

## Methodological Guidance Paper: The Art and Science of Quality Systematic Reviews

**Patricia A. Alexander**  
*University of Maryland*

*The purpose of this article is to overview various challenges that prospective authors of quality systematic reviews should be prepared to address. These challenges pertain to all phases of the review process: from posing a critical question worthy of pursuit and executing a search procedure that is appropriately framed and transparently recorded, to discerning patterns and trends within the resulting data that speak directly to the critical question framing the review. For each of these challenges, suggestions are offered as to how authors might respond so as to enhance the quality of the review process and increase the value of findings for educational research, practice, and policymaking.*

**KEYWORDS:** systematic review, literature search, critical questions

In their inaugural issue of the *Review of Educational Research* (RER), the newly appointed editorial team of Murphy, Knight, and Dowd (2017, p. 4) lauded review articles as the scholarly community's efforts to "look deeper into the edifice of education research to the bricks and mortar with which that structure is built." Moreover, they held that review articles are uniquely positioned to afford a new vantage point on an educational topic or question. Because of that new vantage point, a quality review "has the potential to shape the future of research and practice" (p. 4). The importance of literature reviews to scholarly communities is evidenced by the fact that review journals are among the most highly cited publications in the field, with RER topping that list. Yet, with such importance comes heightened responsibility to ensure quality of every review privileged to fill the pages of those journals. That responsibility rests not only on the shoulders of editors and editorial boards but first and foremost on the backs of authors who labor to carry out those reviews.

To assist authors in this process, my goal in this methodological guidance paper is to articulate various challenges that authors undertaking a systematic review of the literature are likely to confront. The need for methodological guidance in the development of systematic reviews is by no means new or specific to

**TABLE 1**

*Challenges confronted in conducting a systematic review*

---

Framing challenges

- Finding a “critical question” worthy of review
  - Situating the review theoretically
  - Demonstrating the merits of the review
  - Considering the maturity of the field of inquiry
- Posing an unanswered but answerable critical question
  - Establishing what is already known and what needs to be known
  - Articulating an answerable question

Procedural challenges

- Formulating appropriate search parameters
  - Specifying the corpus of search terms
  - Establishing the right time frame
  - Delimiting the search
  - Extending search results
- Establishing justifiable inclusion and exclusion criteria
  - Adhering to quality standards
  - Tying back to critical questions

Consolidating and summarizing challenges

- Recording the basics
- Charting other relevant characteristics
- Considering potentially informative groupings

Interpreting and communicating challenges

- Recognizing meaningful outcomes
  - Capturing significant patterns and trends
  - Communicating contributions
    - Acknowledging delimitations and limitations
    - Returning to overall purpose and critical question
- 

educational research. In fact, there was a growing awareness within the medical professions that systematic reviews, as with meta-analytic reviews, were often poorly conducted. For that reason, Moher, Liberati, Tetzlaff, and Altman (2009) set out to establish guidelines for the medical profession engaged in such reviews. Those guidelines, commonly referred to as the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, touch on certain points I overview here as challenges for the educational research community.

These challenges, summarized in Table 1, arise as authors frame, conduct, and conclude their reviews, and when they attempt to communicate their findings to diverse audiences. For each of these possible challenges, I proffer certain responses that authors may wish to consider. Let me state at the outset that I do not presume that the challenges discussed herein represent an exhaustive list or that all apply to every systematic review authors initiate. Nonetheless, I contend that

these challenges are a good starting point for both novice and experienced authors to ponder as they take on the responsibility of conducting a quality systematic review. Moreover, while I have conveyed this process in an orderly and ordered manner, I grant that, in reality, conducting a systematic review is often iterative and nonlinear, as authors encounter unexpected barriers, backtrack in their thinking, or explore alternative paths to completion.

Throughout this methodological guidance paper, I use the phrase *systematic review* to designate both a process and a product. In effect, the process of conducting a review systematically entails an orchestrated search of literature intended to address a particular question or issue of importance to the field. The identified literature is then analyzed and synthesized in a manner that is methodical, logical, and transparent, and the conclusions and implications that arise from this process are, thus, well-grounded in that examined literature. The word *science* in the title of this methodological guidance paper is meant to highlight these very characteristics, albeit in a manner that remains broad and encompassing of varied methodological approaches and diverse theoretical orientations toward the nature of scientific inquiry.

Of course, the quality of a systematic review cannot be relegated to science alone. As I will endeavor to show, there is an *art* to systematic reviews that break new ground, offer novel insights, or instigate critical dialogue. In other words, a quality systematic review is so much more than a mere exercise in locating, charting, or tallying literature, even in a rigorous manner. Such reviews require a critical eye—the ability to recognize the significant but often less discernible elements that comprise the “bricks and mortar” of educational research. They also call for a creative hand—the ability to synthesize and communicate the importance of what was learned in an articulate and appropriate manner. Ultimately, it is the confluence of art and science that shapes a quality systematic review and marks it as unique form of scholarly writing. It is also why such reviews can prove so challenging to those who dare to undertake one.

