

Overviews in Education Research: A Systematic Review and Analysis

Joshua R. Polanin
Development Services Group

Brandy R. Maynard and Nathaniel A. Dell
Saint Louis University

Overviews, or syntheses of research syntheses, have become a popular approach to synthesizing the rapidly expanding body of research and systematic reviews. Despite their popularity, few guidelines exist and the state of the field in education is unclear. The purpose of this study is to describe the prevalence and current state of overviews of education research and to provide further guidance for conducting overviews and advance the evolution of overview methods. A comprehensive search across multiple online databases and gray literature repositories yielded 25 total education–related overviews. Our analysis revealed that many commonly reported aspects of systematic reviews, such as the search, screen, and coding procedures, were regularly unreported. Only a handful of overview authors discussed the synthesis technique and few authors acknowledged the overlap of included systematic reviews. Suggestions and preliminary guidelines for improving the rigor and utility of overviews are provided.

KEYWORDS: overview, review of reviews, meta-analysis, education research

Research synthesis, a rigorous approach to cumulate evidence, has become an important technique to manage, integrate, and summarize the burgeoning research industry (Cooper, 2010). Researchers use syntheses to generate new knowledge and identify gaps in extant literature (Lipsey & Wilson, 2001; Pigott, 2012). Policymakers and practitioners increasingly rely on systematic reviews to inform funding allocations and practice (Gough, Oliver, & Thomas, 2012). The research synthesis industry is efficient and expanding, nearly doubling each year in the social sciences (Williams, 2012). Bastian, Glasziou, and Chalmers (2010) estimated that 11 systematic reviews are published daily in the online database MEDLINE alone.

Largely as a result of the increase in systematic reviews, researchers have begun to synthesize the syntheses. This method of research synthesis (Becker &

Oxman, 2008), where the review is the unit being synthesized rather than the primary study, offers another means to precis the ever-increasing amount of research generated (Bastian, Glasziou, & Chalmers, 2010). These syntheses can produce answers to unique and important questions that other research methods cannot (Cooper & Koenka, 2012) and are often robust to sample or scale variations, resulting in utility and practicality for policymakers and practitioners above and beyond systematic reviews. This method has been referred to by different terms, including meta-meta-analysis (Hattie, 2009; Kazrin, Durac, & Agteros, 1979), meta-synthesis (Cobb, Lehmann, Newman-Gonchar, & Alwell, 2009), overview (Pieper, Antoine, Morfeld, Mathes, & Eikermann, 2014a), overview of reviews (Cooper & Koenka, 2012), review of reviews (Maag, 2006), second-order meta-analysis (Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011), tertiary review (Torgerson, 2007), and umbrella review (Thomson, Russell, Becker, Klassen, & Hartling, 2010).

We adopt the terminology used by the Cochrane Collaboration and refer to a synthesis of reviews as an overview (Becker & Oxman, 2008). Overviews are becoming increasingly common in health sciences (Pieper, Buechter, Jerinic, & Eikermann, 2012; Thomson et al., 2010), and overviews' results are extending into other areas including the social and education sciences (Cooper & Koenka, 2012). As such, overviews have the potential to shape education policy and provide guidance to researchers and practitioners alike.

Examples of influential overviews are easily identified in the literature. For example, Higgins, Xiao, and Katsipataki (2012) synthesized 45 systematic reviews on the effects of digital technology on children's learning. The authors provided a comprehensive summary of the findings as well as clear and direct recommendations to practitioners based on the totality of the studies. The overview provided clarity to the discrepant systematic review results, making them easier to interpret and implement in practice. Consequently, Higgins et al.'s overview has already been cited 15 times since it was published. Torgerson's (2007) overview, cited 27 times since publication, combined 14 systematic reviews on the effects of literacy training. The reviews were grouped into content areas where specific conclusions could be drawn about each of the varying programs or intervention styles. The authors suggested that, based on the overview findings, specific literacy training may be more appropriate for differing groups of students or intervention styles, a finding that would not be possible with a traditional or systematic review that focuses on one or small set of studies. Finally, the largest education research overview conducted to date (Hattie, 2009) synthesized over 800 reviews related to academic achievement and has been cited 113 times since 2009. The overview synthesized well over 10,000 primary studies, and therefore, its conclusions may be more robust to sample and intervention variation. Moreover, the overview is able to provide comparisons across the reviews and thus make suggestions and inferences not possible using individual reviews.

Although overviews are becoming prevalent and may offer advantages over traditional research syntheses, overviews are a relatively nascent and undeveloped synthesis method that pose unique methodological challenges (Cooper & Koenka, 2012; Thomson et al., 2010) and may be problematic (Hartling, Chisholm, Thomson, & Dryden, 2012; Pieper et al., 2012). It is unclear to what

extent overviews are being conducted in education research, the methods used to conduct education overviews and synthesize results, or how valid this research method is. The purpose of this study, therefore, is to examine the prevalence of education overviews, assess the current state of overviews in education, outline the unique challenges and contributions overviews may provide above and beyond systematic reviews, and provide preliminary guidelines to education researchers based on the review's results.

Unique Contributions From Overviews

Overviews can make unique contributions to the knowledge base above and beyond systematic reviews and be advantageous to policymakers, practitioners, and researchers alike. Overviews can provide a broader summary of evidence for use by stakeholders and researchers (Cooper & Koenka, 2012) and can be used to examine trends and changes in research over time. Overviews allow for the research problem to be defined in a broader way, capture a variety of interventions being used to treat similar conditions, or be used to identify variation in the types of outcomes, problems, populations, or contexts of the same intervention (Becker & Oxman, 2008; Cooper & Koenka, 2012).

Another advantage of the overview is the ability to compare and contrast results across multiple systematic reviews. The rapid growth of the systematic review industry means that reviews on the same topic will occur, and those reviews may result in varying conclusions. It might be difficult to discern concordance or discordance among reviews without the context of an overview. One illustrative example is from the literature on school bullying prevention programs. Merrell, Gueldner, Ross, and Isava (2008) synthesized 15 studies across 15 different outcomes. The results for the reduction in bullying perpetration indicated only a small, nonsignificant intervention effect, $k = 8$, $d = .04$. On the other hand, Ttofi and Farrington (2011) synthesized 89 studies of 53 evaluations divided into only two main constructs, bullying perpetration and bullying victimization. The average results for the reduction in bullying perpetration indicated a larger and statistically significant treatment effect, $k = 41$, $d = .17$. By synthesizing these two reviews, overview authors have the opportunity to compare and contrast the variation between these discordant reviews based on differences in research questions, populations, methods, or other characteristics (Pieper et al., 2012). As such, identifying and examining discrepancies and agreements across reviews provide valuable evidence that could be used by stakeholders and researchers to advance scientific knowledge and practice.

A third advantage offered by overviews is the ability to conduct a network meta-analysis (Ioannidis, 2009). A network meta-analysis is applicable when multiple interventions and control groups are compared. This analysis allows the researcher to understand differences across interventions or comparisons, even when direct comparisons were not made within the reviews. A relevant yet hypothetical example derives from literature on the impact of various interventions to increase math test scores. One systematic review collects studies that tests curriculum changes, another synthesizes the effects of teacher professional development, and a third review includes studies that examine the effectiveness of curriculum changes compared with teacher professional development and both of

those types to a control group. Using a network meta-analysis, one is able to compare *across* each of the various combinations of interventions in addition to the simple comparison of intervention versus control. Cooper and Koenka (2012) described one such scenario in the context of medical research. To date, however, researchers have not attempted an analysis of this kind in education, but it is likely a logical next step.

A final advantage to conducting an overview is that it elucidates when systematic reviews need an update. The Campbell Collaboration (2014), the leading producer of systematic reviews in the social sciences, suggests that reviews be updated at least every 3 to 4 years. The Cochrane Collaboration, the leading producer of systematic reviews in the medical sciences, suggests an update may be necessary even sooner (Higgins & Green, 2008). An overview that synthesizes the corpus of systematic reviews will recognize if an update to a specific field is required. Pieper, Antoine, Neugebauer, and Eikermann (2014c) provided a helpful framework to assess whether a particular systematic review is up to date, which could be used across reviews.

Taken together, overviews can be useful and enlightening to inform policy, practice, and research. The use of overviews, however, relies on their validity, applicability, and methodological rigor. As such, the community must maintain high standards for such overviews, similar to the way that methodologists have argued for higher standards in systematic reviews (Moher, Liberati, Tetzlaff, & Altman, 2009).

Conducting an Overview

The conduct and organization of an overview, in many ways, is very similar to a systematic review. Cooper and Koenka (2012) suggested that an overview mirrors the steps of a systematic review, following the suggestions of Cooper (2010) or Lipsey and Wilson (2001). As illustrated in Table 1, the parallels between the two methods are striking, and overview researchers, in the face of few guidelines, would do well to simply follow the methodological suggestions of systematic reviewers. The need to formulate a well-conceived research question, search the literature for relevant studies, extract data from studies, evaluate the studies, analyze and integrate the outcomes, and interpret and present the evidence are the basis of rigorous synthesis methods (Cooper, 2010). Although the major steps of conducting an overview are analogous to conducting a systematic review, important differences remain within these steps that are critical to consider.

