on identifying and analyzing sources that would provide additional information about the ways the organizations that Slavin contrasts address these three areas of transparency.

The sites reviewed were the What Works Clearinghouse (WWC) of the Institute of Education Sciences (IES) funded by the U.S. Department of Education within the U.S. government (<http://ies.ed.gov/ncee/wwc/overview>); the Best Evidence Encyclopedia (BEE; <http://www.bestevidence.org>) of the Center for Data-Driven Reform in Education (CDDRE), a U.S. Department of Education-funded research center at John Hopkins University (<http://www.cddre.org/>); the Comprehensive School Reform Quality (CSRQ) Center at the American Institutes for Research (<http://www.csrq.org>); the Campbell Collaboration (C2), a largely voluntary international organization (<http://www.campbellcollaboration.org>); and the Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre of the Social Science Research Unit at the Institute of Education, University of London, funded in part by the Economic and Social Research Council in the United Kingdom (<http://eppi.ioe.ac.uk/cms>).

The overarching question that guided our work in exploring the texts on the five sites was, What kinds of goals and criteria for synthesis work are available to readers of the website of each organization? To identify patterns within and across the texts collected from organization websites, we constructed a set of semantic relationships (Spradley, 1980) that permitted us to uncover patterns of cultural knowledge and assumptions about the work of synthesis held by members of each research synthesis organization. Four semantic relationships (cf. Spradley, 1980) were used to make transparent different areas of information we needed to engage in the contrastive analysis across sites.

- X is a source of literature for potential inclusion in synthesis.
- X is a way of collecting sources of inclusion.
- X is a type of review conducted.
- X is a type of inclusion criterion.

Using these semantic relationships to guide our critical reading of each site and grounding our analysis in the discourse used by members to represent their goals, purposes, criteria, and processes of research and synthesis, we were able to identify the actions and processes involved in the work (Agar, 1994; Lakoff & Johnson, 1980). The analysis of each site was then placed in matrices in order to construct cumulative records of inscribed practices and goals across the five organizations conducting systematic major synthesis efforts (e.g., Miles & Huberman, 1994). We then used the matrices to conduct contrastive analyses across the sites. The initial level of analysis included contrasting the kinds of literature considered for inclusion in syntheses by each site. The range of literature we identified, and the ways in which this literature was used in syntheses, led us to question Slavin's critiques of the sites and subsequently to reexamine the contrasts Slavin inscribed for the five organizations. Our response, therefore, is grounded in an empirical analysis of texts and intertextual relationships (e.g., Bazerman, 2006; Bloome, 1992) across organizations, which laid a foundation for critically reading and thus assessing Slavin's representations of the work across organizations.

As our analysis in the next section will show, this step enabled us to uncover a series of inconsistencies in the way Slavin represents the approaches taken by these organizations. After uncovering gaps in Slavin's representation of the work of the organizations he cites, which we viewed as a problem of transparency in his critiques of their research methods and designs, we then returned to websites of the five organizations to explore in more detail the synthesis criteria used by these sites. This sequence of iterative processes of analysis, anchored in Slavin's article and extending to information beyond what he provides, enabled us to identify information needed to fill in the gaps in the key arguments in his article. Furthermore, as our analysis demonstrates, uncovering these gaps raises challenges not only for readers but also for those whose work is being represented.

One caveat needs to be presented in the spirit of transparency. The data we identified for inclusion in the matrices that we used for contrastive analysis illustrate information available and may not represent the full range of work undertaken, given the complex layers of information on each website. Therefore, we view the information provided as an illustrative, not as a comprehensive, review of the scope of work of each organization. That level of analysis, given our ethnographic perspective, would require a more comprehensive form of triangulation (Corsaro, 1981) involving interactions with the organizations to identify the fit between our representations of their work and their own cultural understandings of such work (Green, 1983). Given the short time frame for responding, this step was not possible.

Contrasting Ways of Collecting Literature for Potential Inclusion in Syntheses: What Is Relevant to Whom?

The first contrast focuses on the ways of collecting literature that the sites stated they used to identify potential studies for

Table 1
X Is a Way of Gathering Literature for Potential Inclusion in Synthesis

Literature-Gathering Criteria	U.S. Only			International	
	What Works Clearinghouse	Best Evidence Encyclopedia	Comprehensive School Reform Quality Center	Campbell Collaboration	Evidence for Policy and Practice Information and Co-ordinating Centre
Statement about how literature was gathered	Yes	Yes	Yes	Yes	Yes
Include a broad range of sources	Yes	Yes	Yes	Yes	Yes
Gathered by organization staff (researchers)	Yes	Yes	Yes	Yes	Yes
Collected by teams consisting of members with different expertise	Yes: primary investigator, project coordinator, research analysts	Does not specify who conducts synthesis	Yes	Yes	Yes
Submitted to organization by members of the field and the public	Yes	No	No	Proposal for a synthesis or review submitted	No
User involvement in selecting topics and literature for review	No	No	Yes	Yes	Yes
Verified by participants in program for the organization	Not specified. Depends on effect size as criteria?	Not specified. Depends on effect size as criteria?	Interviews at sites	Yes, at topic selection stages	Reviewed by those involved in identifying area of needed synthesis