Let me now turn to those challenges, which I have organized under the banners of framing, procedural, consolidating and summarizing, and interpreting and communicating. *Framing* challenges pertain to those that are encountered at the outset, when the idea of undertaking a systematic review is first contemplated. In contrast, *procedural* challenges arise when the search for relevant literature begins and continues until the final collection of works has been selected. Then, there are the *consolidating* and *summarizing* challenges, which entail the creation of the literature database through the analysis, organization, and cataloguing of key features and elements of the identified literature. Finally, the challenges that come with *interpreting* and *communicating* pertain to the discernment of relevant patterns and significant trends within the literature and the effective articulation of their nature and importance to others. As I describe these challenges, I will also draw on my own experiences and those of my current and former students as we engaged in reviewing the literature systematically. My goal is not to be egocentric or to place undue emphasis on my own field of educational psychology but, rather, to illustrate the somewhat abstract challenges through personal experiences I know firsthand.

## Framing Challenges

### *Finding a “Critical Question” Worthy of Review*

It is perhaps self-evident to say that every quality systematic review begins with a good research question. The quandary comes in knowing what counts as a *good* question. There is no simple formula that can be applied to make this determination, but there are certainly general guidelines that can be helpful.

### *Situating the Review Theoretically*

All educational research occurs within a context that serves to shape both the process and product, whether that context is explicitly acknowledged or not. One aspect of that overall context is the theoretical orientation that authors bring to the research endeavor. When it comes to systematic reviews, for instance, authors’ theoretical orientations are likely reflected in the topics or issues they find intriguing and worthy of pursuit. These theoretical orientations can also influence what critical questions are posed and how, which bodies of literature are examined, and, ultimately, what emerges as findings. Even if authors take a broad, interdisciplinary, or even critical view of the constructs and variables they are pursuing, that theoretical context will nonetheless play a role in what is unearthed.

For example, in the 1980s, I was deeply immersed in the study of text-processing strategies and concomitantly in the study of the knowledge acquired from reading. During this timeframe, information-processing theory was a prevailing theoretical framework for understanding strategic processing and knowledge acquisition. In my research, I became aware that within these related literatures there was no explicit consideration given to the reciprocal nature of readers’ strategic behaviors and their knowledge of the topic or domain in question. That moment of awareness led me to undertake the well-cited review on “The Interaction of Domain-Specific and Strategic Knowledge in Academic Performance” (Alexander & Judy, 1988<sup>1</sup>).

What this example is meant to illustrate is that the critical questions authors put to the literature will unquestionably be shaped by their theoretical perspective. What is posed as a viable and important question to researchers operating within one theoretical framework will not necessarily be seen as viable or important to researchers coming from a different framework. Later, I will discuss the notion of “delimitations” or the boundaries authors inevitably construct around their reviews. The theoretical context is perhaps the most general and potentially influential of those delimitations. Thus, authors of systematic reviews need to be cognizant of the theoretical context in which they are functioning and sensitive to how that context affects each phase of the review process.

### *Demonstrating the Merits of the Review*

With the context for the systematic review established, authors still must ensure that there is a value to the question or questions they pose—a value that extends beyond their own personal curiosity or self-serving interests. Moreover, authors must be prepared to convincingly communicate that broader value to readers. For educational journals like *RER*, that broader value should have educational

importance or relevance, as well. Murphy et al. (2017) refer to such valuable, important, and relevant questions as *critical questions*.

As is also true when framing empirical investigations, it is not enough to declare that a particular question is critical simply because it has not been subjected to review before. The more determinative issue is *why* this question warrants a meticulous review of the educational literature at this point in time. Moreover, those asking this question of the literature must not assume they already know the answer and are, therefore, using the guise of a systematic review to share their “truth” with others. Should that occur, the authors have crossed into the realm of an argumentative review. There is certainly a time and place for using the literature to argue or persuade. However, such laudable goals are not well suited to a systematic review as conceptualized here. That is because the question or questions that frame the interrogation of the literature for a systematic review should represent a honest “need to know” for the authors and for the wider educational research community.

Yet, almost by default, this “need to know” must represent some knowledge or exposure to the topic or issue on the authors’ part, and perhaps a cursory exploration of the literature that helps ensure that the proposed question has not already been thoroughly addressed. An initial exposure to the literature was P. Karen Murphy’s path to a highly cited review on motivation terminology (Murphy & Alexander, 2000). As part of a doctoral course, Murphy was exposed to various readings on specific motivation constructs. In her analysis of those readings, she struggled to find explicit or consistent definitions for many of the constructs being considered. A critical question regarding how key constructs were defined in the motivation literature and the consistency of those definitions was thus formulated and became the catalyst for a systematic review. This systematic review concluded with a glossary for 17 motivation terms and working definitions derived from the explicit and implicit information that had been charted.

### *Considering the Maturity of the Field of Inquiry*

Even if a question appears to be meritorious on the surface, this does not ensure that it meets the standards for a systematic review. Research on the topic or issue that undergirds the critical question must be sufficiently mature to support an in-depth interrogation but not so long established that there are relatively few novel questions left to pose. That was a situation my doctoral student, Anisha Singh, confronted recently when weighing the possibility of a systematic review on the effects of listening to audiobooks versus reading them in print on comprehension. What she found was that there were very few empirical studies in which participants of any age read *and* listened to comparable texts. Thus, despite the growing popularity of audiobooks and the theoretical discussions of the perceived benefits and difficulties associated with audiobooks, the empirical literature was not mature enough to handle her question. Even broadening her search to varied fields such as media production and acoustics did not alleviate the problem.

