Abstract

Background In fields such as medicine, psychology, and education, systematic reviews of the literature critically appraise and summarize research to inform policy and practice. We argue that now is an appropriate time in the development of the field of engineering education to both support systematic reviews and benefit from them. More reviews of prior work conducted more systematically would help advance the field by lowering the barrier for both researchers and practitioners to access the literature, enabling more objective critique of past efforts, identifying gaps, and proposing new directions for research.

Purpose The purpose of this article is to introduce the methodology of systematic reviews to the field of engineering education and to adapt existing resources on systematic reviews to engineering education and other developing interdisciplinary fields.

Scope/Method This article is primarily a narrative review of the literature on conducting systematic reviews. Methods are adapted to engineering education and similar developing interdisciplinary fields. To offer concrete, pertinent examples, we also conducted a systematic review of systematic review articles published on engineering education topics since 1990. Fourteen exemplars are presented in this article and used to illustrate systematic review procedures.

Conclusions Systematic reviews can benefit the field of engineering education by synthesizing prior work, by better informing practice, and by identifying important new directions for research. Engineering education researchers should consider including systematic reviews in their repertoire of methodologies.

Keywords methods; review; systematic review

Introduction

Scholarly contributions must be situated in the context of prior work; however, the number of articles published each year is increasing rapidly, with no signs of abating. Individual researchers cannot read all of the literature required to stay current. In response, diverse fields are developing systematic review approaches to synthesizing primary studies as a research methodology in and of itself. For example, in fields such as medicine, psychology, and education, systematic reviews of the literature are conducted to critically appraise, summarize, and attempt to “reconcile the evidence in order to inform policy and practice” (Petticrew & Roberts, 2006, p. 15). Systematic reviews make important contributions to the evidence base.
of a discipline by providing synthesized reviews on important issues or topics (Gough, Oliver, & Thomas, 2012). Reviews can demonstrate gaps in recent work or highlight areas where a concept is accepted as true but little evidence exists to support it (Petticrew & Roberts, 2006). Systematic review purposes and procedures may overlap with narrative reviews, yet they constitute a distinctive approach to synthesizing the literature. Specifically, systematic reviews follow transparent, methodical, and reproducible procedures that might be grouped broadly into two arenas: (1) selecting a collection of appropriate studies that will address the review question from the vast and rapidly increasing knowledge base and (2) extracting trends, patterns, relationships, and the overall picture from the collected studies. While procedures to select a collection of appropriate studies are coalescing, procedures to make meaning from a set of primary studies are still emerging. One set of quantitative meaning-making procedures, i.e., meta-analysis, has emerged to study efficacy of interventions, but newer qualitative and quantitative synthesis procedures for meaning making are still evolving.

Now is an appropriate time in the development of the field of engineering education to both support systematic reviews and benefit from them. Borrego (2007b) has applied Fensham’s (2004) criteria for evaluating development of science education research disciplines to engineering education and found dramatic improvements are still needed to develop its foundation. Specifically, the improvements needed in engineering education include:

- Progression: evidence that researchers are informed by previous studies and build upon or deepen understanding
- Model publications: publications that other researchers hold up as models of conduct and presentation of research studies in the field
- Seminal publications: publications recognized as important or definitive because they marked new directions or provided new insights

Systematic reviews have strong potential to become model and seminal publications on which engineering education research can support progress(ion) by uncovering patterns, connections, relationships, and trends across multiple studies. For example, Johnson, Johnson, and Smith’s (1991, 1998) systematic reviews of active and collaborative learning are highly cited, seminal publications providing the foundation for a variety of active learning interventions within and beyond engineering education.

Further, engineering education is an interdisciplinary field, drawing on multiple – often more mature – fields (e.g., cognitive sciences, physical sciences, education, life sciences, organizational development and change, traditional engineering disciplines) for its content, theoretical frameworks, research methodologies, and key findings (Borrego, Streveler, Miller, & Smith, 2008; Jesiek, Newswander, & Borrego, 2009; Johri & Olds, 2011; Lucena, Downey, Jesiek, & Elber, 2008). Identifying, codifying, and synthesizing across these multiple fields will provide crucial foundations for advancement and reduce the likelihood of “reinventing the wheel.” Although there has been work on engineering education for over 100 years (Lohmann & Froyd, 2010), research in engineering education is in its infancy. In this emerging state (Haghighi, 2005; Johri & Olds, 2011), providing wide-ranging, transparent, repeatable reviews would help provide key underpinnings and help focus on productive directions likely to accelerate development of the field. Finally, a systematic review of systematic reviews in engineering education (described briefly later in this article) demonstrates growing interest in this type of synthesis among engineering education researchers. Providing clearer criteria and
standards for conducting and evaluating systematic reviews would improve potential contributions. In sum, more reviews of existing work across multiple fields, conducted more systematically, will help advance the field of engineering education by lowering barriers for both researchers and practitioners to access relevant findings, by enabling more objective critique of past efforts, by identifying gaps, and by proposing fruitful directions for research.

**Purpose**

This article describes systematic reviews as a methodology to promote development of the field of engineering education and adapts existing resources on systematic reviews to engineering education and other developing interdisciplinary fields. Several guides to conducting systematic reviews already exist (e.g., Gough et al., 2012; Petticrew & Roberts, 2006). However, these guides were written for other disciplines with greater degrees of scholarly maturity and greater breadth and depth of research — and often for different purposes (e.g., efficacy of interventions). Our principal contribution lies in adapting these procedures specifically to engineering education through collaboration between experienced engineering education researchers and a librarian whose area of scholarly expertise is conducting systematic reviews. While we do not explicitly address other developing fields that may be at a similar stage, we anticipate this article can be a potential resource to similar fields that rely on more established disciplines for theory and research methods. Methodologically, this article contributes a primarily narrative review of the literature on conducting systematic reviews, specifically for engineering education.

In the following sections, we describe the roles and procedures of systematic reviews in other disciplines and compare the primary purposes and types of reviews. Then we describe the steps in conducting and reporting the results of a systematic review, including establishing validity and reliability. We followed systematic review procedures to identify high-quality systematic reviews of topics related to engineering education. These 14 systematic reviews serve as concrete examples for applying the various steps of systematic review methods in engineering education. (It is beyond the scope of this article to include complete findings and synthesis from our systematic review; rather, we hope the examples will offer specific models for engineering education researchers.)