inclusion in their syntheses. This contrast is one that is directly related to Slavin’s arguments and is based on our examination of the semantic relationship that X is a way of gathering literature for potential inclusion in synthesis. This contrastive analysis sets the frame for grounding additional steps we undertook to understand the differences and similarities in practices identified across the five organizations. In reading critically (i.e., analyzing texts) across the five websites, what became visible was the great variation in the kinds of studies collected as potential texts to be used in reviews and the ways in which they were collected, screened, and used for program evaluation research syntheses.

Each website foregrounds information about the processes involved in collecting and screening literature; however, there are significant differences in the scope of work undertaken by the different organizations and in the specificity about the ways the work of synthesizing is accomplished. Table 1 illustrates the ways the five organizations state that they gathered the research studies for potential inclusion in program evaluation research syntheses and subsequent program ratings.

As indicated in Table 1, the contrastive analysis of literature collection practices across the five sites led to the identification of a broader set of issues about ways of collecting literature than an analysis of any one site would have permitted. All five organizations argue that any program evaluation synthesis should make transparent how the literature was gathered and screened and that studies for potential inclusion should be collected from a broad

range of sources. They argue that such information is needed so that the audiences interested in their findings, for example, funders, educators, and policy makers, among others, can make informed decisions about the program or area reviewed.

These arguments led us to pause our analysis of the information in Table 1 to examine the search strategies used across the sites to locate research studies in scholarly journals, electronic databases, web-based repositories, the Internet, specialist websites, dissertations, hand searches of journals, reports, and other published and unpublished materials. While foregrounding different kinds of sources, all five organizations have a common goal—to uncover the broadest scope of potential sources for the systematic and comprehensive review. This information is presented in Table 2. This analysis of the differences in strategies of identifying literature raised questions for us as readers about the scope of work represented in a synthesis and the need for transparency at the level of literature search.

Having identified the types of searches undertaken and the differences among the ways of conducting searches, we then returned to Table 1 to examine who engaged in the processes of identifying potential articles for inclusion and, in some instances, who was involved in setting the research agenda. Although the literature in all organizations is identified and gathered primarily by the researchers in the organization (or associated with the organization, as in C2), there are marked differences in the scope and kinds of participation by those outside the organizations. This

Table 2
X Is a Kind of Source for Research Studies Explicitly Mentioned on the Websites

Source of Research Studies	U.S. Only			International	
	What Works Clearing house	Best Evidence Encyclopedia ^a	Comprehensive School Reform Quality Center	Campbell Collaboration	Evidence for Policy and Practice Information and Co-ordinating Centre
Published research	x	x	x	x	x
Journals	x	x	x	x	x
Unpublished research	x	x	x		x
Dissertations		x			
Government and university reports		x			
Public/users	x		x		x
Program developers			x		
Within program studies		x			
Personal contacts		x		x	
Reference databases		x		x	x
Hand searches of journals		x		x	x
Web-based repositories		x		?	x
International sources				x	

^aFor Best Evidence Encyclopedia, categories marked with an “x” are inclusive of Slavin’s earlier work (Slavin, 1986, 1995; Wasik & Slavin, 1993). The website does not provide this information; under Methods it only lists an earlier version of the article reviewed here.

analysis led us back to Slavin’s text to examine how and in what ways he acknowledges these differences.

In his article, Slavin argues that diversity becomes confusing to the readers of syntheses and thus argues for a limited set of criteria, or what he calls suggestions, for creating a common standard to make the research syntheses more understandable and reliable to those seeking information about program effectiveness. However, as we found in our search of the different sites, organizations were concerned about diversity but approached their responses in markedly different ways, ones that viewed readers as needing transparency in review practices rather than a single standard. The following quote from the EPPI website captures succinctly alternative ways of making visible the credibility of the sources and the processes involved in synthesis:

Whatever the approach to searching, it is important that reviewers keep a “search log” to record the detail of how searching was undertaken. For example, which journals, websites and databases were searched and how and when, along with the list of search terms used and the combinations in which they were applied to the databases. Enough detail should be reported so that the search process could, in theory, be replicated. Keeping a record of all this information will make the review methods explicit and transparent in the final report, so that readers can make a judgment about the quality of the reviews findings. (EPPI-Centre, 2008)

This argument leaves the decision about credibility and scope of information afforded readers in the hands of those seeking the information and in their ability to assess the quality of the information provided. This approach does not rule out Slavin’s recommendations for reporting effect size in studies with randomized, quasi-experimental, and matched designs. Rather, underlying the EPPI argument for transparency is the view that readers can assess the quality of different approaches in order to make informed decisions about programs.