### *Posing an Unanswered but Answerable Critical Question*

Once an initial judgment has been made about the viability of a topic or issue for in-depth review, framing challenges still remain.

### *Establishing What Is Already Known and What Needs to Be Known*

One of those challenges is to ensure that the rationale for the current pursuit is lucidly and strongly made. This rationale should speak directly to the educational relevance of the topic or issue at the heart of the systematic review. It should also establish the need for this current interrogation of the literature by offering a concise and informative overview of what has already been empirically established and what still remains to be explored. It is this articulation of the extant literature and the important gaps in understanding that persist that demonstrates the “critical” nature of the question or questions authors propose to address. Shaw (2010, p. 39) reiterates the importance of gathering evidence about what is already known about the area of inquiry—a process that is “implicit in the word, ‘re-search.’” In effect, the act of framing a systematic review requires researching the literature to ensure that there is a novel yet significant question to explore.

### *Articulating an Answerable Question*

In empirical studies, the theoretical framework typically concludes with research questions that arise from identified gaps in a field of study. These questions are sometimes accompanied by hypotheses that explain what researchers expect to unearth based on what is already known. Similarly, the critical questions in systematic reviews should not only build on authors’ sense of unanswered questions in the literature, but they should also reflect authors’ expectations that there is, in fact, information within the extant literature that can help address those lingering questions. In this way, the “good” question for a systematic review is unanswered but seemingly answerable. Of course, the added challenge for those conducting a systematic review is that they are not free to design and conduct a new study that would directly address their critical questions. They are limited to what the existing literature is able to reveal. Thus, authors are wise to keep that major constraint in mind when they finally put their questions into words, being careful to ask a question that is potentially answerable by a review of the literature, which by its nature offers no definitive or causal outcomes.

## **Procedural Challenges**

With the framing of the systematic review in place, the process of carrying out what has been promised begins. There are multiple steps in the process that must be carefully planned, explicitly justified, and thoroughly documented. Overall, the procedural actions that ensue encompass the *what*, *where*, *when*, and *how* of a systematic review. Moreover, each of those actions introduces its own set of challenges that authors of systematic reviews must be prepared to address.

### *Formulating Appropriate Search Parameters*

Among the hallmarks of a systematic review are the parameters that authors set for the search of the literature. Those parameters include the search terms used to identify potentially relevant works, the databases to which those search terms will be applied, alternative avenues for locating pertinent literature, and more. Although this phase of the systematic review may seem somewhat more prescribed, there are many crucial decisions that authors must make that will markedly influence the outcomes, beginning with the search terms that are set.

Similarly, the choice of databases to search can differentially shape the outcome of systematic reviews. That is because there are certain databases (e.g., ERIC or Google Scholar) that are broader and more encompassing than others (e.g., PsycInfo or Web of Science). The use of these more encompassing databases opens the door to more diverse research documents that might not be considered under more traditional literature searches. At the same time, tapping the broader databases can bring works of much more variable quality into the initial document pool.

### *Specifying the Corpus of Search Terms*

The process of building the initial pool of scholarly writings that will undergo further examination begins with the delineating of terms to be applied to selected databases. The better researchers are at specifying those search terms, the better search engines can operate in finding potentially relevant documents. That means authors must be thorough in stipulating the key constructs and variables associated with their critical questions, as well as any other alternative terms that have been used for those constructs and variables in other domains of inquiry. Especially when aspects of the critical question have been investigated in diverse fields, there can be quite different words used to signify the same construct or variable. For example, when Dumas, Alexander, and Grossnickle (2013) undertook a review of relational reasoning, they had to determine what general terms had been used in several relevant domains (e.g., philosophy, neuroscience, and mathematics/science) to represent this cognitive process (e.g., higher order thinking, comparison/contrast, and patterning) or its specific forms (anomaly, outlier, aberration, or exceptionality). That delineation then became the basis for their literature search.

### *Establishing the Right Timeframe*

Beyond the vital decision of which terms should guide the literature search, authors must make a series of related judgments that will directly affect what emerges as the initial database for their review. One of those determinations is the timeframe for the search, which should be logically set based on what is known about the sphere of inquiry and any pertinent events, factors, or products. For example, when Lauren Singer Trakhman initiated a systematic review on the effects of reading in print or digitally on comprehension (Singer & Alexander, 2017<sup>2</sup>), there were several conditions that proved especially salient.

One condition that Singer and Alexander (2017) had to consider was the rapid changes in digital technologies over the past decade. Those changes were important to their literature search because drastic differences in the technology used to deliver text could dramatically affect reading comprehension, thus confounding findings from older and new empirical studies. Another determinative event was the publication of a literature review in 1992 (Dillon, 1992) that addressed a similar critical question. By combining these two factors, Singer Trakhman decided to begin her search in 1992, and explained the reasons for her decision in the final document. As this latter point suggests, whatever timeframe authors ultimately set for their literature search must be accompanied by a justification that conveys the events, factors, or products that informed their decision.