**Systematic Reviews in Other Disciplines**

Over the past decade, the number of systematic reviews published in medicine, psychology, and education has dramatically increased. A search of Scopus, a comprehensive abstract and citation database of peer-reviewed literature, demonstrated that the most prolific of the disciplines is medicine, with 8,243 reviews published in 2012 alone. Systematic reviews in medicine have become highly valued types of evidence for clinical practices and public policies. The Cochrane Collaboration, an international, nonprofit organization, is leading efforts to inform clinical practices and public policy in medicine by publishing systematic reviews and meta-analyses. Other disciplines also produce reviews; the trends in engineering, psychology, education, and engineering articles with education topics are shown in Figure 1. Although not reaching the scale of health-related fields, reviews in education and psychology are being published at rapidly increasing rates.

In education, three organizations have made strides in creating reviews and developing methods. The Norwegian-based Campbell Collaboration is an international, nonprofit organization funded in 1999 as a related organization of the Cochrane Collaboration. The Campbell Collaboration conducts reviews in education, criminal justice, and social welfare
(Campbell Collaboration, n.d.). They have adapted systematic review methods to accommodate broad disciplinary perspectives, especially those considering social justice and disparities. Another international organization, the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), was founded in 2000 by the Department for Education and Skills, part of the Social Science Research Unit at the Institute of Education, University of London (Andrews, 2005). The EPPI-Centre provides methods and tools for conducting systematic reviews and collects research on education with a global scope in the Database of Education Research. In 2002, the United States Education Sciences Reform Act established the What Works Clearinghouse (Institute of Education Sciences, 2012), which concentrates on effectiveness of education interventions by appraising and synthesizing quantitative studies.

Perhaps the most important consideration in selecting models for systematic reviews in engineering education is the availability of quantitative experimental studies from which to draw for reviews. The more established fields of psychology, education, and medicine have more quantitative experimental studies than are generally available for a given engineering education topic; hence these fields have longer histories of systematic reviews and meta-analyses. As a result, many of the resources describing how to conduct systematic reviews...
tend to privilege positivist and quantitative approaches (e.g., U.S. Department of Education, 2007). In contrast, developing fields, which rely more heavily on qualitative research, are more accommodating of a variety of study designs in applying the logic of systematic reviews (Gough et al., 2012). For example, public health researchers in the United Kingdom have adapted systematic review approaches from clinical medicine to include quasi-experimental designs and qualitative studies while still emphasizing randomized control trials when available (National Institute for Health and Clinical Excellence [NICE], 2009). Public health approaches to systematic reviews may be particularly relevant to engineering education.

The systematic review literature in public health also identified a number of limitations to systematic reviews that overemphasize tightly controlled studies. Slavin (1984) criticized meta-analysis methods that rely strictly on randomized control trials for focusing too much on research design and not enough on the quality of the study; he noted that well-conducted quasi-experiments may yield more useful information than poorly executed randomized control trials. Publication bias favoring significant results tends to overinflate the effects of an intervention (Petticrew & Roberts, 2002). Petticrew and Roberts (2002) noted that emphasis on efficacy of interventions does not take into account the complexity of interactions around public health issues. Public health often deals with cost effectiveness and the gap between efficacy and adoption of effective behaviors, and similar issues are receiving increased attention in engineering education in terms of adoption of research-based instructional strategies (e.g., Borrego, Froyd, & Hall, 2010; Montfort, Brown, & Pegg, 2012). Thus, for many reasons a public health approach to systematic reviews might be most informative to engineering education.

A second important difference between engineering education and psychology, medicine, and education is the interdisciplinary nature of engineering education. As a newer field, engineering education has a greater tendency to draw its literature and theory base from related fields, including more established cognitive psychology and industrial and organizational psychology. Particularly if the scope is not limited to studies of engineers or students majoring in engineering, the majority of sources in a systematic review undertaken by an engineering education researcher may come from other sources than engineering education journals and conferences. The availability of experimental studies and the norms of the field(s) from which sources are drawn are important considerations in the inclusion criteria and analysis approaches selected for a systematic review.

Later in this article we will compare and contrast systematic reviews from other types of reviews. First, we provide stronger motivation for conducting reviews and describe the steps in conducting a systematic review to help readers understand the nature of systematic reviews. (Readers seeking an overview may skim the headings and tables or skip ahead.)

**Purposes and Goals of Reviews**

Researchers may have several different goals in mind when undertaking a review, including “to enable the researcher both to map and to assess the existing intellectual territory, and to specify a research question to develop the existing body of knowledge further” (Tranfield, Denyer, & Smart, 2003, p. 207); “to evaluate the competing theoretical approaches to conceptual representations” (Kiefer & Pulvermüller, 2012, p. 805); and “to identify and describe the extent to which theory or theoretical frameworks informed the development and evaluation of decision support technologies” (Durand, Stiel, Boivin, & Elwyn, 2008, p. 125).

Guides on systematic studies offer multiple purposes. For example, Gough et al. (2012) succinctly stated reasons that reviews of prior work are needed:
1. Any individual research study may be fallible, either by chance, or because of how it was designed and conducted or reported.

2. Any individual study may have limited relevance because of its scope and context.

3. A review provides a more comprehensive and stronger picture based on many studies and settings rather than a single study.

4. The task of keeping abreast of all previous and new research is usually too large for an individual.

5. Findings from a review provide a context for interpreting the results of a new primary study.

6. Undertaking new primary studies without being informed about previous research may result in unnecessary, inappropriate, irrelevant, or unethical research. (p. 3)

Baumeister and Leary (1997) included the following purposes for reviews: evaluating intervention efficacy, tracing historical development, describing state of knowledge or practice on a topic, developing or evaluating theory, and identifying opportunities for future research and innovation. Reviewers may seek to address multiple purposes in the same review. Since justification for intervention efficacy is well understood in the engineering education community, we briefly describe other purposes and offer examples in engineering education of which we are aware.

**Trace Historical Development**

Historical reviews trace development of a field or research area chronologically, often with an “ongoing commentary regarding the impact and shortcomings of various contributions” (Baumeister & Leary, 1997, p. 312). Historical development reviews are not particularly common. Since engineering education research has a relatively short history, initial historical development reviews usually encompass the entire timeline of engineering education research. Wankat, Felder, Smith, and Oreovicz (2002) and Jesiek et al. (2009) included historical development of engineering education in the literature review in their publications (neither of which were systematic reviews).

**Describe State of Knowledge or Practice on a Topic**

Many of the earliest systematic reviews were intended as guides for practitioners to synthesize what was currently known about practice in a specific field. Many reviews in engineering education to date have been intended for this purpose (e.g., Froyd & Ohland, 2005; Litzinger, Lattuca, Hadgraft, & Newstetter, 2011; Prince, 2004; Prince & Felder, 2006; Streveler, Litzinger, Miller, & Steif, 2008).