One of the most significant differences between conducting a systematic review compared with an overview is the need for overview authors to consider multiple study levels—the overview level, the review level, and the primary study level—throughout the process and take steps to minimize bias and error at all levels. Overview authors may introduce bias and error through their methodological procedures and by including reviews that contain bias and errors. Authors introduce bias through their own methods and through the inclusion of possibly biased primary studies. Indeed, overview authors compound bias and error when the methods used at the overview, review, and primary study levels are not evaluated. Overview authors therefore must consider not only how they conduct the overview but also how the review authors conducted their review.

TABLE 1

Comparison of steps for conducting a systematic review and overview

Steps	Systematic review	Overviews
Objectives and primary unit of analysis	To synthesize the findings from primary studies	To synthesize the results from research syntheses and primary studies
Eligibility criteria	Primary studies must meet certain criteria	Both primary studies <i>and</i> research syntheses must meet certain criteria
Search	Online databases, reference harvest, author contact, targeted websites, search engine	Same as systematic reviews, but includes the harvesting of references from all included reviews
Screening	Abstract and full-text screening should be reported, preferably conducted by more than one person	Same as systematic review
Data extraction	Information on the primary study: population, setting, intervention (if applicable), comparison (if applicable), outcome, study design, effect size Extraction procedures should be reported, preferably done by more than one person	Information on the overview and included reviews and primary studies. In addition to data collected for systematic review: search criteria (review), screening criteria, data extraction procedures, outcomes, results, moderator or sensitivity analyses, extraction procedures should be reported and preferably done by more than one person
Quality of studies	Study quality assessed for each primary study	Study quality assessed for each systematic review; report quality appraisals for the primary studies in the reviews
Summary of findings table	Include all relevant information extracted from primary studies	Include all relevant information extracted, including summary of primary studies and systematic review procedures
Synthesis: Descriptive	Group studies into relevant categories and report results	Group studies into relevant categories and report results; describe discordance between reviews
Synthesis: Quantitative	Use inverse-variance following set guidelines (Cooper, Hedges, & Valentine, 2009)	Consider using Schmidt and Oh (2013); the procedures should be reported
Subgroup analyses	Use moderator or meta-regression analyses following set guidelines (Cooper et al., 2009)	No prescribed guidelines; consider using moderator or meta-regression techniques
Publication bias	Reported on population of studies	Report results from the individual reviews and consider analyzing the overviews

The importance of taking both the overview and the review level into account during the overview process is first apparent when determining eligibility criteria. Overview authors must explicate eligibility criteria related to their primary unit of analysis—the review—in addition to the eligibility criteria they define for the primary studies included in the reviews. The eligibility criteria, for example, may specify that the overview will include only systematic reviews (i.e., descriptive reviews are excluded) that examined effects of an intervention using randomized controlled trials only. In this case, overview authors must attend to both the design of the reviews (systematic reviews only) as well as the study designs included within the reviews (randomized controlled trials only). In this example, any systematic review that includes primary studies other than randomized-controlled trials would therefore be ineligible for inclusion.

Overview data extraction takes a similar form to coding primary studies for a systematic review, but the information extracted is often quite different. The difference again lies in the multiple levels embedded within the overview process. For an overview, the author extracts data related to the primary unit of analysis—the included reviews. Overview authors must also consider what data they will extract related to the primary studies included in the reviews, and they can choose to include or ignore reporting on primary studies. Ignoring the primary studies, however, likely results in an incomplete portrayal of systematic review findings and thus the credibility of the overview would be questionable. An illustrative example is study design. A systematic review that includes many types of controlled and uncontrolled studies differs from a systematic review that only includes randomized controlled trials. The average effect sizes may be similar across the two reviews, but the internal validity of each primary study differs greatly. Therefore, it is important that overview authors code and report pertinent information about the systematic review as well as information the systematic review reports about the primary studies.

Study quality is another component of the overview process that must be considered at the review and primary study levels. For systematic reviewers, it is critical to assess study quality and risk of bias of included studies because problems with the design and execution of primary studies have implications for the inferences gleaned from the review (Valentine & Cooper, 2008). Overview authors must also consider the quality of the included reviews as well as the quality of the primary studies constituting the reviews. High-quality systematic reviews may include many or mostly low-quality primary studies. The validity of the conclusions drawn across included systematic reviews relies on the quality of the overview, reviews, and primary studies.

The final component of the overview process to consider is in the synthesis of the results. Similar to a systematic review author, the overview author can elect to describe each study individually, conduct a descriptive synthesis, or quantitatively synthesize the results of the reviews using meta-analytic techniques. The criticisms of descriptive and vote counting methods of synthesizing outcomes (see Cooper, 2010) apply equally to overviews. In terms of a quantitative synthesis of results, overview authors are faced with more complexity than review authors. Overview authors may choose to extract and synthesize primary study level effect

sizes (where available) and variances or they may choose to extract the average effect size calculated and reported by review authors. If the overview authors choose to quantitatively synthesize mean effects across reviews, little guidance is available; however, Schmidt and Oh (2013) described methods for second-order meta-analysis when each meta-analysis reports the results from a random-effects model.

Current State of Overview Methods

Increase in the demand and production of research syntheses engenders the need for more credible methods of synthesizing evidence (Cooper, 2010). As a result, the methods of research synthesis have been advancing dramatically over the past 20 years. Multiple research articles, books, and journals devoted to research synthesis methods have been published to improve the practice of research synthesis and advance the science of research synthesis methods (Shadish & Lecy, 2015). Although significant empirical work has been undertaken to inform and improve research synthesis methods to minimize bias and error in the review process and increase the credibility and validity of review findings (Cooper, 2010; Moher et al., 2009; What Works Clearinghouse, 2015), limited research or guidance is available to the overview author.

The Cochrane Collaboration endorses overviews and has published guidance on the conduct and reporting of health-related overviews (Becker & Oxman, 2008). Cooper and Koenka (2012) and Thomson et al. (2010) offered a description of the steps and methods overview researchers have adapted from other methods and the challenges inherent in the overview process. The What Works Clearinghouse's (2015) guidelines explicitly discuss reviewing education-related topics, but focus exclusively on reviewing primary studies. Limited extant empirical inquiry in education regarding overview methods is available, however, and the limited empirical work published in this field is primarily in medicine and health (Thomson et al., 2013).

Pieper and colleagues (Pieper et al., 2012, 2014a; Pieper, Antoine, Neugebauer, & Eikermann, 2014c) published a series of studies examining the rigor, overlap, and up-to-dateness of overviews in the health sciences. They found in their review of 126 overviews that there was much heterogeneity in the conduct of overviews, and many overviews lacked methodological rigor. Moreover, only about half of the overviews considered overlap of reviews, with the possibility of certain primary studies being included more than once, which gives disproportionate statistical power to those primary studies (Pieper et al., 2014c). Up-to-dateness is another characteristic of overviews that has been examined empirically, with findings pointing to overview authors' lack of attention to whether the reviews are providing the most up-to-date evidence (Pieper et al., 2014a). Thomson et al.'s (2013) review of 29 overviews concluded that considerable work is still needed on the methods of overview research.

Cooper and Koenka (2012) and others (Pieper et al., 2012, 2014c; Thomson et al., 2013) have called for researchers to examine and advance overview methods. Despite these calls, however, overview methods in education research have largely been overlooked. The purpose of this study, therefore, is to build on prior studies to further elucidate overview methods and expand this research into the

field of education. By examining the extant overviews of education research, we can describe the prevalence and current state of overviews in education, compare the overview methods used by education researchers to that of health sciences, and begin to provide further guidance to conducting overviews of education research and advance the evolution of overview methods. The research questions guiding this study are the following: (a) To what extent are overviews being conducted in the area of education related research with preschool to postsecondary student populations? (b) To what extent are methodological characteristics being reported in overviews of education? (c) What methods are overview authors using to conduct overviews? We conclude by suggesting preliminary overview methodological guidelines based on the answers to these questions.

Method

Systematic review procedures were employed to search, select, and extract data from overviews that meet eligibility criteria for this study. We followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines where applicable (Moher et al., 2009).

Eligibility Criteria

Eligible overviews must have aimed to synthesize more than one empirical education-related review (e.g., narrative, systematic review, or meta-analysis) with preschool, primary, secondary, or postsecondary students, including special or general populations. For the purposes of this study, we considered the overview focused on an education related topic if the overview authors explicitly reported that the focus was on education in the title or the abstract, or if at least 50% of the included reviews synthesized effects of school-based interventions or education-related outcomes. We did not restrict our search to any time frame, and we searched for both published and unpublished reports; however, we included only English language reports. Relevant authors were contacted to inquire about potential missing studies.

Search Procedures

Seven electronic databases were searched in September 2014 to identify eligible overviews: Academic Search Premier, Education Complete, ERIC, ProQuest Dissertations and Theses, PsychINFO, Science Direct, and Social Sciences Citation Index. Keyword searches within each electronic database included variations of the following keyword terms: “meta-review,” “umbrella review,” “review of review,” “overview of review,” “meta-meta-analysis,” “overview,” “meta-analysis of meta-analyses,” “synthesis of review,” and “synthesis of systematic review.” Hand-searching of reference lists and forward citation searching using Google Scholar were conducted with articles identified during the search process as well as with the following articles identified prior to the search: Cooper and Koenka (2012), Thomson et al. (2010, 2013), Pieper, Buechter, Jerinic, and Eikermann (2012), Pieper, Antoine, Morfeld, Mathes, and Eikermann (2014a, 2014c). We searched the gray literature using Google Scholar. The full search strategy for each electronic database is available in Appendix A (available in the online version of the journal).