### *Delimiting the Search*

The same contemplation and justification applied to the timeframe needs to go into the delimitations that are typical parts of the search parameters. Unlike *limitations*, which are the unfortunate and unplanned factors that arise during a study and which can negatively affect results, *delimitations* refer to the purposeful restrictions that researchers place on their search plan, with an awareness of what those restrictions may mean for the outcomes they report. As part of a well-explicated search procedure, I would strongly suggest that authors not only justify the specific delimitations they have instituted but also have consideration of the potential consequences of those decisions. For example, it is quite common for authors in English-speaking countries who are not fluent in other languages to restrict their search to works published in English. While understandable, this delimitation does mean that relevant research published in other languages will be overlooked, potentially giving greater weight to research occurring in Western countries or among researchers who are skilled at academic writing in English.

Another crucial delimitation pertains to what forms and sources of scholarly writings authors will consider in their search and analysis. This particular procedural concern brings up the challenge of *if* and *when* gray literature will be incorporated in authors' systematic reviews. The term *gray literature* refers broadly to any nontraditional research documents that are found outside of traditional commercial, organizational, or academic publishing venues, most notably journals. Common forms of gray literature include dissertations and theses, governmental or organizational reports or white papers, working papers from research centers or laboratories, and conference proceedings. Determining what manner of gray literature (if any) might be part of a systematic review is a difficult challenge and should never be decided out of hand. Rather, that decision should be based on a rationale that is made transparent to readers. But what issues or factors should be part of this determination?

Generally speaking, there are pros and cons that are associated with these alternative, nontraditional sources and forms of research writings. Among the "cons" that are often discussed are (a) the variable quality of research found within the gray literature and (b) the difficulty in accessing these nontraditional works (Rothstein & Hopewell, 2009). On the "pro" side, there is the repeated argument about publication bias within certain fields or for particular topics (Garousi, Felderer, & Mäntylä, 2019; Rothstein, Sutton, & Borenstein, 2006). Indeed, the mainstream publication outlets rarely publish studies with nonsignificant effects. This can lead to the projection of a slanted view with regard to the effectiveness of a given technique or procedure.

Although each systematic review represents a unique case that must be judged on its own merits, let me suggest a few points to consider when resolving the issue of traditional versus gray literatures. First, within certain domains, such as medicine, engineering, physics, and technology, changes can occur at such a rapid rate that there is a need to access research prepublication. Thus, when the critical questions for educational researchers touch on such rapidly changing domains, there may be good reason to access working papers, conference papers, dissertation studies, and the like. Moreover, in particular educational research communities, published conference proceedings are *de rigueur*. If authors were undertaking a



review pertaining to intelligent tutoring systems, for example, many viable sources would take the form of published conference proceedings. Another reason for considering gray literature would be the relative newness of a field of study. For a still nascent field, remaining solely in the traditional literature may result in an underrepresentation of pertinent research.

### *Extending Search Results*

For the relative novice at systematic reviews, there are any number of articles, books, reports, and websites that offer guidelines for conducting the search portion of the review (Hart, 2018; Okoli & Schabram, 2010; Shaw, 2010). One common recommendation in these sources is to look beyond the results of the database searches to ensure that important and pertinent research documents have not been overlooked. This is a useful recommendation since the algorithms that search engines employ are not perfect. Therefore, works that do not contain the search terms in a fashion or form the search algorithm can recognize will not be catalogued. For that reason, it is suggested that authors should take several additional steps to maximize their initial pool of possibly useful documents. I recommend that three actions be routinely executed: referential backtracking, researcher checking, and journal scouring or hand searching. *Referential backtracking* entails examining the references from highly salient documents to identify research that may fit within search parameters. It is not uncommon to use chapters, reviews, or research reports as the bases for referential backtracking even though those documents fall outside the authors' search parameters (e.g., empirical studies).

Also, authors may find that the name of certain researchers or research groups appear with some frequency in search results. When that occurs, authors can employ a *researcher checking* strategy. For this strategy, authors look up the publication record of those identified contributors to determine if any additional works warrant further analysis. One additional source of potentially relevant documents are the journals that appear with some regularity in research results. *Journal scouring* or *journal hand searching* involves examining the contents of those journals for a specified period of time. Which of these search strategies authors employ and the number and characteristics of the documents located by each technique should be documented and reported as part of transparent search procedures.

### *Establishing Justifiable Inclusion and Exclusion Criteria*

When the aforementioned challenges have been addressed, the outcome should be a viable pool of research documents that merit subsequent scrutiny. It is that subsequent scrutiny that will culminate into the final collection of research documents that will serve as the database for the systematic review. Setting the criteria that will be applied consistently to all documents in the initial pool to ascertain what ultimately becomes the bases for the systematic review is, unquestionably, one of the more significant challenges that authors face. In effect, authors must decide what standards will guide their decisions as to which works inhabiting the search pool will be retained and which will be purged. Authors must also be prepared to justify those standards and implement them in a reliable and replicable manner.