**Evaluate or Develop Theory**

Engineering education research has been criticized for lack of connections between related studies, and many scholars have called for deeper engagement with theory to promote cumulative growth of understanding (Beddoes & Borrego, 2011; Koro-Ljungberg & Douglas, 2008). Engineering education has few, if any, of its own theories; researchers typically borrow theories (as well as methods) from many other fields including but not limited to psychology, education, sociology, and women’s studies. Systematic reviews can support both evaluating existing theories and developing new ones, as described further in the following subsections.
**Evaluate theory**  Engineering educators are often skeptical that research results or theories developed by other fields are relevant to engineering education settings (Borrego, 2007a; National Research Council, 2011; Wankat et al., 2002). At least two potential approaches may improve relevance. One is to replicate studies with engineering education participants. The other is through scholarly synthesis and translation procedures such as those developed for systematic reviews. Systematic reviewers in engineering education might, for example, combine primary studies conducted with engineering and nonengineering students to consider whether a particular motivation theory applies to engineering undergraduates and what additional evidence is needed for the theory’s relevance to engineering students. We did not identify any engineering education reviews that took the approach of evaluating theory.

**Develop theory**  The ability to develop and evaluate theoretical frameworks depends, in part, on the extent to which authors apply theoretical frameworks that have been constructed and how their applications inform refinement of the frameworks. Construction of a framework requires assembly of published literature into patterns; in turn, framework construction requires knowledge of appropriate literature across multiple disciplines to include published partial patterns. After initial framework development, the refinement and improvement of the framework are informed by applications. Strengths of the framework can be established, and opportunities for improvement can be identified. Applications may show where inclusion of more factors, more elaborate descriptions of interactions among the factors, or more sophisticated analysis of factor interaction would improve the framework. Framework refinement and improvement are less likely if the framework is never or infrequently applied. Frequency of application of a framework depends on how many researchers know about the framework and existence of models for applying the framework; if frequency of application can be increased, it is likely that theory will develop faster. As a result, theoretical development in an emerging field can be catalyzed if authors are aware of reviews on which they can rely for the best available answers to questions about theoretical frameworks and their applications. Examples of engineering education-related reviews that sought to develop theory include Crismond and Adams (2012), Henderson, Finkelstein, and Beach (2010), and Mehalik and Schunn (2007).

**Identify Opportunities for Future Research or Innovations**  All reviews, regardless of their primary goal, should also identify opportunities for future work. For example, in highlighting gaps in past work, some reviews emphasized areas of opportunity (Beddoes & Borrego, 2011; Koro-Ljungberg & Douglas, 2008). Few engineering education reviews that we identified, though, offered specific implications for future work.

**Steps in Conducting a Systematic Review**  

Now that we have established some of the reasons a researcher may undertake a review, we now describe systematic review procedures in detail.

**Deciding to Do a Systematic Review**  Since the data source for a systematic review is primary studies conducted by others, the quantity, quality, and accessibility of primary studies will determine whether a systematic review can be conducted. The quality of a systematic review relies on consistency and appropriateness between goals, research questions, selection criteria, and synthesis approaches. Developing goals, criteria, and approaches is likely to involve an iterative process to align procedures with the type and number of primary studies available.
Protocols often change throughout the course of a systematic review, particularly during literature searching and appraisal of the primary sources. Petticrew and Roberts (2006) recommended conducting a scoping review to estimate the number and accessibility of sources and decide whether to conduct a systematic review at all. The results of a scoping review should be included in a proposal if one is required (e.g., for funding or dissertation work) before starting a systematic review. Petticrew and Roberts (2006) provided specific situations that could warrant or particularly benefit from systematic reviews:

When there is uncertainty, for example about the effectiveness of a policy or a service, and where there has been some previous research on the issue.

In the early stages of development of a policy, when evidence of likely effects of an intervention is required.

When it is known that there is a wide range of research on a subject but where key questions remain unanswered – such as questions about treatment, prevention, diagnosis, or etiology, or questions about people’s experiences.

When a general overall picture of the evidence in a topic area is needed to direct future research efforts.

When an accurate picture of past research, and past methodological research is required to promote the development of new methodologies. (p. 21)

Identifying Scope and Research Questions

As with other research methods, well-defined questions that accurately reflect the intent of the researchers suggest criteria to guide design decisions later in the process (Petticrew & Roberts, 2006). For example, later in the process, researchers will have to articulate criteria for excluding or including studies in the collection to be analyzed. If the question is phrased as “What is known about X?”, it will be difficult to determine if a particular study should be included or excluded, since virtually every study could potentially shed some light on the question. The question might be refined in several different ways:

Is X effective (under particular conditions)? or What factors influence effectiveness of X?

What learning outcomes are associated with X?

What factors contribute to increasing or decreasing likelihood of X?

How has X been operationalized, measured, or assessed?

What methods have been used to teach X, increase learning of X, or increase application of X?

What theoretical frameworks and perspectives have been applied to X?

What research and evaluation methods have been used to study X?

For this example, X might be a topic or construct of interest, such as global competency, entrepreneurship, thermodynamics, motivation, diversity, teamwork, or instructor beliefs about teaching.

The EPPI-Centre (2010) advised systematic reviewers to spend time conceptualizing the research question so that it will guide all other stages of the review process:
As with any piece of research, defining the research question for a systematic review is the most important stage of the process, as it provides the framework for all the other stages. The question being asked by a review will determine the method of review and the studies that are considered by the review. The question is likely to include implicit assumptions about the topic and thus any underlying conceptual framework (or logic model) that will be used to interpret and understand the research evidence in the review should be made explicit. (p. 4)

More specifically, Petticrew and Roberts (2006) advised that “what works?” review questions are best answered by synthesis of quantitative experimental studies, while “what matters?” review questions are best addressed through synthesis of qualitative studies (2006, p. 57).

To ensure that relevant parameters are defined in a research question and in later stages of analysis, the National Institute for Health and Clinical Excellence offered the PICO framework (Table 1). Salleh, Mendes, and Grundy (2011) used the PICO framework in describing the interventions for their systematic review of pair programming in computer science education.