Study Selection and Data Extraction

Titles and abstracts of the overviews found through the search procedures were screened for relevance by one author. If the report appeared to be eligible, or if there was any question as to the appropriateness of the report at this stage, the full text document was obtained and independently screened by two authors using a screening instrument to determine inclusion. Any discrepancies between authors were discussed and resolved through consensus, and when needed, a third author reviewed the study.

Studies that met inclusion criteria were coded using an author-developed data extraction instrument comprising the following sections: (a) bibliographic information; (b) overview characteristics and methods; (c) information overview authors provided about included reviews; (d) overlap, quality, and up-to-dateness of included reviews; (e) overview synthesis methods; and (f) Google Scholar citation rates. The data extraction instrument, available from the authors, was pilot tested with five of the included overviews by two authors and adjustments to the coding form were made. Two authors then independently coded the remaining overviews using a FileMaker Pro database (Apple, Inc., 2014). Initial interrater agreement was 92.5%. Discrepancies between the two coders were discussed and resolved through consensus, and when needed, a third reviewer was consulted.

Analytic Procedures

The studies were analyzed descriptively for purposes of reporting. We aimed to elucidate all aspects of the overviews including where the overviews failed to collectively report important methodological details. We calculated the percentage of characteristics reported across different methodological aspects. We also elected to test for differences in the reported methodological characteristics across time, and we hypothesized that overviews published more recently would report more information. To investigate whether reporting of methodological characteristics varied across time, we selected 8 of the 17 coded characteristics to be summed (i.e., databases reported, keywords reported, time frame reported, abstract screening process, full-text screening process, coding procedures reported, gray literature searched, eligibility criteria reported) within each overview. We choose these eight characteristics because (a) the characteristics could be used in each study (i.e., all overviews searched online databases but not all overviews conducted a quantitative synthesis) and (b) the characteristics could be easily reported in the overview. A Pearson product-moment correlation was estimated between the summation and the time variable. We also split the sample into recent and early overviews and calculated a *t* test. We created all figures using Microsoft EXCEL and conducted analyses using base R (R Core Team, 2015).

Results

Search Results

Titles and abstracts of the 6,566 citations retrieved from electronic searches of bibliographic databases and additional citations reviewed from reference lists of prior reviews and forward citation searching were screened

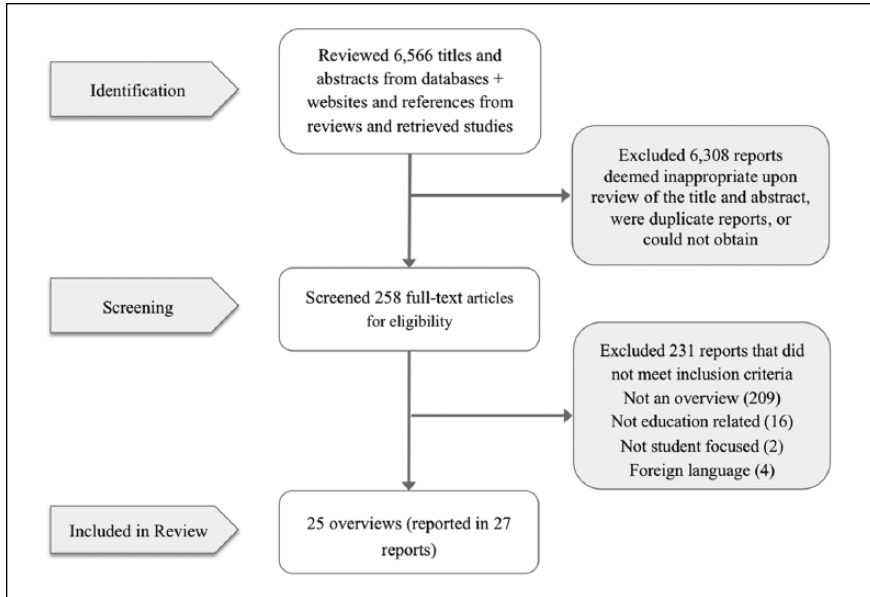


FIGURE 1. *Flow chart of search and selection process.*

for relevance. Of those, the full texts of 258 reports were retrieved and independently screened for inclusion by two authors. Twenty-five overviews reported in 27 reports met eligibility criteria for this study. The 231 reports deemed ineligible were excluded for the following reasons: not being an overview ($n = 209$), not being education related ($n = 16$), not focused on Pre-K through postsecondary populations ($n = 2$), or not published in English ($n = 4$). A list of excluded studies is available in Appendix B (available in the online version of the journal). See Figure 1 for a flow chart summarizing the search and selection process.

Descriptive Characteristics of Included Overviews

The majority of included overviews were conducted with the purpose of examining effects of interventions (80%). Five authors reported conducting the overview for other purposes: to examine effects of schooling, describe the status of knowledge in reading, explore relationships of various education-related factors, and to consolidate and examine the results of the meta-analyses compared with work of other researchers. The overviews covered a variety of topics, including learning and educational achievement, science education training, instructional systems and design, curriculum, literacy and reading, technology, social skills training, school-based health promotion, mental health and social/emotional learning and well-being, social skills training, special education interventions, and self-determination. Although some overviews were found in the gray literature, most (80%) were published in peer-reviewed journals. Publication dates

ranged from 1983 to 2012 ($SD = 9.63$). The overviews included a median of 30 reviews (range 5 to 800) and no overviews included primary studies. Yearly citations rates were high for the included overviews; the median overview received 6.85 citations per year, and the top overview had been cited more than 1,800 times (Lipsey & Wilson, 1993).

Methodological Characteristics of Overviews

We investigated the reported methodological characteristics of overviews and identified 17 distinct methodological criteria: online database searching, hand searching, reference harvesting, author contacts, targeted website search, keywords identified, time frame searched and included, gray literature sources, abstract screening, full-text screening, coding procedures, eligibility criteria, synthesis technique, meta-analysis procedures, adjustments for varying meta-analysis techniques, subgroup analyses, and publication bias. Table 2 delineates many of these characteristics across each of the overviews.

Notably, many aspects that systematic reviews routinely report were lacking in the overviews. Although the online databases were often reported (64%), other critical aspects of the search, such as reference harvesting (48%), author contacting (16%), and hand searching (40%), were reported less than half the time. Only about a third of the overviews reported any of the keywords used to search the online databases (36%), and 40% of overviews reported the allowable time frame of reviews. With regard to the eligibility criteria, about half (56%) of the overviews reported at least some criteria, but few studies (16%) reported every important eligibility criterion. In addition, few overviews detailed how they screened abstracts (24%) and full-text reports (28%), or how they extracted data from the reviews (28%).

Reporting Trends Across Time

Across the eight eligible categories, only two overviews (8%) reported all eight methodological characteristics (Tamim et al., 2011; Torgerson, 2007). Across the 25 overviews, an average of 3.56 ($SD = 2.60$) of the eight characteristics were reported. One possibility for the lack of methodological reporting is that a portion of the overviews was published prior to widespread acceptance and use of reporting standards and guidelines. We therefore sought to determine the relationship between time and the reported methodological characteristics. Figure 2 illustrates the reporting trend across time. It is clear that overview authors have begun to report more methodological aspects of their studies, especially compared with those conducted in the 1980s and early 1990s. The Pearson product-moment correlation between the number of reported methodological characteristics and the year is positive and statistically significant ($r = .41, p = .04$). In addition, we also dichotomized the sample into recent (2000–2015) and early (1983–1999) overviews. The number of reported methodological characteristics is greater in more recent overviews ($M = 4.31, SD = 2.63$) compared with earlier reviews ($M = 2.75, SD = 2.42$), but the difference is not statistically significant, $t(23) = 1.55, p = .13$. Nevertheless, it is clear that methodological reporting has improved ($d = .60$), although there is still need for improvement.