### *Adhering to Quality Standards*

When it comes to systematic literature reviews, the quality of the studies included matters to the trends and patterns that are ultimately identified. However, while quality does matter, it can be especially challenging to fairly and thoughtfully judge what counts as quality research and what does not. That is one reason that authors sometimes elect to stay with traditional versus gray publications, since those works have typically been peer reviewed—a process that involves quality judgments. Yet, even in those instances, authors of systematic reviews cannot simply accept the results of their searches as representing quality, without further analysis of the individual works. When gray literature is part of the search parameters, even more scrutiny of the research that finds its way into the “pool” is warranted. The challenge, of course, is to establish sensible criteria for making that determination without becoming overly restrictive or overly permissive.

In this regard, what I recommend to authors is to be guided by their critical questions and what they are attempting to extract as data from the literature to address those questions. Minimally, a quality study will provide data relevant to authors’ critical questions, and do so in a way that indicates the reliability and validity of those data. Let me clarify this recommendation by referring to a recent review. Specifically, Emily Grossnickle (2016) sought to bring greater clarity to the construct of curiosity by disentangling it from seemingly related constructs, most notably interest. Even staying with traditional publication venues, her initial search resulted in a pool of 851 articles. To cull the pool, she established several inclusion/exclusion criteria. First, she considered only empirical studies, excluding any theoretical or discussion pieces. Additionally, in keeping with her purpose, she retained only studies that were related to educational contexts or that addressed educational or academic outcomes. Because of her particular focus on curiosity in relation to interest, she also decided to keep studies that expressly looked at *both* curiosity and interest even if they were not undertaken in an educational context. Finally, if there was no dedicated measure or no measurement of curiosity, studies were removed. At each of these decision points, Grossnickle provided an example of a specific study that had been accordingly excluded. Her final set of studies meeting all inclusion criteria was 39 articles.

### *Tying Back to Critical Questions*

What the preceding example illustrates is that each inclusion or exclusion criterion authors set should relate meaningfully to the critical question authors are putting to the literature. While the Grossnickle (2016) study was expressly about specific educational constructs, authors posing questions pertaining to special populations, developmental differences, measurement concerns, the effectiveness of various techniques or interventions, the influence of social-contextual conditions, and much more should follow a similar procedure. In effect, whatever constructs or variables are central to the authors’ critical question must be adequately addressed in the research that is ultimately retained. Tangentially related documents, regardless of their quality as research, should be considered for elimination.

By adopting this guideline, authors are not attempting to broadly judge the quality of a research study or document—which Cooper (2010) and others (Glass,

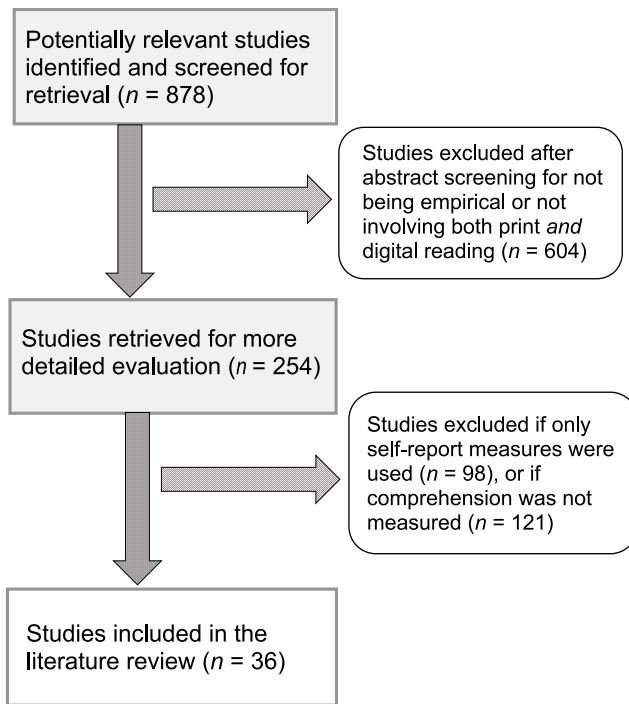


FIGURE 1. *Version of chart from Singer and Alexander (2017, p. 1012) summarizing the steps the authors took in the culling process with data as to the number of initially identified works that were retained or excluded.*

1976; Hattie & Hansford, 1984; Okoli & Schabram, 2010) regard as a precarious venture. Instead, criteria and decisions remain targeted on dimensions that matter for authors' particular purposes as framed in their critical questions. What is essential, however, is that regardless of the criteria that authors set, they should meticulously document what is removed at each step in the culling process and provide an example that further illuminates what kind of research was directly affected. Recently, those publishing systematic reviews in *RER* have begun to include tables or charts that visually summarize each step in the culling process, along with the total number of works that were removed at each step. I am sharing one example based on the Singer and Alexander (2017) review mentioned earlier (Figure 1). Guidelines for generating such a flow diagram for systematic reviews (and meta-analyses) can be found on the PRISMA website (<http://www.prisma-statement.org/>).