The National Institute for Health and Clinical Excellence (2009) offered detailed advice on writing different types of questions for effectiveness, correlation, cost effectiveness, and qualitative systematic reviews. Effectiveness (of an intervention) and qualitative reviews are the two most relevant to engineering education. For effectiveness reviews, researchers should consider outcomes, implementation issues, context, and other factors that could influence the outcomes. For qualitative reviews, the Institute advised to avoid overly broad questions, e.g.,

---

**Table 1** PICO framework for Structuring Research Questions and Analysis for Systematic Reviews

<table>
<thead>
<tr>
<th>PICO Framework component</th>
<th>Engineering education examples and considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population: Which populations are we interested in? How can they best be described? Do any subgroups need to be considered?</td>
<td>Engineering students, faculty members or practicing engineers Year in school or years of work experience Specific engineering discipline(s), course(s) Gender, race/ethnicity, disability status Institution type, country</td>
</tr>
<tr>
<td>Intervention: Which intervention, activity or approach should be used?</td>
<td>Careful definition of what is and is not considered with respect to the intervention: duration incentives (e.g., required course or elective) curriculum facilitator training requirements</td>
</tr>
<tr>
<td>Comparison: What is/are the main alternative/s to the intervention being considered?</td>
<td>Careful definition of any control or comparison group designs to include or exclude, e.g., is a control group required, considering control groups are uncommon in engineering education studies? What is the alternative to the intervention, e.g. lecture-only course or no intervention?</td>
</tr>
</tbody>
</table>

The experience of engineering students in design courses? Table 2 includes examples adapted for engineering education.

**Defining Inclusion Criteria**

Once the purpose, research questions, and scope have been clearly defined, the next step is to develop inclusion and exclusion criteria for selecting primary studies. Published systematic review procedures describe at least three types of inclusion criteria. The first type is criteria for selecting databases that will be searched for articles. More detail on selecting databases is provided in the following section. The second type is a set of combinations of search words or phrases and logical connectors (AND, OR ...) to winnow all the articles in a database down to a small enough set that more detailed inclusion criteria can be examined. In our opinion, in their math education systematic review, Li and Ma (2010) were exemplary in clearly defining terms and showing how search terms and research questions related to inclusion criteria. The third type of criteria guides selection of articles that will be analyzed. In all cases the purpose and research questions for the systematic review should guide inclusion criteria.

In summary, authors should determine the inclusion criteria and the process for identifying the articles that will address the purpose of systematic reviews and the research questions. Extensive documentation of the protocol and criteria will allow other researchers to replicate the systematic search process.

We identified two helpful examples of inclusion criteria in systematic reviews related to engineering education. Schmid, Bernard, et al. (2009) selected articles that

- Address the impact of computer technology (including CBI [computer-based instruction], CMC [computer-mediated communication], CAI [computer-assisted instruction], simulations, e-learning) on students’ achievements (academic performance).
- Be conducted in formal post-secondary educational settings (i.e., a course) leading to a certificate, diploma, or degree.
- Represent CI [classroom instruction], blended, or computer lab-based experiments, but not distance education environments. (p. 99)
Whereas, Li and Ma (2010) applied the following criteria:

- The study uses CT [computer technology] for instructional (or learning) purposes.
- Participants of the study are students in regular classrooms in grades K-12.
- The study employs an experimental or quasi-experimental design (as defined later).
- The study is published during 1990 to 2006 (without restriction to geographical area or language).
- The study uses mathematics achievement as outcome.
- The study reports quantitative data in sufficient detail for the calculation of an intervention effect size. (p. 222)

Implementation of inclusion or exclusion criteria is often formalized as part of a protocol, which assists researchers in executing the procedures, and in documentation of systematic review methods. Gough (2004) emphasized the importance of well-documented and systematic procedures. Both the focus of the review and the inclusion criteria should be explicit. Criteria should reflect the focus and assumptions of the research question (EPPI-Centre, 2010). Inclusion criteria should be free of bias; in other words, they should not intentionally or unintentionally exclude undesirable or inconclusive results (Petticrew & Roberts, 2006).

D. A. Cook and West (2012) explain:

Regardless of the actual criteria selected, it is important to clearly define these criteria both conceptually (often by using a formal definition from a dictionary, theory or previous review) and operationally (by using detailed explanations and elaborations that help reviewers recognize the key concepts as reported by authors in published articles). Although some operational definitions will be defined from the outset, many of these may actually emerge during the process of the review as reviewers come across articles of uncertain inclusion status. Such cases should be discussed by the group with the goal not only of deciding on the inclusion or exclusion of that article, but also of defining a rule that will determine the triage of similar articles in the future. Such decisions, along with brief examples of what should and should not be included, can be catalogued in an explanatory document. Although the conceptual definitions should remain unchanged, the explanatory document and the operational definitions it contains often continue to evolve throughout the review process. (p. 947)

Developing and refining systematic review protocols is an iterative process as review team members begin to apply criteria and discuss and reconcile disagreements at each phase. Cook and West (2012) recommended testing the inclusion criteria through a pilot phase and resolving all discrepancies before moving onto later phases.

Certainly the point of systematic searching is to ensure that all relevant sources are identified. The sampling logic (Gall, Gall, & Borg, 2007) is to identify all relevant sources, rather than a subset that is either representative or particularly informative (although a subset may be selected for analysis due to time or other resource constraints). Professional judgment enters into all phases of research, and systematic reviews are no exception. If researchers are aware of valuable sources that do not fit the inclusion criteria, then the criteria should be adjusted or additional search procedures implemented.
Finding and Cataloging Sources

Several databases can be used to locate primary resources, as shown in Table 3. The principal type of resource is subject (or disciplinary) databases such as ERIC, Education Full Text, Compendex, and Inspec. When selecting databases, it is important to include fields other than engineering education that may be publishing relevant research on the topic of interest, for example, psychology, communication, and computer science. Other more general databases such as JSTOR and Scopus should also be considered. Searching a wide variety of databases ensures that relevant studies are located despite the discipline in which they were published and indexed. However, these databases do not cover the often overlooked gray literature, which includes any research not published by commercial publishers such as many conference proceedings, books, book chapters, dissertations, theses, government documents, white papers, and proposals (Thompson, 2001). In engineering education as in other fields, most conference papers are not indexed in databases such as those listed in Table 3; a notable exception is the ASEE/IEEE Frontiers in Education conference. Another useful source of gray literature for engineering education systematic reviews may be evaluation reports (if they can be obtained; evaluation reports are not often posted online) that present original empirical data and conclusions. Reports, such as those published by national academies and other advisory organizations, written to draw attention to an important issue are another source of gray literature commonly cited in engineering education. While committee reports might be an important component of a narrative literature review, they are less relevant to systematic reviews such as those seeking to understand a specific intervention.