TABLE 2

Methodological characteristics of included overviews

Author (DoP)	Type; reviews	Purpose	Overview search characteristics										Overview screen/code characteristics				
			DB	HS	RE	AU	WS	SE	KW	TF	GL	AS	FS	CO	EL		
Anderson (1983)	J; 8	2	✓										0				5
Browne, Gafni, Roberts, Byrne, and Majumdar (2004)	J; 23	1	✓		✓	✓							1, 2, 5				1, 3, 5, 6
Cobb, Lehmann, Newman-Gonchar, & Alwell (2009)	J; 7	1	✓		✓				✓				5	✓			3, 4, 6
Cook et al. (2008)	J; 5	1	✓		✓				✓				3	✓			1, 2, 3, 4, 5, 6
Diekstra (2008)	R; 19	1	✓		✓				✓				0				1, 2, 3, 4, 5, 6
Forness (2001)	J; 24	1											0				0
Fraser, Walberg, Welch, and Hattie (1987a)	J; NA	3		✓									0				4, 5, 6
Fraser, Walberg, Welch, and Hattie (1987b)	J; 170	3		✓									1, 6				5
Fraser, Walberg, Welch, and Hattie (1987c)	J; 134	3	✓										5				1, 3, 5, 6
Green, Howes, Waters, Maher, and Oberklaid (2005)	J; 8	1	✓		✓			✓					2, 3				1, 2, 3, 4, 5, 6
Gresham (1998)	J; 9	1											0				0
Guthrie, Seifert, and Mosberg (1983)	J; 293	3	✓		✓				✓				3			✓	3

(continued)

TABLE 2 (continued)

Author (DoP)	Type; reviews	Purpose	Overview search characteristics											Overview screen/code characteristics				
			DB	HS	RE	AU	WS	SE	KW	TF	GL	AS	FS	CO	EL			
Hattie (1992)	J; 134	3	√															1, 5, 6
Hattie (2009)	B; 800*	1								1, 2, 3, 4, 5								0
Higgins et al. (2012)	R; 45	1							√									1, 5
Kulik and Kulik (1989)	J; 59	1								0								0
Lipsey and Wilson (1993)	J; 176	1	√							1, 2, 3, 4, 5, 6								0
Lister-Sharp, Chapman, Stewart-Brown, and Sowden (1999)	G; 32	1	√	√	√				√	1, 3, 4, 5				√				1, 2, 3, 4, 5, 6
Maag (2006)	J; 13	1		√					√									1, 2, 6
Sipe and Curlette (1997)	J; 103	2	√	√					√	0				√				5, 6
Tamim et al. (2011)	J; 25	1	√	√					√	1, 2, 4, 5				√				2, 3, 5, 6
Tennant, Goens, Barlow, Day, and Stewart-Brown (2007)	J; 27	1	√	√					√	0				√				1, 5
Torgerson (2007)	J; 14	2	√	√					√	3				√				1, 2, 3, 5, 6
Wang, Haertel, and Walberg (1993)	G; 92	1	√							1, 3, 5								1, 2
Weare and Nind (2011)	J; 52	1	√	√	√				√	1, 2, 3, 4				√				1, 3, 4, 5, 6

Note. * = The author (Hattie) did not provide the exact number of reviews, but rather indicated there were over 800 reviews in his overview. √ = Reported information; Author represents the reference; DoP = date of publication; Type = publication type (J = journal article, R = research organization report, B = book, G = government report); Reviews = number of included reviews; Search characteristics (DB = reported database search, HS = hand search, RE = reference harvesting, AU = contact authors, WS = targeted website, SE = search engine, KW = keywords, TF = time frame, GL = gray literature [1 = government report, 2 = private report, 3 = books, 4 = conference presentation, 5 = dissertations and theses, 6 = other]); Screen/Code characteristics (AS = abstract screening, FS = full-text screening, CO = coding process reported, EL = eligibility criteria [1 = population, 2 = setting, 3 = intervention/relationship, 4 = design, 5 = outcome, 6 = other]).

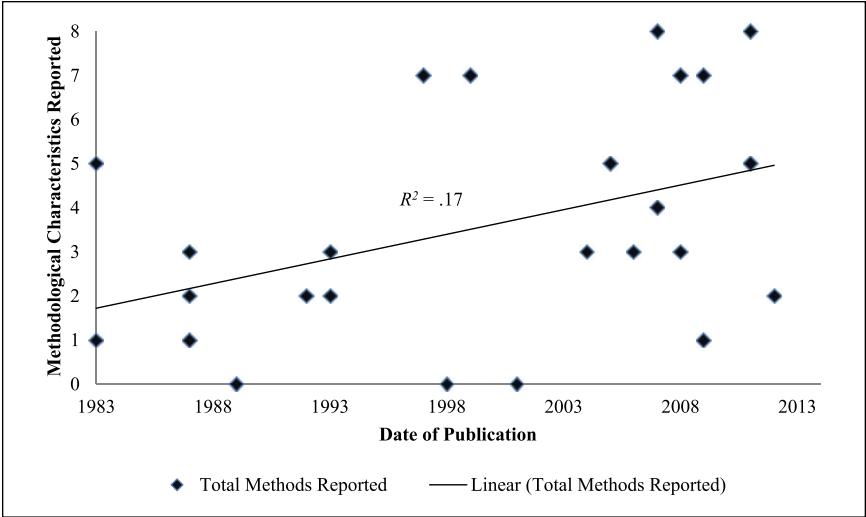


FIGURE 2. Number of methodological characteristics of overviews reported over time.

Synthesis Methods

We also investigated the synthesis methods used by overview authors (Table 3). Of the 25 overviews included, 11 (44%) used a descriptive review technique, electing to summarize each review textually and without quantitative analysis, whereas 14 of the 25 overviews (56%) used a quantitative analytic technique. Of these 14, four elected to average the review results using a simple nonweighted average (16%), and five chose to weight the results by the number of included reviews (20%). Five of the overviews did not report how they synthesized the results (20%). Of the overviews that conducted a quantitative analysis, only a small proportion conducted subgroup (16%) or publication bias (12%) analyses. Notably, none of the studies that used a quantitative technique considered or adjusted for various meta-analytic models and none of the studies used the technique proposed by Schmidt and Oh (2013). Taken together, the methodological aspects of the meta-analyses lacked clear and consistent reporting.

Reported Characteristics of Reviews Included in Overviews

We also evaluated the reported characteristics of reviews included in the overviews (Table 4). We examined 13 distinct and important categories: publication type, number of included studies, time frame searched, databases searched, search and screen procedures, coding procedures, study designs included, study quality, outcome type, analysis procedures, average effects, moderator/sensitivity analyses, and publication bias.

The results of coding these aspects revealed systematic deficiencies across the overviews. For example, only one overview (4%) reported the search and

TABLE 3*Methods used to synthesize reviews in included overviews*

Method	<i>n</i>	Percentage
Synthesis method		
Descriptive	11	44
Quantitative	14	56
Quantitative		
N/A: Descriptive review	11	44
Did not report	5	20
Simple average	4	16
Weighted by number of reviews	5	20
Adjusted for varying models		
N/A: Descriptive review	11	44
Yes	0	0
No	14	56
Subgroup analysis		
N/A: Descriptive review	11	44
Yes	4	16
No	10	40
Publication bias analysis		
N/A: Descriptive review	11	44
Yes	3	12
No	11	44

screening procedures used by the included reviews, meaning only one overview identified how each included systematic review conducted their search and screening of included primary studies. Along the same theme, only two overviews identified the coding procedures of the included systematic reviews (8%). Also somewhat surprisingly, only two overviews reported the databases searched by each of the reviews (8%), and only one overview reported the time frames of included studies in the reviews (4%). Some other aspects of the reviews, however, were more consistently reported by the overview authors. Most overviews reported the number of studies included in each review (60%). A majority of studies reported the types of outcomes coded within each review (64%), the analysis procedures of the reviews (84%), and the average results (76%).

We also examined whether overview authors limited their study to include only systematic reviews (as a means of controlling for quality), or assessed the quality of included reviews and, if so, what they used to assess quality. Of the 25 overviews included in this review, six (24%) overview authors limited the inclusion of reviews to systematic reviews. Eight of the overviews assessed and reported review quality in some way. One overview author used the QUORUM (Torgerson, 2007), one used the Critical Appraisal Skills Program (Tennant et al., 2007), and the remaining reviews used author-developed tools.

TABLE 4*Review characteristics reported in the overviews*

Author (DoP)	Reported review characteristics in the overview												
	T	NS	Tf	SS	Db	Co	SD	Qu	Ou	An	AE	M/S	PB
Anderson (1983)									√	√	√		
Browne et al. (2004)	√	√						√	√	√	√		√
Cobb et al. (2009)	√	√	√				√	√	√	√			√
Cook et al. (2008)	√	√					√	√		√	√		√
Dijkstra (2008)		√							√	√			
Forness (2001)		√							√	√	√		
Fraser et al. (1987a)													
Fraser et al. (1987b)													
Fraser et al. (1987c)									√	√	√		
Green et al. (2005)		√							√	√			
Gresham (1998)										√	√		
Guthrie et al. (1983)													
Hattie (1992)		√							√	√	√		
Hattie (2009)		√								√	√		
S. Higgins et al. (2012)									√	√	√	√	
Kulik and Kulik (1989)		√							√	√	√		
Lipsey and Wilson (1993)		√							√	√	√		
Lister-Sharp et al. (1999)	√	√			√		√	√	√	√	√		√
Maag (2006)		√					√			√	√		
Sipe and Curlette (1997)	√				√	√		√	√	√	√	√	√
Tamim et al. (2011)		√								√	√		
Tennant et al. (2007)							√		√	√	√		
Torgerson (2007)	√	√					√	√	√	√	√		√
Wang et al. (1993)											√		
Weare and Nind (2011)	√	√		√		√	√	√	√	√	√	√	√
Percentage reported	28	60	4	4	8	8	28	28	64	84	76	12	28

Note. √ = Reported information; Author = Author represents the reference; DoP = date of publication; T = publication type; NS = number of studies; TF = time frame; SS = search/screen strategy; DB = search databases; CO = coding strategy; SD = study design; Qu = study quality; Ou = outcome type; AN = analysis procedure; AE = average effect; M/S = moderator/sensitivity analyses; PB = publication bias; Purpose (1 = To assess the effects of effectiveness reviews, 2 = To review or measure quality or methodological issues, 3 = Other).