The differences in reviews identified in the EPPI argument suggested the need to examine the range of reviews conducted across sites and how the differences in kinds of reviews shape what those interested in improving the quality of education have available through these different sites. We viewed this process as necessary to further contextualize the critique that Slavin makes about what other organizations do or do not do. Table 3, showing the kinds of reviews the organizations do or advocate, represents the range of review approaches we identified.

The differences in types of reviews undertaken by each organization reflect differences in what counts as evidence to members of the organizations. It also speaks to the variety of information available through reviews, because each approach provides a particular level of information and layers of analysis of evidence about programs. What makes this a telling case (Mitchell, 1984) are not only the differences in types of reviews but the fact that all of the sites provide criteria for transparency of their work and all detail the basis for selection of articles. One way of understanding the similarities and differences is to see the range of review forms as signaling different values about what counts as evidence across these organizations. It also raises questions about the question–method relationships each uses and, through this, what types of information are made visible or invisible by the particular type of review promoted or expected by each organization.

These questions about what counts as evidence within and across traditions of work relevant to program evaluation are not ones that we ask solely as a result of this review. Underlying the questions we raise are recent conceptual reviews on the international level, including the 2007 *Yearbook of the National Society for the Study of Education*, “Evidence and Decision Making,” edited by Pamela Moss (2007); Moss, Girard, and Haniford’s (2006) chapter, “Validity in Educational Assessment,” in the *Review of Research in Education* (Green & Luke, 2006), and chapters in AERA’s

Table 3
X Is a Kind of Synthesis or Review: Ways Sites Describe the Kinds of Reviews They Do/Advocate

Type of Review	U.S. Only			International	
	What Works Clearinghouse	Best Evidence Encyclopedia	Comprehensive School Reform Quality Center	Campbell Collaboration	Evidence for Policy and Practice Information and Co-ordinating Centre
Systematic review	x	x	x	x	x
Statistical meta-analysis	x	x		x	x
Quantitative synthesis				x	
Traditional review		(in Slavin, 1986)			
Best evidence synthesis		x			
Qualitative meta-synthesis				x	
Narrative review				x	
Narrative conceptual synthesis					x
Meta-ethnographic synthesis					x
Quantitative/qualitative integrated syntheses across schools in comprehensive school reform			x		
Commissioned reviews				x	
Invitational reviews				x	
Unsolicited reviews				x	

Handbook of Complementary Methods in Education Research (Green, Camilli, & Elmore, 2006) by Kushner and Adelman (2006) on program evaluation and McNeil and Coppola (2006) and Goertz (2006) on policy analysis. All of these publications raise issues related to program evaluation research and point to the need to examine how different layers of analyses may be needed to identify, evaluate, and understand the complex factors that support and/or constrain program effectiveness and its relationships to student learning, not just achievement measured on tests.

Contrasting the Criteria for Research Included in Syntheses: Convergences, Divergences, and Gaps in Slavin’s Article

The analyses presented above identify a common goal of transparency across the five organizations but a difference in how transparency is accomplished and in what types of reviews are undertaken. In this section, we present a closer examination of specific dimensions of the review process, drawing once again on AERA’s (2006) Standards for Reporting on Empirical Social Science Research in AERA Publications. The particular standards we used for this contrastive analysis focus on formulating the problem, research design features, and reporting of outcomes as an issue of measurement and classification. This contrastive analysis provides a means of identifying how the suggestions in Slavin’s article for conducting major syntheses of program evaluation research converge or contrast with the criteria of the four other organizations Slavin critiques as not meeting the criteria he proposes. This section, therefore, considers both the commonalities and differences across the five organizations and, where relevant, brings forward Slavin’s contrast in order to assess the accuracy of his representations.

Table 4 represents a cumulative profile of criteria for inclusion of research identified across sites and serves as a data set for contrastive analysis of what counts as evidence for inclusion of studies in syntheses of program evaluation research. The information presented is the result of searches of selected terms on the websites, searches that often took us to linked websites that contribute to, or partner with, each organization.

As indicated in Table 4, in the section Problem Formulation, in discussing the scope of work, purposes, and methodologies for research syntheses, the WWC and the BEE do not address the issue of problem formulation, making assumptions that the topics selected for syntheses of program evaluation are important and will be relevant to educational researchers, practitioners, and policy makers. CSRQ, C2, and EPPI, on the other hand, include extensive discussions about how the problems to be addressed in reviews are conceptualized, selected, and contextualized. CSRQ, whose work is focused on schoolwide reforms, uses the criteria of the U.S. Department of Education to select the reform program to be reviewed (<http://www.csrq.org/aboutcsr.asp>). Meanwhile, C2 and EPPI provide explicit guidelines for including the statement of the problem, locating it historically and in the research field, providing contextual information about the program, and stating the significance and relevance of the research problem and/or the program being reviewed.