When conducting a search, it is essential for systematic reviewers to meet with a librarian or information scientist specializing in their discipline to get expert advice on searching (McGowan & Sampson, 2005). Many librarians are now aware of systematic reviews and can provide assistance beyond the search process. This expertise is most commonly found in medical libraries because of the increasing prevalence of systematic reviews in medicine. Librarians can make recommendations for effective search procedures, sources to search, and

<table>
<thead>
<tr>
<th>Table 3 Databases for Systematic Review Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of database</td>
</tr>
<tr>
<td>Subject specific databases</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>General databases</td>
</tr>
<tr>
<td>Journal databases</td>
</tr>
<tr>
<td>Gray literature databases</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Borrego, Foster, & Froyd
tracking citations. Many academic librarians are tenure-track faculty who may be interested in coauthoring a review, particularly if they contribute substantially to finding and cataloging sources (or at other stages). As with any interdisciplinary collaboration, open communication about benefits and contributions is advised (Borrego & Newswander, 2008).

Other than database searching, three additional search approaches should be considered: contacting experts, citation searching, and hand searching (Papaioannou, Sutton, Carroll, Booth, & Wong, 2010). Contacting experts involves selecting specific experts, professional organizations, and relevant listservs to advertise and request articles matching inclusion criteria. Citation and hand searching are described in the following paragraphs.

Citation searching, or snowball sampling, involves reviewing the lists of works cited by sources already identified (cited references) as well as the later articles that cite the sources already identified (citing references). The National Institute for Health and Clinical Excellence (2009) recommended citation searching in cases when other search steps have not yielded enough useful studies. Many electronic databases such as Scopus and Web of Knowledge allow for easy identification and searching of both cited references and citing references, provided these sources are also indexed by the database. For example, Web of Knowledge includes a feature that can help researchers focus on the most highly cited primary sources. Similarly, previous reviews in the topic or a related area, if they exist, can be used to identify additional sources, criteria, search terms, and databases (NICE, 2009).

Hand searching is reading the table of contents of journals or other potential sources (NICE, 2009), a task that is required by Cochrane Collaboration guidelines. Hand searching has been the most common method of prior systematic reviews in engineering education research. Wankat (1999, 2004), Whitin and Sheppard (2004), and Koro-Ljungberg and Douglas (2008) hand searched issues of this journal, while Beddoes, Jesiek, and Borrego (2010) and Osorio and Osorio (2002) hand searched the contents of multiple engineering education journals. Many of these reviews have attempted to map the field of engineering education research in terms of general research topics or methodological approaches. In addition to pointing out that a single journal does not sufficiently represent an entire field, we posit that engineering education research (and interest in related fields) has both expanded to include literature from many related fields and coalesced around key issues to the extent that more systematic reviews of specific topics are possible and needed. Many of the primary engineering education journals are now indexed by databases listed above and in Table 3. Ulrich’s periodicals directory (Serial Solutions, n.d.) provides information on which specific databases index a particular journal.

After identifying articles using the various search procedures described above and removing duplicates, the number of sources included and excluded at each phase should be tabulated. One option for presenting this information is shown in Figure 2, which was adapted from the PRISMA flowchart (Liberati et al., 2009). Examples of tables presenting this information can be found in Benitti (2012), Salleh et al. (2011), Meese and McMahon (2012), and Monasor, Vizcaíno, Piattini, and Caballero (2010).

Critique and Appraisal

After the primary sources have been selected, identified, and sorted, the next step is to systematically assess the quality of each primary study.

Cook and West (2012) described an intermediate step of abstracting important details from each source. The PICO framework (Table 1) can be used to create a template specific to the review topic that summarizes information on the outcomes, study design, and
conclusions. These summaries should include details of how success or impact was defined and the conclusion drawn by the authors (as relevant to the focus of the review). It is also important to include the concrete details that led to the assessment of quality, such as sources of bias or missing details. As in earlier phases, abstracting should have a clear protocol and specific rules should be created while discrepancies and borderline cases are discussed and resolved by review team members. Lou (2004) and Tamim, Bernard, Borokhovski, Abrami, and Schmid (2011) presented lists of the features they coded in their engineering education-related systematic reviews.

Details gathered during abstracting allow systematic reviewers to assess the quality of each study and its conclusions. Assessing the quality of primary studies means applying criteria for quality research consistent with the methodology selected by the authors. A number of sources described these criteria in terms directly relevant to engineering education research (Borrego, Douglas, & Amelink, 2009; Koro-Ljungberg & Douglas, 2008; Shavelson & Towne, 2002; Streveler, Borrego, & Smith, 2007). In the literature specific to systematic reviews, quality is framed as internal consistency of the study, or its fit, transparency, and appropriateness. In other words, did the researchers clearly describe their procedures and

**Figure 2** Flowchart for articles, adapted from PRISMA (Liberati et al., 2009).
Table 4 Questions to Consider When Constructing Research Questions and Protocols for Effectiveness Reviews

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>How valid and appropriate are the outcome measures used to assess effectiveness (e.g., self-report vs. observations)?</td>
</tr>
<tr>
<td>How does the content of an intervention or program influence effectiveness?</td>
</tr>
<tr>
<td>How does the way an intervention is carried out influence effectiveness?</td>
</tr>
<tr>
<td>Does effectiveness depend on the job title or position of the intervener or other factors such as their age, gender, or ethnicity?</td>
</tr>
<tr>
<td>Does the site or setting influence effectiveness?</td>
</tr>
<tr>
<td>Does the intensity, length, or frequency of the intervention influence its effectiveness or duration of effect?</td>
</tr>
<tr>
<td>How does effectiveness vary according to the age, gender, class, and ethnicity of the target audience? Is there any differential impact on between different population groups?</td>
</tr>
<tr>
<td>Are there any factors that prevent – or support – effective implementation?</td>
</tr>
</tbody>
</table>


decisions, and were these procedures appropriate to the questions and theoretical framework identified (Petticrew & Roberts, 2006)? Additionally, some guides included consideration of fit to the systematic review question as part of the overall quality assessment of the individual studies (EPPI-Centre, 2010; Gough, 2004). Table 4 gives more detail about aspects of quality specific to intervention studies. Salleh et al. (2011) and Tondeur et al. (2012) provided additional detail on the specific quality appraisal procedures used in their systematic reviews.