Overlap and Up-to-Dateness of Reviews in Overviews

In the majority of the overviews (68%), overview authors did not address the issue of overlap in any way, four of the authors acknowledged that reviews included some of the same primary studies, and three authors accounted for overlap in some way (e.g., removed highly overlapping reviews from analysis). Only

one author provided a matrix of all primary studies included in each review, clearly identifying the primary studies that were included in multiple reviews.

In terms of how current, or up-to-date, the overviews were, we assessed publication lag by calculating the difference between the mean publication year of the included reviews and the publication year of the overview. We also calculated the proportion of reviews published more than 5 years prior to the overview (Pieper et al., 2014c). The mean publication lag was 7.67 years. Many of the overviews included a large proportion of older reviews. The proportion of included reviews published 5 or more years prior to the overview ranged from 0% to 100%, with a median of 69%.

Discussion

The purpose of the present study was to examine the extent that overviews are conducted in education, to assess the methodological characteristics of education overviews, and to provide recommendations to advance overview methods and reporting standards given the current state of the evidence. Our systematic and comprehensive search of the education literature yielded 25 overviews for inclusion. Overall, our findings suggest a serious lack of methodological reporting and use of rigorous methods for conducting overviews, even when considering the fact that a portion of the reviews were conducted prior to widespread adoption of reporting guidelines for primary studies and systematic reviews. Many overviews failed to provide specific details about the search, screen, and coding procedures and a large portion of the overviews did not report many aspects of the eligibility criteria. Overall, we identified three major concerns about overviews in education research: (a) lack of reporting of methods used and characteristics of included reviews and primary studies, (b) sparse attention to overlap across reviews, and (c) underreporting of procedures used to synthesize the reviews. Although disturbing, these results should not be surprising given the previously conducted studies of overviews' findings in the health and medical fields (Hartling et al., 2012; Pieper et al., 2012; Thomson et al., 2013) that found similarly lacking methodological rigor and reporting of key information.

One of the most concerning results from the present study is related to the reporting by overview authors, both in terms of reporting the overview methodology used and reporting information about the included reviews and primary studies. This omission is especially apparent with regard to the selection and coding of reviews, which are crucial to the validity of a review. The practice of omitting crucial study selection and coding details is akin to omitting how participants are selected and data collected for primary studies. Compared with results of similar items assessed in studies of overviews in other fields, fewer education overviews (24%) reported methods for study selection compared with 49% found in Hartling et al.'s (2012) and Li et al.'s (2012) reviews in health. A smaller proportion of education overviews (28%) were also found to report data extraction procedures compared with 60% in the Hartling et al. (2012) and 44% in the Li et al.'s (2012) medical review of overview study. As seen in Figure 2, more recent overviews reported more methodological characteristics than those conducted in prior decades, yet better reporting is still needed.

Although methodological characteristics are important to inform the quality and validity of overviews, reporting characteristics of the reviews and primary studies are also important to the quality, validity, and relevance of the overview. Of the overviews included in this study, overview authors provided insufficient information about the included reviews and rarely provided basic information about the primary studies. Most overview authors did not report quality indicators of the included reviews; however, a greater proportion of education overviews reported quality indicators (28%) compared with Li et al.'s (2012) review where only 7% of the overviews reported quality indicators, but fewer than what was found by Hartling et al.'s (2012) study where 36% reported quality indicators. When primary study information was present, it was usually in regard to the number of included studies or average effect size. The ability to judge the validity of an overview rests in the methods and quality of the reviews and the primary studies included in those reviews. A serious deficit of information regarding the included reviews and primary studies inhibits assessment of the validity and, ultimately, the relevance of an overview.

Overlap of primary studies across included reviews within an overview is another area of concern and has implications for the validity of overviews. When conducting a primary study synthesis, it is well-established that two reports of the same study should not be included, as that would cause a duplication of data; review authors are well aware of the need to ensure independence of effect sizes. Overview authors must be aware of a similar problem when conducting an overview, assess the amount of overlap between reviews, and handle overlap if problematic. Similar to findings in health care overviews (Pieper et al., 2014a), most education overview authors did not assess or address the issue of overlap in any way, and only three accounted for the overlap they found. Cooper and Koenka (2012) summarized various strategies overview authors have used to handle overlap, including selecting the review that is most rigorous, contains the most evidence, provides the most complete description, is the most recent, or is published in a peer-reviewed journal. Overview authors may also choose to disregard overlap all together and include all reviews in the overview. It is not clear which, if any, is the most appropriate approach to handling overlap. Each approach may be justifiable depending on the overview, although none of the approaches are likely to be completely adequate (Cooper & Koenka, 2012). It is clear that the problem of overlap in overviews has not been well addressed and methodological work in this area is needed.

In examining synthesis methods used in the overviews, about half of the overviews employed a descriptive synthesis method, providing a textual summary of the reviews and results from each included review. Summarizing the results of the included reviews was the primary focus, rather than on identifying and analyzing the discordance between reviews. Simply summarizing the results of each study is problematic in the same way traditional descriptive syntheses are problematic for synthesizing primary studies (Combs, Ketchen, Crook, & Roth, 2011). A descriptive overview should maintain a similar methodological rigor to quantitative overviews despite the lack of a quantitative synthesis. At the very least, descriptive overviews can provide a comprehensive portrait of the systematic reviews available.

About half of the overviews, on the other hand, quantitatively synthesized the included reviews in some way. This is far greater than the proportion of overviews in Hartling et al.'s (2012) study, where they found that only 3% of the overviews conducted a quantitative analysis. Of the overviews in the present study that conducted a quantitative synthesis, few reported the specific statistical procedures used to synthesize the results, none corrected for or commented on the diversity of meta-analytic models, and none used the statistical procedures for combining effect sizes across reviews suggested by Schmidt and Oh (2013). Moreover, few studies conducted sensitivity analyses, such as publication bias analyses to evaluate the validity of the population of reviews. Subgroup and moderator analyses were also used infrequently. Given the potential advantages to quantitatively synthesizing reviews, it is important that methods for synthesizing results of reviews be developed and tested.

Although the number of education overviews to date is far fewer than the number of overviews of health related research (Hartling et al., 2012; Li et al., 2012; Pieper et al., 2012), overviews of education research reflect data from hundreds of primary studies and represent hundreds of thousands of students. Indeed, the median number of reviews included in the overviews was 30, with each review including anywhere from 5 to 800 studies. Moreover, overviews are highly cited and as a result have the potential to affect policy, practice, and future research. Using Google Scholar as the source, the overviews in this study were cited a median of 88 times as of February 2015, with one overview cited more than 1,800 times (Lipsey & Wilson, 1993). Given the potential impact of overviews on practice and policy, it is essential that overviews are conducted in a rigorous way to minimize bias and error and provide the most valid results possible. Unfortunately, limited guidance is available to inform the conduct and reporting of overviews (Cooper & Koenka, 2012) and there is a need for more clear conduct and reporting guidelines for overviews similar to those that have been developed for systematic reviews (Hartling et al., 2012; Pieper et al., 2012, 2014a; Thomson et al., 2010).

Preliminary Conduct and Reporting Guidelines for Overviews

Results from the present study revealed significant deficiencies in the conduct and reporting of education research overviews. To ensure the validity and utility of overviews to inform education practice and policy, it is important that the conduct and reporting of overviews improve. Because the nature of an overview follows a similar structure and tone as a systematic review, simply following the standards developed for systematic reviews will greatly improve future overviews. Although the conduct and reporting of overviews can be guided, in large part, by established conduct and reporting guidelines of systematic review methods, important differences remain between a systematic review and an overview that require a distinct set of guidelines. Building on the recommendations made by Cooper and Koenka (2012) and the Cochrane Collaboration (Becker & Oxman, 2008) along with several other sources (Campbell Collaboration, 2014; Chandler, Churchill, Higgins, Lasserson, & Tovey, 2013; Moher et al., 2009; Pieper et al., 2012, 2014a; Pieper, Antoine, Neugebauer, & Eikermann, 2014b, 2014c; Smith, Devane, Begley, & Clarke, 2011; Thomson et al., 2010), we offer the following Preliminary Conduct

and Reporting Guidelines for Overviews. A summary of the Preliminary Conduct and Reporting Guidelines for Overviews can be found in Table 5.

Title and Abstract

The reporting standards for systematic reviews related to titles and abstracts can be applied to overviews. As specified in the PRISMA (Moher et al., 2009) and Meta-Analysis Reporting Standards (MARS; American Psychological Association, 2010) reporting standards, the title should identify the type of study being reported, specifically the title should identify the report as a systematic review or a meta-analysis. Along the same lines, we recommend that the title of an overview also clearly identify the type of study being reported. A number of different terms currently exist to identify the type of study we refer to as an overview; however, we encourage the field to be consistent in the terminology and adopt the term overview of reviews, or more simply overview, used by Cochrane and several others conducting research in this area (e.g., Becker & Oxman, 2008; Cooper & Koenka, 2012; Pieper et al., 2014a; Thomson et al., 2013).

Also similar to PRISMA and MARS reporting standards, we recommend that abstracts for overviews use a structured format and provide a summary of the key components of the study: background and purpose; method (including eligibility criteria, data sources, synthesis method); results (including sample size, characteristics of included reviews and primary studies, quality assessment); and conclusions (including implications and limitations). Although none of the 25 included overviews used a structured format, Weare and Nind (2011) explicitly stated each of the key components in their abstract.