The issue of transparency of problem formulation and program contextualization is directly related to the kinds of research design features that determine which research studies are included in the pool of literature used in the synthesis. As illustrated in the second cluster of Table 4, Research Design, we found that all organizations include randomized, experimental, and quasi-experimental designs with comparison and control

Table 4
X is a Kind of Criterion for Inclusion and Exclusion of Research Studies in Syntheses

Screening Criteria	U.S. Only			International	
	What Works Clearinghouse	Best Evidence Encyclopedia	Comprehensive School Reform Quality Center	Campbell Collaboration	Evidence for Policy and Practice Information and Co-ordinating Centre
Problem formulation					
Topic definition	Yes	Implied	Yes ^f	Yes	Yes
Statement of research problem	Not specified	Not specified	Yes	Yes	Yes
Locate question in history or field	Not stated	Not stated	In history	Yes	Yes
Statement of significance	Not stated	Not stated	Reform effort	In proposal	In review ^e
Research design					
Minimum duration of study	Not specified	12 weeks	1+ year	Not specified	Not specified
Relevant student sample	Yes	Yes	Assumed	Yes	Yes
Randomized design	Yes	Yes	Yes ^b	Yes ^a	Yes ^b
Experimental designs	Yes	Yes	Yes	Yes ^a	Yes ^b
Quasi-experimental design	Yes	Yes	Yes	Yes ^a	Yes ^b
Comparison groups	Yes	Yes	Yes	Yes ^a	Yes ^b
Control group	Yes	Yes	Yes	Yes ^a	Yes ^b
Qualitative research design	No	No	Yes ^a	Yes ^a	Yes ^b
No attrition problem	Yes				
No disruption problem	Yes				
Reporting of outcomes					
Targeted outcome evaluations	Yes	Yes	Yes	Yes	Yes
Achievement outcomes	Yes	Yes	Yes	Yes	Yes
Multiple outcomes ^d			Yes	Yes	Yes
Statistical conclusion validity ^c	Yes	Yes	Yes	Yes	Yes
Intervention fidelity	Yes				
Formative intervention			Yes		
Evidence of causal validity	Yes		Yes		
Statistical outcome measures	Yes	Yes	Yes	Yes	Yes
Effect sizes	Yes	Yes	Yes	Yes	Yes
Program ratings	Yes	Yes	Yes		

^aDepends on proposed review.

^bNot exclusive.

^cStatistical measures, including sampling variance estimated appropriately; key statistical tests are verifiable.

^dFor example, graduation rates, family or community involvement, student attendance.

^eAs defined by advisory groups and researchers.

^fEvaluation criteria for school reform model set by the U.S. Department of Education.

groups, despite Slavin's critiques to the contrary. What Slavin does not make transparent in his critiques of others is that although the WWC and the BEE call for randomized and matched designs as the primary kinds of research to be included in syntheses, the other three sites include those designs but not as exclusive kinds of studies that count as important for understanding and evaluating educational programs and their effects. In addition to the designs advocated by Slavin, CSRQ, C2, and EPPI all consider qualitative research as important in synthesis and program review efforts. This inclusion of qualitative research is directly related to the goals of these three organizations to provide information that feeds back into the program or school reform (CSRQ) and that can be used by decision makers at individual, local, district, state, and other levels. The omission of qualitative data raises questions about how the outcomes of the types of reviews supported by the WWC and the BEE can be used by administrators at the district and school levels and by teachers

and parents to decide if the program is one that fits their populations, fits within their educational goals, or will support student achievement in their local contexts.

The last cluster in Table 4, Reporting of Outcomes, makes visible what kinds of outcomes are expected, measured, and reported in program evaluation syntheses across sites. This cluster of information makes visible that what counts as an outcome to be reported is related to the ways the different organizations formulate the problem and make choices about primary research designs to be included. Although all five organizations engaged in synthesis work include outcome evaluations relevant to the specified topic, provide statements about the adequacy of statistical measures, and estimate effect sizes (exclusively or as part of other estimates), the outcomes reported differ. All organizations include student achievement outcomes in evaluating research on particular research programs, but the last three organizations, CSRQ, C2, and EPPI, which are more directly responsive to local