### **Consolidating and Summarizing Challenges**

With the final corpus identified, the ensuing challenges for authors pertain to capturing the essentials of documents that address their critical questions. The

most common method of encapsulating those data are tables or charts in which key content is cataloged. For the sake of simplicity, I will refer to such visual consolidations as *charts*. However, let me make it clear that there is no requirement that any chart be present for a literature review to qualify as a systematic review. Nonetheless, the presence of some summary graphic adds to the transparency of the review, just as delineation of search procedures contributes to its replicability.

### *Recording the Basics*

The purpose for the chart, as stated, is to encapsulate or condense the review information into a more readable and easily analyzed form. To achieve this encapsulation, it is useful to chart data in some abbreviated but comprehensible way. This can be done by using codes or symbols within certain informational categories, accompanied by a legend that explains those codes. For example, if authors chart the participants' grade levels in the identified studies, they could use designations like PreK, K, 1–12, or E, M, HS, or C. Even with such abbreviations, the summary chart for a systematic review can be quite extensive, covering many pages. When the length of these charts prove problematic, authors can make these available as supplemental documents or data that are shared through the journal's online repository.

Those basics should also include source details (e.g., authors, year of publication or access) for all works analyzed. The full reference for those sources should be provided either within the main reference list with special denotation, or in a separate reference list that is solely for the documents in the final pool. This latter option is typically pursued when the number of documents within the final pool is extensive and authors are unable to cite all of them in the body of the article. Further, if the documents in the final pool include both traditional and gray literature, authors may wish to indicate those classes (e.g., T = traditional; G = gray) along with source type. Here again, a code could be used to chart such information concisely (e.g., J = journal, GR = government report, D = dissertation, or CP = conference proceeding).

### *Charting Other Relevant Characteristics*

Beyond the aforementioned source information, other data pertinent to the critical questions posed should appear in the summary chart. This is an essential step in the process, since any conclusions or insights authors reach should be grounded in the evidence derived from their analysis of identified documents. When analyzing selected works, for instance, it can be important to capture information on the theories framing the study; the main constructs of interest; the age/grade, gender, ethnicity, and other relevant characteristics of participants; the methodologies and measures researchers employed; and the key findings they report. For example, should authors' critical question concern how certain constructs are conceptualized within the educational research literature, whether select populations are overrepresented or underrepresented with regard to an educational program or intervention, or whether approaches to measurement have a bearing on outcomes reported in the literature, then information directly related to those issues should be part of the summary chart.

As a case in point, Dinsmore, Alexander, and Loughlin (2008) conducted a systematic review in order to explore the theoretical and empirical boundaries between three constructs that had often been used variably in the literature: metacognition, self-regulation, and self-regulated learning. After executing the search procedure, these researchers identified 255 journal articles that met their inclusion criteria. In keeping with their critical question, Dinsmore et al. determined that it was imperative to analyze *if* and *how* those core constructs were defined. Following a procedure used by Murphy and Alexander (2000), Dinsmore et al. (2008) indicated whether constructs in those articles were explicitly defined (E), or implicitly derived (I).

When an implicit definition was indicated, Dinsmore et al. (2008) also coded whether the authors derived the definition from the conceptual content provided (C), from the specific works referenced (R), or based on the measures applied (M). Additionally, they conducted a content analysis of the explicit definitions for those three central terms. In their chart for the 255 articles, which was hyperlinked to the article, Dinsmore et al. not only provided the aforementioned definition codes but also listed the key terms authors used when explicitly defining metacognition, self-regulation, or self-regulation learning. For these researchers, such charted content was crucial if they were to address the critical question they initially sought to investigate.

#### *Considering Potentially Informative Groupings*

As suggested, there are occasions when the documents in the final research pool do not constitute a single cohesive grouping but appear to fall into distinct clusters. In some instances, authors are aware of those groupings from the outset, as when they make the determination to include both traditional and gray literature in search parameters. In other situations, pertinent groupings may be less obvious from the outset and only begin to emerge once the charting of the identified research gets underway. In either case, authors should explore the charted data for these seemingly potential groupings thoroughly to ascertain whether collapsing data across the groups could well mask pattern or trends of significance.

For example, in a recent review conducted on relational reasoning among groups engaged in problem solving, Jablansky and Alexander (2019) gathered articles that included both college students and adults working on problems as part of their chosen profession, and elementary students engaged in problem solving as part of classroom instruction. As the charting of the articles began, the authors found that there was more separating these groups than their ages or grade levels. There were differences in the kind of problems being solved, the nature of the collaborations described, the discourse and reasoning patterns identified, and the outcomes documented. Consequently, the researchers found it important to treat these two groups separately when drawing conclusions and offering implications.

#### **Interpreting and Communicating Challenges**

Once the identified documents have been charted, the database for the systematic review has been generated. Whatever conclusions authors subsequently reach or whatever interpretations they ultimately put forward should have some

grounding in that data they have extracted from the literature. This grounding is vital because it serves as a source of justification for whatever claims are made and diminishes the influence that personal biases may have on the insights garnered and the outcomes reported. I speak of diminishing the likelihood of personal biases because whatever inquiry researchers undertake will be shaped by their beliefs, experiences, and perspectives. For some educational researchers, acknowledging that bias is an essential element in the review process.