A major consideration in quality of primary studies is limitations and sources of bias in the studies, particularly those evaluating efficacy of interventions. The existence of publication bias, meaning that positive results are more likely to be published than negative or inconclusive results, is widely acknowledged. Sources of bias within individual studies should also be considered. Participants who self-select to be part of a program, intervention, or survey are likely to be more receptive than any comparison groups. Materials provided to a treatment group may contaminate the comparison group if students share resources, e.g., Web sites. High attrition from programs and studies weakens the validity of the results. Self-reporting may overestimate attitudes and behaviors, particularly when the instruments do not conceal socially desirable responses.

A second major consideration for intervention studies is fidelity of implementation and context. When considering whether a particular intervention works across studies, it is important to understand whether and how different groups are implementing similar interventions. Various aspects of implementation may impact outcomes, and it is important to define the intervention in question. Similarly, contextual characteristics related to the type of institution, location, and socioeconomics of the participants and setting should be considered (What Works Clearinghouse, 2011).

In the systematic review process and report, particularly when drawing conclusions about effectiveness of interventions, authors should formally or informally weight the studies on the basis of application of these quality criteria. When a substantial number of controlled studies are available, reviewers may formalize quality scores in terms of number of explicit criteria met and display this in summary tables using a notation such as +, ++, or ++++. Additionally, Petticrew and Roberts (2006) recommend describing the quality considerations for each source in the text, particularly if research designs vary.
Synthesis

Most of the value of a systematic review derives from a thorough synthesis of the primary sources that have been collected. All of the steps described previously have led to supporting the review synthesis. Cook and West (2012) described the aim of synthesis in systematic reviews:

Synthesis involves pooling and exploring the results to provide a ‘bottom-line’ statement regarding what the evidence supports and what gaps remain in our current understanding. This requires reviewers to organise and interpret the evidence, anticipating and answering readers’ questions about this topic, while simultaneously providing transparency that allows readers to verify the interpretations and arrive at their own conclusions. (p. 950)

Petticrew and Roberts (2006) outlined the basic steps in synthesis:

1. Organize the studies, e.g., by outcome, population, level of analysis, study design
2. Critique within studies
3. Critique across studies

These steps are discussed in the following sections, along with related considerations.

Mapping

The first step, organizing the studies and reporting certain characteristics, can be accomplished by mapping. Mapping, as the name implies, maps the work that has been done in a field or topic area. It uses keywords to describe populations, theoretical frameworks, methods, countries, journals, disciplinary affiliations, and other details as relevant. Deciding which characteristics to report should be guided by the review question. Mapping may be viewed as summarizing the results of the abstracting process described above. Additionally, mapping “produces a useful product in its own right to describe what research has also been undertaken in a field (as defined by the inclusion criteria) and so can inform policies for future research” (Gough, 2004, p. 56). A number of studies have focused on the goal of mapping the body of engineering education research (Jesiek et al., 2011; Wankat, 1999, 2004; Whitin & Sheppard, 2004). Monasar et al. (2010) presented extensive mapping results in the context of a systematic review. Mapping can also help inform decisions about where to focus the remaining analysis (NICE, 2009). While many previously published reviews have stopped at mapping, the systematic review literature does not consider mapping to be a full synthesis of the primary studies without subsequent critique steps.

Critique within studies

The second step focuses on presenting the assessment of quality for each study in turn. Even if detailed tables are included, at least some of the report text should be used to describe each study. The level of detail can range from the amount of text that fits in a table to lengthy summaries. The level of description should be balanced for space considerations (Petticrew & Roberts, 2006). The EPPI-Centre (2010) advises:

It is valuable to be explicit about how studies are singled out for description in a review and to be systematic when presenting detail of different studies so that each study is given standard treatment at write-up. It is even more valuable if the rationale for presenting certain studies and their results includes a measure of the quality and relevance of the studies producing those results. . . . The aim is to make the links between the detail of the studies found and the reviewers’ conclusions clear. (pp. 14–15)
This extended textual description of primary studies is important because it highlights differences in study quality and discusses how sources of bias may have influenced the study results (Petticrew & Roberts, 2006).

**Tables** Publications of systematic reviews typically include one or more tables to summarize populations, methods, and results of the sources reviewed. Petticrew and Roberts (2006) explained:

Clear, detailed tables increase the transparency of the review. They show the reader which data have been extracted from which studies, and they clarify the contribution made by each study to the overall synthesis. Tables can also reveal something of the review’s wider context; that is, they can be used not just to present details of each study’s methods and findings, but to organize and present information on the context, setting, and population within which each study was conducted. In short, clear, well-organized tables show how each study’s data contribute to the reviewer’s final conclusions, and they allow the reader to assess whether they would have come to similar conclusions, based on the same set of data. (p. 165)

To present important characteristics of the studies as well as the critical appraisal of their quality, multiple tables may be needed. Sources should be listed in tables according to some sort of logical organization; alphabetical listing by first author is often used to help readers navigate between accompanying text and tables. One example is Table 6 in this article. Other strong examples to which we point readers are Becker and Park (2011), Benitti (2012), Falchikov and Goldfinch (2000), Koro-Ljungberg and Douglas (2008), and Salleh et al. (2011). Since the review is likely to include a substantial number of studies, they should be categorized into groups based on methods or other salient concepts. Some examples of meta-analysis summary tables, which highlight potential factors, are presented in Falchikov and Goldfinch (2000) and Li and Ma (2010). The results of meta-analysis are also often presented in plots, as in Schmid et al. (2009) and Springer, Stanne, and Donovan (1999). Some examples of qualitative meta-ethnographies that summarize themes in tables are Meese and McMahon (2012) and Tondeur et al. (2012). As in earlier stages of the systematic review, it is important to define and communicate the criteria for grouping primary studies to avoid the appearance of bias, e.g., to support a particular belief (Petticrew & Roberts, 2006).

**Critique across studies** The third step is the heart of synthesis and the major contribution of systematic reviews. D. A. Cook and West (2012) asserted that “The most informative aspect of many reviews is not the average result across studies, but, rather, the exploration of why results differ from study to study. An explanation of between study inconsistency should be part of all systematic reviews” (p. 951).

A number of specific synthesis methods are available, and choice of method is informed by the nature of primary studies and goal of the review. For example, Gough et al. (2012) listed meta-analysis and other quantitative methods, thematic summaries, framework synthesis, thematic synthesis, meta-ethnography, and mixed methods synthesis as separate methods. Although most systematic review guides give equal coverage to quantitative and qualitative synthesis methods, in this article we focus on methods that do not require a large number of primary studies for statistical analysis (to test hypotheses or efficacy). We selected this focus because we believe that, in most research areas related to engineering students or engineering content, there are simply not enough studies of similar design to support meta-analysis. This is not to say that meta-analysis cannot inform engineering education research; rather, if such
a study were conducted at this point, it would draw heavily from psychology or education sources and closely follow the procedures previously developed and extensively documented for those fields (Institute of Education Sciences, 2012; B. Mullen, Driskell, & Salas, 1998; Petticrew & Roberts, 2006).