Table 5
Slavin's Descriptions and Critiques of the Work of Others

Site	Synthesis Criteria for Each Site	Slavin's Critique
What Works Clearinghouse (WWC)/Institute of Education Sciences	Randomized experiments; high-quality matched/controlled quasi-experiments	(1) Duration required varies; (2) slow pace; (3) emphasize "many small, brief experiments"; (4) rejected a study (by Slavin) "with little potential to bias outcomes" based on "technicality" (grades of the program did not match exactly); (5) focus on randomized studies without attention to sample size and other design elements that also have potential to introduce bias can lead to illogical conclusions; (6) WWC specifies its inclusion and synthesis procedures in great detail, but in practice it allows considerable variation from one topic area to the next on key issues, such as the minimum study duration required for inclusion; (7) both the WWC and the BEE require at least one randomized experiment with a positive effect for a program to receive the highest rating. However, the BEE allows this to be a randomized quasi-experiment (RQE), whereas the WWC does not—instead, the WWC recomputes analyses in RQEs to control for clustering, which almost invariably makes analyses nonsignificant, regardless of effect sizes or sample sizes; (8) potentially, the WWC is the most influential of the synthesis efforts for policy because it alone carries the endorsement of the U.S. Department of Education
Best Evidence Encyclopedia (BEE)	Inclusion criteria similar to WWC; 12-week program duration	
Comprehensive School Reform Quality Center (CSRQ)	Review methods quite different from those of the WWC and of the BEE; emphasis on numbers of studies and statistical significance rather than randomized evaluations and effect sizes	Does not look at effect sizes, like the BEE does; the CSRQ emphasizes the number of statistically significant positive results; it does not take random assignment into account
Campbell Collaboration	Strong emphasis on randomized controlled trials	Randomized studies are few in number, and many are very small, very brief, very artificial, and/or very old; given the increasingly common finding that in studies in education randomized and well-matched studies tend to produce similar effect sizes (see Torgerson, 2007), the rationale for restricting attention to randomized studies alone is diminished
Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI)	Researchers come up with own standards	EPPI reviews vary widely in breadth, focus, and methodology; the focus is on variables rather than programs

sites, include additional outcome measures such as graduation rates, family or community involvement, student attendance, and teacher and student learning.

On Groundings of Critiques: Uncovering Issues of Warranting Claims

When we took Slavin's critiques of the four organizations out of the context of the article and clustered them by organization, as indicated in Table 5, Slavin's Descriptions and Critiques of the Work of Others, we were able to make visible an overarching argument that he crafted through these critiques. In this section, we demonstrate how we analyzed Slavin's critiques of others by examining the order in which they are presented and then by considering the weight of the critique given the number of times a particular organization is the subject of Slavin's critique.

In describing the synthesis work of each of the five organizations, Slavin first introduces the work of the WWC, stating that "the WWC reviews all emphasize randomized experiments but include high-quality matched quasi-experiments in a lower category" (p. 6). Although a positive statement, given Slavin's proposed criteria, this statement is surrounded by two sentences that criticize the work of the WWC—the first a critique about the variation in requirements and the second a critique about the WWC's slow pace and "inability to meet its own expectations":

The WWC specifies its inclusion and synthesis procedures in great detail, but in *practice* [italics added] it allows considerable variation from one topic area to the next on key issues, such as the minimum study duration required for inclusion. All of the WWC reviews emphasize randomized experiments but include high-quality

matched quasi-experiments in a lower category. The WWC has suffered from an inability to meet its own expectations in terms of completion of reviews. (p. 6)

The placement of the positive statement between the two negatives in Slavin's article had the impact for us of calling into question the procedures specified on the WWC website. However, our analysis of the WWC website led us to a series of technical working papers (<http://ies.ed.gov/ncee/wwc/twp.asp>) designed to make transparent to reviewers and users or readers what counts as an eligible study or an appropriate process in conducting a review. The existence of these papers led us to question Slavin's critique, given that they specify conditions under which a study might vary from the randomized experiments and how equivalence is to be assessed. The level of specificity in these working papers suggests that they have a force of being a standard, rather than a practice, for which Slavin critiques the WWC. Slavin's statement might lead readers to conclude that there are not well-defined rationales for practices used.

The challenge to the WWC's process is further visible in how Slavin positions his own organization, the BEE, in relationship to the WWC:

[The BEE] initially intended to use the WWC as its source of information on proven programs, but because of the WWC's slow pace, CDDRE researchers created their own set of reviews, using standards and procedures similar to those of the WWC. At this writing, the BEE has completed reviews of elementary math, middle and high school math, middle and high school reading, and reading programs for English language learners. Its website contains links to reviews by the CSRQ on comprehensive school reform and other reviews on several topics. The BEE includes easy-to-read "educator's summaries" of reviews, both those written by CDDRE staff and those written by other reviewers. (p. 6)

In this statement, Slavin positions the BEE as more efficient and effective in conducting reviews. However, when we considered this statement in light of the contrastive analyses presented previously, we grew concerned, because the BEE's website has limited transparency of processes used to identify and select articles for inclusion for each of these reviews, and the website does not specify the problem being examined.