Over the years, I have found that this culminating phase of the review process, while less labor-intensive than the prior phases, can prove the most challenging for authors. That is because this is the phase where the art of a systematic review becomes most evident. At this juncture, authors must step back from their data to gain a perspective on precisely what they have learned from their review that speaks directly to their critical questions and to educational theory and practice.

### *Recognizing Meaningful Outcomes*

As I will try to demonstrate, the process of gaining a perspective on what was learned from a systematic review can occur at multiple grain sizes—from the construct or variable level to overarching trends. Regardless of grain size, however, authors cannot derive meaningful conclusions from their systematic reviews only by engaging in simple arithmetic. In effect, the process of reaching conclusions from these reviews is much more than merely tallying the number of elements that appear in a summary chart. Regrettably, treating data in the summary chart as an arithmetical exercise is a too common occurrence and a barrier to achieving quality in systematic reviews. Certainly, as part of the conclusions, reference will be made to the frequency of particular features or elements that authors charted. However, such numbers are not the conclusion per se but the prelude to meaningful outcomes.

In fact, in a systematic review, it is necessary to delve deeply into the data produced. This in-depth analysis allows authors not only to chronicle what is present or absent within the literature but also to unveil what discrepancies or ambiguities seem to exist. In essence, it is through such probing that the truly meaningful outcomes of the literature review are revealed.

### *Capturing Significant Patterns and Trends*

Even at the construct or variable level, authors should consider how study elements operate interactively and not merely in isolation. In effect, much of the power in conclusions comes from grasping how the constructs and variables function interactively in a way that better informs the critical questions guiding the review. In the Alexander and Judy (1988) review, for instance, it was the attempt to probe the literature to determine *if* and *how* the constructs of domain-specific strategies and knowledge were interrelated that afforded the greatest insights. Similarly, Singer and Alexander (2017) were able to bring clarity to the inconsistent findings regarding comprehension effects when reading in print and digitally by noting the link between the length of texts read and the increasing benefits of print. Further, in their recently completed review, Jablansky and Alexander (2019) discerned that the relational reasoning patterns of K–12 students engaged in group problem solving were distinctly different from the reasoning patterns of groups

working on problems specific to their professional domains. These distinctive patterns appeared to be related to a number of factors, including group members' differential knowledge and investment in the problem and the significance they placed on the outcome of their shared problem solving.

It should be noted that key findings do not always pertain to the interrelations among constructs or variables at any one time but rather to shifts or developments that seemingly transpire across time. For example, how particular constructs have been differentially conceptualized or measured; what reformations in teacher preparation or induction programs have been introduced; how specific cognitive, socioemotional, or neurological conditions are identified and educationally addressed; or what significant curricular developments have occurred within academic domains are the kinds of critical questions that may call for developmental or historical analysis. By viewing their data through a developmental lens, authors may be better able to recognize contextual forces that help explain their findings.

For instance, consider how national and state policies regarding identified student populations have altered the characterization of and attention to those populations within the educational system, or how advancements in cognitive, neurological, or socioemotional assessments have expanded what is understood about those students. In such instances, the failure to examine the resulting database without positioning the constructs or variables in some temporal or contextual framework could well obscure crucial patterns that warrant discussion and explication.

### *Communicating Contributions*

The closing set of challenges I would like to address have to do with communicating the outcome of the systematic review to those with perhaps less in-depth knowledge of the content but likely with some interest in or curiosity about the topic. Although this discussion will center on capturing the implications for theory and practice, I cannot overlook the relevant concerns about the quality of the written document as a whole. As an editor and reviewer for various educational journals, it is sadly too common for me to receive a manuscript that is so grammatically and structurally flawed that its message cannot be deciphered or so prosaic that the power of its message is lost. Quality of the writing and careful preparation of the document cannot be underestimated for any research, including for systematic reviews. Thus, it is imperative for authors to familiarize themselves with the guidelines prepared for their target journal, including any word limits or unique features that are requested. They should be certain to be explicit when defining all key terminology, transparent in the procedures employed, and clear in justifying choices made. I have personally found it important to title manuscripts well maintaining the focus while capturing readers' attention. Moreover, abstracts should be given due consideration, since both title and abstract are often the means by which specific works are discoverable in literature searches.

### *Acknowledging Delimitations and Limitations*

Before authors begin recounting the insights garnered from their systematic review—insights that speak directly to the guiding questions they initially

posed—they should first revisit the study’s delimitations and any limitations that arose. For example, although certain delimitations were deemed necessary from the outset of the review process, it could be that the patterns revealed through analysis suggest that those intentional constraints may have exerted unforeseen effects. Likewise, there may have been unanticipated issues (i.e., limitations) identified during the course of the review that should be acknowledged. A sign of a quality systematic review, as with any good piece of empiricism, is that authors are forthcoming in discussing any decision or event that may have adversely affected the review process and, thus, the findings reported. While acknowledging possible shortcomings, authors should take care not to overshadow the contributions of the systematic review they have completed.