Introduction and Research Questions

The structure of the introduction for a report of an overview is also very similar to any other report of an empirical study. The introduction should provide a summary of the problem under study, why the problem is important, and discussion of prior research and theory related to the problem under investigation. The introduction of an overview, however, differs from reports of primary studies and systematic reviews and meta-analyses in that the introduction should provide a strong rationale for the need and appropriateness of synthesizing multiple reviews as opposed to conducting a first-order review by synthesizing the primary studies. The introduction should also include a clear statement of the research questions, and if appropriate, research hypotheses. Ideally, if the overview is addressing a question related to effectiveness of interventions, the research question should follow the PICOS format, specifying the population, intervention, comparison condition, outcomes, and study design. Of the overviews included in the present study, Torgerson's (2007) review on literacy learning in English provides a good example of summarizing the literature appropriately and stating specific research objectives.

Overview Methods

Standards for the conduct and reporting of data sources and search procedures can be wholly adopted from systematic review conduct and reporting standards.

TABLE 5

Preliminary guidelines for the conduct and reporting of overviews

Item name	Conduct standard	Reporting standard
Title	N/A	Identify the report as an overview, preferably identifying whether descriptive or quantitative synthesis methods were used.
Abstract	N/A	Provide a structured summary of the overview, including background and purpose, methods and analytic procedures, summary of findings, and conclusions, including implications and limitations.
Introduction	N/A	Provide a clear rationale for the overview, discussion of problem under investigation, and prior research.
Research questions	Formulate clear research questions appropriate to one of several purposes of an overview. When applicable, use the PICOS format to formulate the research questions.	State specific research questions, preferably using the PICOS format when applicable.
Method		
Research plan	Plan the overview in advance, specifying purpose, research questions, and methods for the conduct of the overview.	Report whether a review protocol exists and provide information about how it can be accessed.
Eligibility criteria	Define in advance the eligibility criteria for the reviews as well as the primary studies included in the reviews using PICOS and report characteristics. Publication status must not be used to exclude otherwise eligible reviews.	Clearly report the inclusion and exclusion criteria for reviews and any criteria related to primary studies and provide rationale.
Information sources	Search using multiples sources, including at least two electronic library databases, gray literature sources, within overviews of a similar topic, reference lists of included reviews, forward citation searches of included reviews and primary studies.	Report all sources searched, search strategies within those sources and the dates searched.
Search procedures	Develop specific search procedures for each source, preferably with assistance from an information retrieval specialist.	Report in detail all search procedures for each source searched in such a way that the search could be repeated. This can be provided in supplementary material.

(continued)

TABLE 5 (continued)

Item name	Conduct standard	Reporting standard
Study selection	<p>Include relevant eligibility criteria related to both reviews and primary studies included in targeted reviews. Use at least two independent reviewers to screen studies and make final eligibility decisions and document process.</p>	<p>Clearly describe the study selection process at each stage and provide a PRIMSA flow chart to identify the results of each stage of the study selection process. Report reasons for exclusion of reviews.</p>
Data collection	<p>Use at least two reviewers to extract data from reviews using a structured data collection form. Collect data on characteristics of reviews and pertinent data related to the primary studies included in the review.</p>	<p>Describe characteristics of the included reviews and pertinent information about the primary studies in a table of characteristics of reviews.</p>
Assessment of methodological quality		
Quality and risk of bias	<p>Assess the quality and risk of bias of each included review. Also assess quality and risk of bias of included primary studies by conducting a new assessment or taking into account the process used by the review authors for assessing quality and risk of bias of primary studies.</p>	<p>Report results of quality and risk of bias assessment of reviews and quality and risk of bias of included studies (as assessed by review authors or overview authors).</p>
Overlap	<p>Determine strategies for assessing and handling overlap a priori, assess overlap of included reviews (see Pieper et al., 2012, for ways to assess overlap), and handle the overlap if needed (see Cooper & Koenka, 2012 for several ways to handle overlap).</p>	<p>Report results of overlap analysis, minimally using a citation matrix to report the overlap across included reviews. If methods were used to handle the overlap, clearly report how the overlap was handled and what reviews, if any, were excluded due to overlap.</p>
Up-to-dateness	<p>Assess the extent to which the included reviews are up-to-date. Consider including recent primary studies that were conducted after the included review timeframes.</p>	<p>Report the up-to-dateness of the reviews included in the overview.</p>

(continued)

TABLE 5 (continued)

Item name	Conduct standard	Reporting standard
Synthesizing results		
Descriptive synthesis	Minimally, examine and describe the agreement and discordance of included reviews (Cooper & Koenka, 2012).	Provide a detailed reporting of the methods used to synthesize the overviews.
Quantitative synthesis	We refer to Schmidt and Oh (2013) and DerSimonian and Laird (1986) until further developments are made in the quantitative synthesis of overviews.	
Discussion and conclusions		
Summarize the findings	Summarize the findings of the overview, the overall completeness and quality of the evidence, and how the current overview fits into the context of the extant body of research.	
Limitations	Consider and discuss limitations and potential biases at the primary study, review, and overview levels.	
Conclusions	Base conclusions on the evidence of the present overview, taking into account the overall quality of the evidence and limitations at every level. Avoid specific recommendations for practice as factors in addition to the evidence may be important to practitioners and decision makers. Discuss implications for future research.	

Note. These preliminary guidelines were developed using several sources: Becker and Oxman (2008); Campbell Collaboration (2014); Chandler et al. (2013); Cooper and Koenka (2012); Moher et al. (2009); Pieper et al. (2012, 2014a); Pieper, Antoine, Neugebauer, and Eikermann (2014b, 2014c); Thomson et al. (2010).

Additional nuance that overview authors should consider include the need to add information sources and search terms to capture reviews rather than primary studies. Overview authors should also consider contacting both review and primary study authors.

Search and study selection. The search for overviews should follow familiar guidelines present in any systematic review report. Major databases should be searched thoroughly and systematically and the authors would do well to track the quantity and type of overviews retrieved from each. Although publication bias may or may not be an issue in terms of a review being published, it is nevertheless important to search the gray literature. A thorough search of Google Scholar is one place to start, but conference abstracts and relevant research firms are also informative. Should the overview author intend to include primary studies in addition to reviews, these searches should be tailored to retrieve both types of studies. Finally, the overview author should consult a librarian or information retrieval specialist when planning the search.

Study selection procedures for overviews are similar to procedures for selecting primary studies for a systematic review. We recommend using at least two independent reviewers at each stage of the selection process, with transparent reporting of these procedures and decisions at each stage of the selection process. It is important when determining eligibility criteria for study selection that overview authors take care to determine study design criteria at the review level as well as the primary study level. Considering the type of review design (limiting to only systematic reviews and how this is defined) or limiting reviews that include only randomized controlled trials or other study designs are examples of review and methodological characteristics that will need to be considered when setting eligibility criteria for an overview. For instance, Diekstra's (2008) overview on school-based social and emotional education programs included in the present study provided a thorough and clear description of the eligibility criteria.

Data collection. It is standard practice in a systematic review and meta-analysis to use a predetermined coding document and two independent coders. The recommendation for an overview should follow suit, and the overview author should consider conducting regular meetings with the coders to ensure coder drift does not occur. In terms of specific information collected from the overview, we suggest that overview authors report information collected by the review authors and characteristics of primary studies included in the reviews. The reviews contain crucial information that affects the validity of the overviews. Including low-quality or biased reviews relegates the overview to lower quality and biases the results of the overview. Audiences must be able to ascertain aspects of the population, and in this case, the reviews are the population. Without such information, it will be difficult to discern differences across overviews accurately. Lister-Sharp et al. (1999), for example, carefully articulated the coding and data extraction procedures.

Assessment of Methodological Quality

The quality and validity of an overview is dependent on the quality of the included reviews and the primary studies included in those reviews. Thus, it is crucial that overview authors assess the quality of the reviews and the primary studies included in those reviews. This is a much more complex task than that faced by systematic review authors. Overview authors should describe the methods used for assessing the quality of the included reviews and the evidence that is included in those reviews. Several tools are available for assessing methodological quality of reviews (e.g., *Assessing the Methodological Quality of Systematic Reviews*; Shea et al., 2007) and primary study evidence (e.g., *Grading of Recommendation, Assessment, Development, and Evaluation*; Guyatt et al., 2008). The newly created Risk of Bias in Systematic Reviews tool is also now available as an option (Whiting et al., 2016). The tool is geared toward medical research, and as such some items may not be appropriate or applicable to education and social science. Another option is to create a tool specific to the topic area using the guidelines proposed in Cooper (2010) or Lipsey and Wilson (2001). Cooper's (2010) text is especially appropriate, and Table 8.1 (p. 222) is an excellent guide. Given the lack of research on quality assessment of systematic reviews, we will not recommend any one tool for evaluating review or primary study quality or risk of bias; however, it is strongly recommended that the overview authors clearly report their method for assessing methodological quality of included reviews and primary studies included in those reviews and provide rationale for the methods they used.