In terms of qualitative methods for synthesizing primary studies, Noblit and Hare (1988) are widely credited with first describing qualitative synthesis as meta-ethnography. More recently developed approaches such as thematic synthesis borrow from meta-ethnography (Thomas & Harden, 2008). Barnett-Page and Thomas (2009) critically reviewed nine different methods for qualitative synthesis of primary studies. Thomas and Harden (2008) present an example of thematic analysis while placing the approach in the context of similar qualitative synthesis techniques. Readers are directed to these sources for more detailed information on the techniques, while we illustrate some of the principles using content analysis approaches.

**Qualitative content analysis** Particularly for syntheses of qualitative studies, another source of guidance is the literature on content analysis, which is often presented as a type of qualitative research in educational research texts. Content analysis is a method for synthesizing meaning from written documents, transcripts, or other media (including audio/visual). The basic procedure is to create mutually exclusive coding categories, categorize the data into these categories, and report frequencies. If the sample size is large enough, descriptive statistics and simple relationships are also reported (Downe-Wamboldt, 1992; Gall et al., 2007). Using a previously developed coding scheme (e.g., from a similar prior study) saves content analysts time in developing a comprehensive set of categories (Gall et al., 2007).

Content analysis is flexible and open-ended; the researcher makes some decisions based on the research question and nature of the data. As with the systematic review methods described in previous sections, content analysts should have a clear purpose and rationale in mind to guide all decisions to adapt the methodology (Leedy & Ormrod, 2005). Leedy and Ormrod (2005) suggested using at least two analysts and reporting interrater reliability (Gall et al., 2007). Using a previously developed coding scheme (e.g., from a similar prior study) saves content analysts time in developing a comprehensive set of categories (Gall et al., 2007).

Content analysis is flexible and open-ended; the researcher makes some decisions based on the research question and nature of the data. As with the systematic review methods described in previous sections, content analysts should have a clear purpose and rationale in mind to guide all decisions to adapt the methodology (Leedy & Ormrod, 2005). Leedy and Ormrod (2005) suggested using at least two analysts and reporting interrater reliability (Gall et al., 2007). Using a previously developed coding scheme (e.g., from a similar prior study) saves content analysts time in developing a comprehensive set of categories (Gall et al., 2007).

Graneheim and Lundman (2004) explained that early content analysis publications, beginning in the 1950s, described strictly quantitative approaches striving for objectivity; more recently, content analysis has expanded to accommodate more interpretive, qualitative perspectives. Petticrew and Roberts (2006) cautioned that “Perhaps the least useful way of dealing with qualitative data in systematic reviews is to turn it into quantitative data” (p. 191). Content analysts differentiate between the visible, obvious manifest content and the underlying meaning, or latent content (Downe-Wamboldt, 1992; Graneheim & Lundman, 2004; Kondracki, Wellman, & Amundson, 2002). When embarking on a new project, the researcher makes a conscious decision about stance and focus with respect to these issues (Graneheim & Lundman, 2004). Educational research textbooks tend to emphasize “objective” (Leedy & Ormrod, 2005, p. 152) study of manifest content in which “text is assumed to be invariant across readers and across time” (Gall et al., 2007, p. 292). Nursing and nutrition education, in contrast, consider latent meaning and broader context (Downe-Wamboldt, 1992; Graneheim & Lundman, 2004; Kondracki et al., 2002). Finally, the most cutting-edge discussion about creating quantitative data from qualitative texts is in mixed methods (e.g., see the Journal of Mixed Methods Research). Again, researchers make decisions...
based on goals and the quality and quantity of primary sources, and these choices should be documented with rationale in the review report.

Finally, relevant theory should be considered in content analysis (Gall et al., 2007), although it may be used in a different way than in primary studies, which may be guided by one theoretical perspective. Content analysts may not want to limit their systematic review to a particular theoretical framework because the aim of the review may be either to compare different theoretical perspectives on the same problem or develop a framework to describe seemingly disparate approaches. Nonetheless, researchers should articulate their stance on the role of theory in a particular systematic review.

Limitations, Validity, and Reliability Concerns

As with any research study, systematic reviews have limitations that arise from both the quality and quantity of the primary studies and the quality of the systematic review procedures (D. J. Cook, Mulrow, & Haynes, 1997).

Most threats to validity and reliability arise from bias. Any situation that prevents a reviewer from discovering primary studies creates bias. Systematic reviewers should consider whether bias influenced their selection of primary studies due to any of the following characteristics of primary studies: strong positive (or negative) results, multiple related publications on the same or related studies, whether they were published in well-indexed journals or conference proceedings, and language of publication (Higgins & Green, 2009). For example, studies identifying statistically significant differences might be more likely to be published than studies finding no statistically significant differences. Similarly, multiple primary sources describing similar studies could skew the results toward one set of original data. Restricting the search to sources published in one language would exclude certain countries and regions. These biases can be reduced by searching thoroughly and including gray literature.

Another type of bias is selective outcome reporting, when primary study authors decide to report only certain outcomes. It is usually difficult for systematic reviewers to guess which results were not reported, but they can note selective outcome reporting as a potential limitation, particularly when primary authors describe their analyses and part of a larger study. Last, reviewers could unintentionally employ selection bias when applying their inclusion and exclusion criteria to primary studies. To avoid selection bias, during the quality appraisal phase, primary studies could be masked for author names, affiliations, and the journal name. Although research has not shown masking to make a significant difference in results of reviews, it does eliminate one source of bias and lessen the overall perception of bias (Berlin, 1997; Jadad et al., 1996).

Publication bias in quantitative studies can be detected using a funnel plot, or a “sample scatter plot of the intervention studies against some measure of each study’s size or precision” (Higgins & Green, 2009, p. 310). Funnel plots were first used in social science research by plotting effect estimates against total sample size (Light & Pillemer, 1984). Each primary study would be plotted as a single point. An asymmetrical funnel plot of a particular research area would indicate a publication bias from not reporting both positive and negative outcomes.