This concern led us to examine where we might find Slavin's conceptual arguments for the kind of synthesis he advocates in his article. Through a search on Google Scholar, we were able to identify an argument that was more situated in the history and methodology of the field. In an article published in 1986 in *ER*, Slavin argued that the "best evidence synthesis" he was proposing was an alternative to both "traditional reviews" and "meta-analyses" and solved the problems of both. This article, however, was not available on the website for either the BEE or the CDDRE, making the guidelines all but invisible without added search beyond the sites.

Moreover, we saw Slavin's critique of the other three organizations—CSRQ, EPPI, and C2—as a further attempt at positioning the BEE as *the* standard. These three organizations are introduced in the article through a contrast of what they do and how that does not meet the suggestions Slavin proposes. For example, he states that CSRQ's "review methods quite different

from those of the WWC and of the BEE, emphasizing numbers of studies and statistical significance rather than randomized evaluations and effect sizes" (p. 6). In stating this contrast, Slavin downplays the importance of CSRQ's review and synthesis work and how it meets the goals of the U.S. Department of Education, which did not impose a single measure of effectiveness and required that the program reviews and syntheses from such reviews include other measures, including attendance and community or family participation, among other outcomes of the school reform models being assessed. Thus, Slavin's discursive choice (Ivanič, 1994, 1998) to critique CSRQ for what it does not do, instead of asking about the appropriateness of what it does do to reach the goals, purposes, and requirements for its program set from inside the organization and from the funding agency, can be viewed as a missed opportunity for learning from CSRQ about factors other than student achievement that may influence students' long-term success in school.

Furthermore, given that C2 and EPPI are treated similarly to CSRQ, in that Slavin only briefly mentions them and focuses on critiquing them instead of making transparent the kind of work they do or how the issues Slavin foregrounds are contextualized in those organizations' work, we argue that Slavin missed additional opportunities to learn from these organizations about international conditions of education that might possibly have led to additional insights into the work of the CDDRE and the BEE. An example of a missed opportunity for learning from the international group of partners that are part of C2 can be seen in how he critiques C2's focus on "randomized controlled trials." His critique focuses on C2's inclusion of small, brief, and artificial trials while excluding "well-matched studies" that can also yield effect sizes similar to those of randomized studies. Although this critique may be fair given his criteria and approach to effect sizes, we leave assessment of this to those who know the effect size literature and are more conversant with the norms for randomized experiments and other statistically based research designs. What his critique signaled to us was that, once again, Slavin is suggesting that there is (or should be) a single standard for synthesis work, which can ignore the webs of related systematic work undertaken by reviewers such as in the case of C2. Without considering how and when a randomized design is used and what may also be present in an evaluation of a program, Slavin once again misses an opportunity to explore how the designs he advocates might be complemented by other designs and how achievement effects may be part of other effects, ones missed by his approach and its tightly focused view of best evidence.

Slavin's presentation of the last organization, EPPI, is very brief and states that the center

commissions a wide range of reviews on programs in many areas of education, such as science education, English teaching, and citizenship education. The Department of Children, Schools, and Families funds groups of reviewers to work in each area and allows them to come up with their own standards; thus EPPI's reviews vary widely in breadth, focus, and methodology. (p. 6)

By ending the introduction of EPPI with a statement that "most EPPI education-related reviews focus on variables (e.g., effects of grammar teaching on writing) rather than on specific programs" (p. 6), Slavin seems to dismiss EPPI's work because it is

not directly relevant to the discussions about syntheses he is presenting in the article. In making this claim, he ignores the possibility that the effects of grammar teaching on writing performance he gives as an example of a variable may involve assessment of student achievement in the context of teaching. In his article, Slavin defines a program as

any set of replicable procedures, materials, professional development, or service configurations that educators could choose to implement to improve student outcomes. A program is distinct from a variable in consisting of a specific, well-specified set of procedures and supports. Class size, assigning homework, or provision of bilingual education are variables, for example, whereas programs typically are based on particular textbooks, computer software, and/or instructional processes and usually have a name and a specific provider, such as a company, university, or individual. (pp. 6–7)

Given that he provides no information about whether, and in what ways, he sought information about the larger context of the variables that were reported, we are not able to understand if the variables and analyses he cites as related to the teaching of writing were, in fact, related to some particular program that might have met his criteria of a program described above. His omission of this information raised the question once again about the transparency of the processes of reviewing studies, and how and in what ways he made the decision that teaching variables are not part of program evaluations. Given the emerging nature of this area of inquiry, we were puzzled by the fact that nowhere in his suggested guidelines did he address the issues of the relationship of teaching and learning processes that are part of formal programs, such as Success for All, and how these relate to or promote student achievement. This omission was particularly puzzling given the six decades of research in this area in AERA and the publication of extensive reviews in the four editions of *Handbook of Research on Teaching*, the most recent of which was published in 2001 (Richardson, 2001).