Overlap. The degree of overlap is a methodological quality issue that needs to be addressed when conducting an overview. Including several reviews that have a high level of overlap could give disproportionate weight to one or a small number of reviews, and thus could bias the results of the overview and lead to erroneous conclusions. Pieper et al. (2014c) described two ways of assessing overlap that overview authors could consider: calculating the "covered area" or the "corrected covered area" (p. 370). Both of these methods use a citation matrix, with the latter method making some adjustments to reduce the influence of a single large review.

Unfortunately, no clear guidance is available on how to best assess or mitigate overlap; however, it is of utmost importance that overview authors plan to assess and handle overlap a priori and examine and report the level of overlap across included reviews. When authors recognize a high level of overlap and choose to handle overlap in some way, it is important that the authors clearly report the methods used to handle overlap and the results of that approach (e.g., clearly identify overviews excluded due to overlap). From the set of included reviews included in this study, Lipsey and Wilson (1993) considered the overlap of primary studies and attempted to ameliorate the issue.

Up-to-dateness. The up-to-dateness of reviews is also important to consider when conducting an overview. Including outdated reviews with older primary studies may not be comparable in terms of relevance or quality of more current reviews

and primary studies, and may disregard more recent studies that have not yet been included in a review. Thus, overviews may be out of date and not reflective of the current state of the evidence. Reflective of the recommendations of Pieper et al. (2014c), we recommend that overview authors attend to the up-to-dateness of the overview. Minimally, authors can examine the age of the studies included in the reviews as well as the reviews themselves and report and discuss the up-to-dateness of the evidence. Calculating the publication lag is another strategy of assessing up-to-dateness (Pieper et al., 2014c). Furthermore, if authors find gaps in the inclusion of more recent evidence, overview authors are encouraged to search for and include recent primary studies in the overview.

Synthesizing Results of Reviews

Methods of synthesizing review results offer unique conduct and reporting challenges over synthesizing primary study results because there are more complexities and little guidance for synthesizing reviews. Nevertheless, there are some key advantages to synthesizing reviews, including the potential to make comparisons among interventions examined in different reviews and the opportunity to employ more sophisticated analyses to allow both direct and indirect comparisons (Thomson et al., 2010). Cooper and Koenka (2012) identified three primary approaches to synthesizing evidence from reviews: examining discordance between reviews, performing second-order meta-analysis, or performing a new meta-analysis by including all of the primary studies that were included in the reviews. We argue, however, that the third option, including all of the primary studies included in the reviews, would then be a new review and not an overview, and we will thus not discuss that option here. Unfortunately, techniques for qualitatively or quantitatively synthesizing reviews are in their infancy and must be further developed. For overview authors who are conducting a descriptive synthesis of reviews, we recommend minimally examining and describing the discordance of included reviews as suggested by Cooper and Koenka (2012). We strongly discourage overview authors from using a vote-counting method, where the authors simply identify the number of reviews that found overall positive effects, null effects, and negative effects.

Although methods for quantitatively synthesizing mean effects across reviews are not well developed, overview authors may have good reason to quantitatively synthesize effects across reviews. When possible, we suggest that overview authors consider quantitatively synthesizing review results. Overview authors must consider, however, the statistical implications of combining average effect size estimates across multiple reviews. To date, only Schmidt and Oh (2013) have put forward procedures for quantitatively synthesizing results from meta-analysis using random effects models. Although random-effects meta-analytic models are becoming commonplace, fixed-effect models are still used (Polanin & Pigott, 2014). Synthesizing random-effects results with fixed-effect results should be treated as cautionary and, at a minimum, discussed as a limitation. We prefer that overview authors not combine these two types of effect size estimates and instead contact study authors for the appropriate results. Alternately, an overview author could estimate the random-effects model using effect sizes reported in the reviews. In addition to techniques for quantitatively synthesizing reviews, overview

authors must consider overlap and take appropriate steps to handle primary study overlap.

Limitations

Although this study is the first to examine education research overviews and contributes to the sparse empirical research on this developing research synthesis method, the findings of this review must be interpreted in light of the study's limitations. The search process, although comprehensive for the subject matter, was constrained to education-related topics. Fields outside of education may conduct superior overviews or regulate the reporting of overviews. This is unlikely, however, and we hope to investigate overviews in other fields in the future. It is also possible that studies are published in languages other than English. We believe the likelihood that many additional overviews exist in other languages is low, but nevertheless we could have missed a few. Although we did not limit our search to the United States, our search yielded only four overviews published outside of the United States. The overviews, however, no doubt included reviews published outside the United States. An additional limitation is our lack of overview quality rating; however, we did code for a variety of overview methodological characteristics that would likely constitute any measure of overview quality and did use the PRISMA checklist to guide our construction of the coding form.

Finally, we are susceptible to the traditional limitations of systematic reviews. Our work should be critiqued as if it were a systematic review and is only as good as the methods we used to collect and synthesize the studies, although we attempted to use systematic review best practices in conducting and reporting the results. Moreover, we are aware of the level of abstraction that comes from dissecting overviews, which are themselves reviews of reviews. We must be cautious when discussing the direct implications of these types of studies, while understanding that researchers are conducting overviews and need guidance.

Conclusion

The overview offers an exciting, yet challenging method for synthesizing and managing the ever-expanding volume of education research. Overviews provide unique opportunities to answer more broad and different research questions than we can answer using primary research or research synthesis methods. The results of this study, however, revealed significant deficiencies in the reporting, conduct, and synthesis of overviews in education research. Thus, caution must be used in interpreting and using results of extant overviews of education research. This study also supports the need for further development of overview methods and quality assessment tools; it is important that empirical work on the methodology for conducting overviews be undertaken to advance this novel synthesis method and inform best practices.

Although conduct and reporting guidelines for systematic reviews are now commonplace and required to be followed by some journals, there has been little guidance for the conduct and reporting of overviews. Due to the added complexity inherent in the multiple levels of an overview, systematic review guidelines are not adequate, and thus, we have offered Preliminary Guidelines for the Conduct and Reporting of Overviews. We hope that the development of overview methods,

particularly methods for quantitatively synthesizing reviews, will follow the rapid progress of systematic review and meta-analytic methods and that these preliminary guidelines are further developed as advances are made. Advancing the science of overview methods will take concerted time and effort, which we believe is necessary given the increase in the use of overview methods and the potential of this method to answer important questions.

Note

Research for the current study was partially supported by an Institute of Education Sciences Postdoctoral Training Fellowship Grant to Vanderbilt University's Peabody Research Institute (R305B100016). Opinions expressed herein do not necessarily reflect the opinions of the Institute of Education Sciences.

References

Studies marked with an asterisk were included in the review.

- American Psychological Association. (2010). *Meta-Analysis Reporting Standards*. Retrieved from <https://www.apa.org/pubs/journals/features/arc-mars-questionnaire.pdf>
- *Anderson, R. D. (1983). A consolidation and appraisal of science meta-analyses. *Journal of Research in Science Teaching*, 20, 497–509. doi:10.1002/tea.3660200511
- Apple, Inc. (2014). Filemaker Pro [computer software]. Santa Clara, CA: Author.
- Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Medicine*, 7(9), e1000326. doi:10.1371/journal.pmed.1000326
- Becker, L. A., & Oxman, A. D. (2008). Overviews of reviews. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. Chichester, England: John Wiley.
- *Browne, G., Gafni, A., Roberts, J., Byrne, C., & Majumdar, B. (2004). Effective/efficient mental health programs for school-age children: A synthesis of reviews. *Social Science & Medicine*, 58, 1367–1384. doi:10.1016/S0277-9536(03)00332-0
- Campbell Collaboration. (2014). Campbell collaboration systematic reviews: Policies and guidelines. *Campbell Systematic Reviews: Supplement 1*. doi:10.4073/csr.2014.1
- Chandler, J., Churchill, R., Higgins, J., Lasserson, T., & Tovey, D. (2013). *Methodological standards for the conduct of new Cochrane Intervention Reviews*. Retrieved from http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards%202.2%2017122012_0.pdf
- *Cobb, B., Lehmann, J., Newman-Gonchar, R., & Alwell, M. (2009). Self-determination for students with disabilities: A narrative metasynthesis. *Career Development for Exceptional Individuals*, 32, 108–114. doi:10.1177/0885728809336654
- Combs, J. G., Ketchen, D. J., Jr., Crook, T. R., & Roth, P. L. (2011). Assessing cumulative evidence within “macro” research: Why meta-analysis should be preferred over vote counting. *Journal of Management Studies*, 48, 178–197. doi:10.1111/j.1467-6486.2009.00899.x
- *Cook, C. R., Gresham, F. M., Kern, L., Barreras, R. B., Thornton, S., & Crews, S. D. (2008). Social skills training for secondary students with emotional and/or behavioral disorders: A review and analysis of the meta-analytic literature. *Journal of Emotional and Behavioral Disorders*, 16, 131–144. doi:10.1177/1063426608314541