### *Returning to Overall Purpose and Critical Question*

Once any pertinent delimitations and limitations have been duly noted, the task for authors is to address their critical questions and to consider the specific outcomes, data patterns, and general trends in relation to those questions. That consideration begins by returning to the rationale the authors forwarded for taking this deep dive into the literature in the first place and the contributions they expected it to make to educational research and practice. Did this effort of interrogating the literature in such a deliberate and organized manner produce the answers or insights that the authors sought? If so, what specific answers to their critical question were derived from this scholarly exercise? What insights about the topic or issue were garnered?

In declaring what they have unearthed through the systematic review process, authors must be certain to explain those findings in relation to theoretical context that was articulated at the outset. As such, authors must take care to convey what has been learned within the theoretical or contextual boundaries they initially established. They must avoid any impulse to overstep those boundaries by voicing claims that cannot be substantiated by the data they have gathered. This is not to imply that there is no place for speculation within a systematic review. To the contrary, it is in this concluding section of the review that authors are urged to make reasoned speculations regarding research and educational policy and practice.

Of course, while there is a place for venturing beyond the data in a systematic review, such venturing must be expressly labeled as speculative, reasoned and reasonable, and informative and constructive to researchers, practitioners, and policymakers who are the intended recipients of the authors’ message. In effect, authors should signal that what they are about to proffer readers are implications *suggested* by the data, which means these implications are consistent with the data, but by no means unassailable or incontrovertible. With this caveat in mind, authors should advance suggestions to researchers about lingering and unresolved issues from their systematic review that could inform the design of subsequent studies about the focal topic. These suggestions could become the foundation for future critical questions posed by others reviewing the literature. Authors should also use this occasion to venture recommendations to educational practitioners and policymakers as to constructs, factors, or conditions that merit their attention. Even though systematic reviews do not resolve into causal claims, the patterns



and trends they produce can, nonetheless, be highly persuasive evidence that educational researchers, practitioners, and policymakers should not ignore.

### Concluding Thoughts

I could not agree more with Murphy et al. (2017) that reviews allow us to peer deeper into the edifice of education research. Moreover, when these reviews are undertaken systematically to address a question of critical importance to education, and when they are conducted rigorously, reported in a transparent manner, and communicated in an illuminating but justified way, they truly have “the potential to shape the future of research and practice” (Murphy et al., 2017, p. 4).

### Notes

<sup>1</sup> Alexander and Judy (1988) is freely accessible through <https://journals.sagepub.com/stoken/default+domain/2G2T4I6S8NKGXBPAEEI3/full>.

<sup>2</sup>Singer and Alexander (2017) is freely accessible through <https://journals.sagepub.com/stoken/default+domain/JXBRY4AJGUAWZXEPC/PC/full>.

### References

- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*, 375–404. doi:10.3102/00346543058004375
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage.
- Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*, 1297–1326. Retrieved from <https://www.ischool.utexas.edu/~adillon/Journals/Reading.htm>
- Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, *20*, 391–409.
- Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review*, *25*, 391–427. doi:10.1007/s10648-013-9224-4
- Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, *106*, 101–121. doi:10.1016/j.infsof.2018.12.003
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. doi:10.3102/0013189X005010003
- Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, *28*(1), 23–60. doi:10.1007/s10648-014-9294-y
- Hart, C. (2018). *Doing a literature review: Releasing the social science research imagination*. London, England: Sage.
- Hattie, J. A., & Hansford, B. C. (1984). Meta-analysis: A reflection on problems. *Australian Journal of Psychology*, *36*, 239–254. doi:10.1080/00049538408255094
- Jablansky, S., & Alexander, P. A. (2019). *Positioning relational reasoning within group problem solving*. Manuscript submitted for publication.

- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Annals of Internal Medicine*, *51*, 264–269. doi:10.7326/0003-4819-151-4-200908180-00135
- Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology*, *25*, 3–53. doi:10.1006/ceps.1999.1019
- Murphy, P. K., Knight, S. L., & Dowd, A. C. (2017). Familiar paths and new directions: Inaugural call for manuscripts. *Review of Educational Research*, *87*, 3–6. doi:10.3102/0034654317691764
- Okoli, C., & Schabram, K. (2010). *A guide to conducting a systematic literature review of information systems research*. Retrieved from <https://pdfs.semanticscholar.org/31dc/753345d5230e421ea817dd7dcdd352e87ea2.pdf>
- Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 103–125). New York, NY: Russell Sage Foundation.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. New York, NY: John Wiley.
- Shaw, R. L. (2010). Conducting literature reviews. In M. Forrester (Ed.), *Doing qualitative research in psychology: A practical guide* (pp. 39–55). London, England: Sage.
- Singer, L. M., & Alexander, P. A. (2017). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research*, *87*, 1007–1041. doi:10.3102/0034654317722961

### Author

PATRICIA A. ALEXANDER is the Jean Mullan Professor of Literacy and Distinguished Scholar/Teacher in the Department of Human Development and Quantitative Methodology at the University of Maryland, College Park, MD, USA; email: [palexand@umd.edu](mailto:palexand@umd.edu). She is currently a member of the *Review of Educational Research* editorial board and conducts research related to text-based learning, academic development, relational reasoning, strategic processing, and learner interest.