The examination of Slavin's pattern of critique in this section uncovered a series of practices that were troubling to us. As discussed above, Slavin's critiques had the cumulative effect of positioning his own work as the standard because it represented the range of suggestions about synthesis work in program evaluation studies proposed in his article. In responding to his practices, we argue that he missed opportunities to learn from the work of the other sites ways in which their transparency of approach and processes of evaluation might inform his own work. In taking the stance of narrowly focused critique, he leaves invisible a broad range of dimensions that other sites have shown to be relevant to understanding student achievement and to bringing about school reforms that support not only student achievement but also teacher development and community and family participation. Thus, in taking a strong stance on what should count as program evaluation research practices, Slavin ignores many of the factors that have been shown to influence student achievement and the larger issue of student success in schooling. Finally, by limiting his critique to the level of program as the unit of analysis, Slavin makes invisible variations in teaching practices, implementation processes, and institutional supports that constitute a program as experienced by students and their teachers.

A Closing and an Opening: Toward an Extended Hermeneutic Conversation

As indicated at the beginning of this article, we had a number of goals in accepting the editor's invitation to write this response: We wanted to learn from Slavin about his perspective in order to help our students learn how to read his perspective; we wanted to learn more about what he calls the rapidly growing area of program evaluation research syntheses, because we have recently become involved in such research on a formative evaluation basis; and we wanted to learn more about the scope of the field from the international organizations that he identified as engaged in this work.

In writing this response as a set of contrastive analyses to uncover the often invisible dimensions of the work of both Slavin and the organizations whose work he cites, we achieved these goals. However, had we simply read his article and critiqued it for what was not transparent from our own theoretical perspective as ethnographers, we would not have been able to learn as much, given the gaps we located in his text. Therefore, by juxtaposing our analyses of the work inscribed on the websites of the organizations, we were able to fill in these gaps and gain an understanding of the complexity of work on program evaluation research, the different ways in which it is being conducted, and how using Slavin's suggested approach would narrow the information that is important to our work in well-defined programs that are being built at graduate schools and community colleges.

Our systematic analysis of Slavin's critiques of the work of others also leads us to caution readers about the possibility of missed opportunities for learning about program evaluation research that we present as an if-then logic: if researchers and users of syntheses accept Slavin's critiques of the five organizations without critically reading the organizations' own descriptions of their work, then these readers will miss the scope of synthesis work undertaken and available both nationally and internationally that might inform their own work and understandings. As our contrastive analyses of Slavin's inscriptions of the organizations and their websites demonstrate, by tracing the sources cited in his article and (re)reading the studies or websites cited, readers can uncover often invisible dimensions of arguments in the field about what counts as program evaluation research or, the larger issue underlying the arguments in Slavin's article, what counts as scientific research (see Eisenhart & Towne, 2003; Shaker & Ruitenberg, 2007; Shavelson & Towne, 2002).

Our exploration of the different organizations' sites led us to questions about the use of limited designs and why such designs were being recommended on some sites but not others. This question became a rich point, which, according to Agar (1994, 1996), is a place where culture happens and learning takes place. Thus, our approach made visible differences in cultural practices and in common practices across particular sites. Agar further argues that when rich points happen, it is possible to learn about your own cultural assumptions and expectations as well as those of the "other" who created a frame clash for you.

One way of viewing this response is that in sharing our frame clashes, and in making visible how we turned these clashes into rich points, we have proposed an alternative way of reading critically, one that makes any text subject to examination. Another way to see the approach to responding we provided is to view it as a particular form of hermeneutical conversation between the

perspectives that Slavin inscribes and our perspective. From this perspective, each site became a partner for us in our contrastive work, making visible its particular views of program evaluation. In this way, as we read across the sites, our analysis of each site became an anchor for, and thus a participant in, our reading of subsequent sites, making possible new rich points for both learning from the sites and learning about our own cultural understandings. Thus, having engaged in our own hermeneutical conversation with Slavin's texts and those of the other organizations he cites, we look forward to new conversations with Slavin through his response as well as with readers of *ER* through future hermeneutical dialogues.

NOTE

¹This standards document was developed by a presidential taskforce consisting of researchers across empirical social science traditions, which was then reviewed and commented on by AERA's leadership in Council meetings and then by the general AERA membership in a series of public meetings and in written responses. The document was then adopted by AERA's Council in 2006. AERA currently has a taskforce considering standards for other forms of research not based in empirical social science traditions.