- Cooper, H. (2010). *Research synthesis and meta-analysis* (4th ed.). Thousand Oaks, CA: Sage.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, *67*, 446–462. doi:10.1037/a0027119
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*, 177–188. doi:10.1016/0197-2456(86)90046-2
- *Diekstra, R. F. (2008). Effectiveness of school-based social and emotional education programmes worldwide. In C. Clouder, B. Dahlin, & R. F. Diekstra (Eds.), *Social and emotional education: An international analysis* (pp. 255–312). Santander, Spain: Fundación Marcelino Botín.
- *Forness, S. R. (2001). Special education and related services: What have we learned from meta-analysis? *Exceptionality*, *9*, 185–197. doi:10.1207/S15327035EX0904_3
- *Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987a). Contextual and transactional influences on science outcomes. *International Journal of Education Research*, *11*, 165–186.
- *Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987b). Identifying the salient facets of a model of student learning: A synthesis of meta analyses. *International Journal of Educational Research*, *11*, 187–212.
- *Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987c). Syntheses of research on factors influencing learning. *International Journal of Education Research*, *11*, 155–164.
- Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews* (1st ed.). London, England: Sage.
- *Green, J., Howes, F., Waters, E., Maher, E., & Oberklaid, F. (2005). Promoting the social and emotional health of primary school-aged children: Reviewing the evidence base for school-based interventions. *International Journal of Mental Health Promotion*, *7*, 30–36. doi:10.1080/14623730.2005.9721872
- *Gresham, F. M. (1998). Social skills training: Should we raze, remodel, or rebuild? *Behavioral Disorders*, *24*, 19–25.
- *Guthrie, J. T., Seifert, M., & Mosberg, L. (1983). Research synthesis in reading: Topics, audiences, and citation rates. *Reading Research Quarterly*, *19*, 16–27. doi:10.2307/747334
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*, 924 - 926. doi: 10.1136/bmj.39489.470347.AD
- Hartling, L., Chisholm, A., Thomson, D., & Dryden, D. M. (2012). A descriptive analysis of overviews of reviews published between 2000 and 2011. *PLoS One*, *7*(11), e49667. doi:10.1371/journal.pone.0049667
- *Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, *36*(1), 5–13.
- *Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Oxon, England: Routledge.
- Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. Chichester, England: John Wiley.

- *Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The impact of digital technology on learning: A summary for the education endowment foundation*. Retrieved from http://educationendowmentfoundation.org.uk/uploads/pdf/The_Impact_of_Digital_Technology_on_Learning_-_Executive_Summary_%282012%29.pdf
- Ioannidis, J. P. A. (2009). Integration of evidence from multiple meta-analyses: A primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Canadian Medical Association Journal*, *181*, 488–493. doi:10.1503/cmaj.081086
- Kazrin, A., Durac, J., & Agteros, T. (1979). Meta-meta analysis: A new method for evaluating therapy outcome. *Behaviour Research and Therapy*, *17*, 397–399. doi:10.1016/0005-7967(79)90011-1
- *Kulik, J. A., & Kulik, C. L. C. (1989). The concept of meta-analysis. *International Journal of Educational Research*, *13*, 227–340. doi:10.1016/0883-0355(89)90052-9
- Li, L., Tian, J., Tian, H., Sun, R., Liu, Y., & Yang, K. (2012). Quality and transparency of overviews of systematic reviews. *Journal of Evidence-Based Medicine*, *5*(3), 166–173.
- *Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209. doi:10.1037//0003-066X.48.12.1181
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- *Lister-Sharp, D., Chapman, S., Stewart-Brown, S., & Sowden, A. (1999). Health promoting schools and health promotion in schools: Two systematic reviews. *Health Technology Assessment*, *3*, 1–207.
- *Lloyd, J. W., Forness, S. R., & Kavale, K. A. (1998). Some methods are more effective than others. *Intervention in School and Clinic*, *33*, 195–200. doi:10.1177/105345129803300401
- *Maag, J. W. (2006). Social skills training for students with emotional and behavioral disorders: A review of reviews. *Behavioral Disorders*, *32*, 4–17.
- Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly*, *23*, 26–42. doi:10.1037/1045-3830.23.1.26
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(6), e1000097. doi:10.1371/journal.pmed1000097
- Pieper, D., Antoine, S. L., Morfeld, J. C., Mathes, T., & Eikermann, M. (2014a). Methodological approaches in conducting overviews: Current state in HTA agencies. *Research Synthesis Methods*, *5*, 187–199. doi:10.1002/jrsm.1107
- Pieper, D., Antoine, S. L., Neugebauer, E. A., & Eikermann, M. (2014b). Systematic review finds overlapping reviews were not mentioned in every other review. *Journal of Clinical Epidemiology*, *67*, 368–375. doi:10.1016/j.jclinepi.2013.11.007
- Pieper, D., Antoine, S. L., Neugebauer, E. A., & Eikermann, M. (2014c). Up-to-dateness of reviews is often neglected in overviews: A systematic review. *Journal of Clinical Epidemiology*, *67*, 1302–1308. doi:10.1016/j.jclinepi.2014.08.008
- Pieper, D., Buechter, R., Jerinic, P., & Eikermann, M. (2012). Overviews of reviews often have limited rigor: A systematic review. *Journal of Clinical Epidemiology*, *65*, 1267–1273. doi:10.1016/j.jclinepi.2012.06.015
- Pigott, T. D. (2012). *Advances in meta-analysis*. New York, NY: Springer.
- Polanin, J. R., & Pigott, T. D. (2014). The use of meta-analytic statistical significance testing. *Research Synthesis Methods*, *6*, 63–73. doi:10.1002/jrsm.1124

- R Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.3) [Software]. Available from: <https://www.r-project.org/>
- Schmidt, F. L., & Oh, I. S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, *121*, 204–218. doi:10.1016/j.obhdp.2013.03.002
- Shadish, W. R., & Lecy, J. D. (2015). The meta-analytic big bang. *Research Synthesis Methods*. Advanced online publication. doi:10.1002/jrsm.1132
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*, 10. doi:10.1186/1471-2288-7-10
- *Sipe, T. A., & Curlette, W. L. (1997). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing meta-analyses. *International Journal of Educational Research*, *25*, 583–698.
- Smith, V., Devane, D., Begley, C. M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology*, *11*, 15. doi:10.1186/1471-2288-11-15
- *Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, *81*, 4–28. doi:10.3102/0034654310393361
- *Tennant, R., Goens, C., Barlow, J., Day, C., & Stewart-Brown, S. (2007). A systematic review of reviews of interventions to promote mental health and prevent mental health problems in children and young people. *Journal of Public Mental Health*, *6*, 25–32. doi:10.1108/17465729200700005
- Thomson, D., Foisy, M., Oleszczuk, M., Wingert, A., Chisholm, A., & Hartling, L. (2013). Overview of reviews in child health: Evidence synthesis and the knowledge base for a specific population. *Evidence-Based Child Health: A Cochrane Review Journal*, *8*, 3–10. doi:10.1002/ebch.1897
- Thomson, D., Russell, K., Becker, L., Klassen, T., & Hartling, L. (2010). The evolution of a new publication type: Steps and challenges of producing overviews of reviews. *Research Synthesis Methods*, *1*, 198–211. doi:10.1002/jrsm.30
- *Torgerson, C. J. (2007). The quality of systematic reviews of effectiveness in literacy learning in English: A “tertiary” review. *Journal of Research in Reading*, *30*, 287–315. doi:10.1111/j.1467-9817.2006.00318.x
- Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, *7*, 27–56. doi:10.1007/s11292-010-9109-1
- Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods*, *13*, 130–149. doi:10.1037/10889X.13.2.130
- *Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*, 249–294. doi:10.3102/00346543063003249
- *Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: What does the evidence say? *Health Promotion International*, *26*, 29–69. doi:10.1093/heapro/dar075

- What Works Clearinghouse. (2015). *Procedures and standards handbook* (Version 3). Washington, DC: Institute of Education Sciences, Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf
- Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., ... Churchill, R. (2016). ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225-234.
- Williams, R. T. (2012). *Using robust standard errors to combine multiple estimates with meta-analysis* (Unpublished doctoral dissertation). Loyola University, Chicago, IL.

Authors

- JOSHUA R. POLANIN, PhD, is a senior research scientist at Development Services Group, 7315 Wisconsin Ave., Suite 800 East, Bethesda, MD, 20814, USA; email: jpolanin@dsgonline.com. He recently completed a Postdoctoral Fellowship at Vanderbilt University's Peabody Research Institute. His methodological research has focused on the use of statistical significance testing in meta-analysis and investigated the potential publication bias in gray literature. He is the managing editor of the Campbell Collaboration's Methods Group, providing methodological expertise to potential Campbell Collaboration reviewers. He has served as the methodological expert and statistician for a large-scale, cluster-randomized trial of bullying prevention programs, and recently started as a methodological consultant on IES' What Works Clearinghouse Postsecondary research reviews.
- BRANDY R. MAYNARD, PhD, is an assistant professor at Saint Louis University, 3550 Lindell Blvd, St. Louis, MO 63103, USA; email: bmaynar1@slu.edu. She is also the cochair of the Campbell Collaboration Social Welfare Coordinating Group. Her major research interests include the etiology and developmental course of academic risk and externalizing behavior problems; evidence-based practice, the use, and quality of research in education and social sciences; and research synthesis and meta-analysis.
- NATHANIEL A. DELL, MA, is a master of social work student and graduate research assistant at Saint Louis University, 3550 Lindell Blvd, St. Louis, MO 63103, USA; email: nathanieldell@gmail.com. He received a master of arts in the humanities from the University of Chicago. His major research interests include ethics of care, community mental health, and narrative identity construction. He also lectures at Southwestern Illinois College in the Division of Liberal Arts.