REFERENCES

- Agar, M. (1994). *Language shock: Understanding the culture of conversation*. New York: Quill.
- Agar, M. (1996). *The professional stranger: An informal introduction to ethnography* (2nd ed.). San Diego: Academic Press.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40.
- Bazerman, C. (2006). Analyzing the multidimensionality of texts in education. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 77–94). Mahwah, NJ: Lawrence Erlbaum for AERA.
- Bloome, D. (1992). A special issue on intertextuality. *Linguistics and Education*, 4(3–4), 255.
- Bloome, D., Carter, S. P., Christian, B. M., Otto, S., & Shuart-Faris, N. (2005). *Discourse analysis and the study of classroom language and literacy events: A microethnographic perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Castanheira, M. L., Crawford, T., Dixon, C. N., & Green, J. L. (2000). Interactional ethnography: An approach to studying the social construction of literate practices. *Linguistics and Education*, 11(4), 353–400.
- Corsaro, W. (1981). Entering the child's world: Research strategies for field study and data collection. In J. L. Green & C. Wallat (Eds.), *Ethnography and language in educational settings* (pp. 117–146). Norwood, NJ: Ablex.
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher*, 32(7), 31–38.
- EPPI-Centre. (2008). *Searching for studies*. Retrieved January 3, 2008, from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=172>
- Goertz, M. E. (2006). Policy analysis: Studying policy implementation. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 701–710). Mahwah, NJ: Lawrence Erlbaum for AERA.
- Green, J. L. (1983). Teaching as a linguistic process: A state of the art. *Review of Research in Education*, 10, 151–252.
- Green, J. L., & Bloome, D. (1997). Ethnography and ethnographers of and in education: A situated perspective. In J. Flood, S. B. Heath, & D. Lapp (Eds.), *Handbook of research on teaching literacy through the communicative and visual arts* (pp. 181–202). New York: International Reading Association & MacMillan.
- Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2006). *Handbook of complementary methods in education research*. Mahwah, NJ: Lawrence Erlbaum for AERA.
- Green, J. L., & Luke, A. (Eds.). (2006). *Review of research in education. Rethinking learning: What counts as learning and what learning counts* (Vol. 30). Washington, DC: AERA.
- Heath, S. B. (1982). Ethnography in education: Defining the essentials. In P. Gillmore & A. A. Glatthorn (Eds.), *Children in and out of school: Ethnography and education* (pp. 33–55). Washington, DC: Center for Applied Linguistics.
- Ivanič, R. (1994). I is for interpersonal: Discoursal construction of writer identities and the teaching of writing. *Linguistics and Education*, 6(1), 3–15.
- Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. Amsterdam: John Benjamins.
- Kelly, G. (2006). Epistemology and educational research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 33–56). Mahwah, NJ: Lawrence Erlbaum for AERA.
- Kushner, S., & Adelman, C. (2006). Program evaluation: A democratic practice. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 711–726). Mahwah, NJ: Lawrence Erlbaum for AERA.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- McNeil, L. M., & Coppola, E. M. (2006). Official and unofficial stories: Getting at the impact of policy on educational practice. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 681–700). Mahwah, NJ: Lawrence Erlbaum for AERA.
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). London: Sage.
- Mitchell, C. J. (1984). Typicality and the case study. In R. F. Ellen (Ed.), *Ethnographic research: A guide to general conduct* (pp. 238–241). New York: Academic Press.
- Moss, P. A. (Ed.). (2007). Evidence and decision making. *Yearbook of the National Society for the Study of Education*, 106(1), 1–406.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109–162.
- Richardson, V. (Ed.). (2001). *Handbook of research on teaching* (4th ed.). Washington, DC: American Educational Research Association.
- Shaker, P., & Ruitenberg, C. (2007). "Scientifically-based research": The art of politics and the distortion of science. *International Journal of Research and Method in Education*, 30(2), 207–219.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5–11.
- Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 9–18.
- Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5–14.
- Spradley, J. (1980). *Participant observation*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Torgerson, C. J. (2007). The quality of systematic reviews of effectiveness in literacy learning in English: A "tertiary" review. *Journal of Research in Reading*, 32, 287–315.
- Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28(2), 178–200.

AUTHORS

JUDITH L. GREEN is a professor at the University of California, Santa Barbara, Department of Education, Santa Barbara, CA 93106; *green@education.ucsb.edu*. Her research focuses on ethnographic and discourse research methodology and the opportunities linguistically and ethnically diverse students and their teachers are afforded in classrooms and other educational settings.

AUDRA SKUKAUSKAITĖ is an associate professor at the University of Texas, Brownsville, Department of Curriculum and Instruction, 80

Fort Brown, Brownsville, TX 78520; *audra.skukauskaite@utb.edu*. Her research focuses on ethnographic, discourse, and interviewing research methodology; the impact of educational reforms on the lives of teachers; and access to educational opportunities for diverse students.

Manuscript received January 11, 2008

Accepted January 14, 2008