The quality of a systematic review is determined primarily by consistency and transparency in selecting and reporting procedures for every step of the review. Readers of the review should have enough information to judge the validity of the search procedures, the identified studies, and the conclusions drawn from them. The reliability of the review can be established primarily through a collaborative process in which multiple team members apply and discuss criteria for inclusion or exclusion and quality assessment.
These limitations of systematic reviews cannot be eliminated, but they can be reduced by employing some of the strategies described in this section, defining and following clear procedures, and reporting decisions and procedures in detail (P. D. Mullen & Ramirez, 2006).

**Advice for Writing the Review**

In many cases, the final step of conducting a systematic review is disseminating results through a report, conference paper, dissertation or thesis, or article. Specifically, Leedy and Ormrod (2005) offered a simple checklist of what to include in a report or article describing a content analysis:

- A description of the body of material studied
- Precise definitions of the characteristics of focus
- The coding or rating procedure
- Tabulations of each characteristic
- A description of patterns that the data reflect (p. 143)

As reinforced in several sections above, procedures, criteria, and evidence should be presented in enough detail to allow readers to assess the quality of the review and to judge for themselves whether conclusions were warranted. Key decisions related to procedures should include the rationale for the review. Important details of the primary studies should be presented and cited in discussion and conclusion sections.

Several of the systematic reviews we identified described their procedures so clearly and thoroughly that the reviews could be replicated by others. Benitti (2012) presented search protocols in both text and tables. Salleh et al. (2011) provided extensive detail on quality appraisal in particular. Tamim et al. (2011) provided the variables they coded (extracted) in their meta-analysis of prior meta-analyses in an appendix.

Finally, a systematic review should conclude with substantive directions for future research and implications for practice (NICE, 2009). The primary conclusions should not be obvious or already known at the outset of the review, e.g., “More research is needed in this important area.” Interpretation should also include clear directions for future research (Gall et al., 2007). Future research directions should summarize, synthesize, and critique gaps in the present body of research and prioritize which directions, topics, or approaches would be most productive. Systematic reviews we identified with relatively strong future work sections include Falchikov and Goldfinch (2000), Lou (2004), and Springer et al. (1999). Implications for practice should be related to the conclusions that can be (cautiously) drawn based on the evidence collected regarding the efficacy of the intervention and its influencing factors. Systematic reviews with relatively strong implications include Falchikov and Goldfinch (2000), Li and Ma (2010), Springer et al. (1999), and Tondeur et al. (2012). Conclusions should be grounded in the findings. Baumeister and Leary (1997) frame their advice in terms of common mistakes and how to avoid them; this advice is presented in Table 5.

**Systematic Identification of Reviews**

Throughout this article, we have cited reviews as examples of how specific systematic review procedures have been applied to engineering education topics; we identified the examples by applying systematic review procedures. We now describe how inclusion criteria, search procedures, critique, and appraisal steps were applied to identify these example reviews.
Our purpose was to identify exemplar systematic reviews on topics of interest to engineering education that engineering education researchers might use to guide their systematic review procedures, as well as to demonstrate systematic review procedures. Our inclusion criteria were:

- Journal article or conference paper published since 1990
- On an “engineering education topic” (engineering or STEM content or participants as topic area)
- Included a methods section with criteria for selecting primary studies

One author with expertise in systematic reviews searched the following databases: ERIC (ProQuest), Academic Search (EBSCO), Education Full Text (EBSCO), Educational Administration Abstracts (EBSCO), Applied Science & Technology Full Text (EBSCO), Science & Technology databases (ProQuest), and Papers First and Proceedings First (FirstSearch). The search used keyword combinations (with truncations) such as [engineer* and educat*] and [“systematic review” or “meta-analysis”] and identified 476 unique articles. We also requested, via engineering education research listservs, nominations of high-quality reviews meeting the criteria listed above. Responses provided 13 additional articles, bringing the total to 489 articles, which were entered into a RefWorks online database. These articles were randomly assigned to the three authors so that each article was screened by two authors. The two screeners disagreed on inclusion or exclusion of approximately 20 articles, and these borderline articles were discussed further until consensus was reached. A total of 37 articles met the inclusion criteria.

### Table 5 Mistakes and How to Avoid Them in Writing Systematic Reviews

<table>
<thead>
<tr>
<th>Mistake</th>
<th>Recommendation to Avoid Mistake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate introduction</td>
<td>Present conceptual and theoretical ideas early so readers can decide whether to read the lengthy review and consider all of the evidence.</td>
</tr>
<tr>
<td>Inadequate coverage of evidence</td>
<td>Include in primary study descriptions, methods, and specific results as well as conclusions of the studies.</td>
</tr>
<tr>
<td>Lack of integration</td>
<td>Relate individual primary studies together through an overarching conceptualization, perspective, or point-of-view (i.e., the take-home message).</td>
</tr>
<tr>
<td>Lack of critical appraisal</td>
<td>Point out and assess flaws and weaknesses of primary studies or groups of primary studies in order to weigh the evidence.</td>
</tr>
<tr>
<td>Failure to adjust conclusions</td>
<td>Temper conclusions based on critique of primary studies so that they are consistent with the overall strength of the evidence.</td>
</tr>
<tr>
<td>Blurring assertion and proof</td>
<td>Clearly distinguish between citations to sources which assert an idea and those that provide empirical evidence for it.</td>
</tr>
<tr>
<td>Selective review of evidence</td>
<td>Seek out and consider counterexamples.</td>
</tr>
<tr>
<td>Focusing on the researchers rather than the research</td>
<td>Deemphasize researcher names by putting citations in parentheses rather than in text.</td>
</tr>
<tr>
<td>Stopping at the present</td>
<td>Include implications for future research.</td>
</tr>
</tbody>
</table>

*Note: Adapted from Baumeister and Leary (1997).*
We adapted the quality appraisal form from Campbell Collaboration, and again, two authors conducted a quality appraisal of each paper. The final item on the quality appraisal form was an overall score for the systematic review. Eight articles were rated as “excellent” (on a three-point scale of poor, good, excellent) by both reviewers, and an additional six rated as excellent by one reviewer were identified through discussion to consensus. These 14 articles are presented in Table 6, which includes information on the topics and methods of each.

In sum, this brief description of our systematic review of systematic reviews demonstrates how some of the steps in conducting systematic reviews can be implemented in engineering education. We have identified 14 high-quality reviews on engineering education topics that can serve as models for engineering education researchers conducting their own systematic reviews. This set of exemplars included nearly equal numbers of quantitative (meta-analyses of quantitative primary studies) and qualitative (qualitative synthesis or meta-ethnographies of both quantitative and qualitative primary studies) synthesis approaches. We note that most of the exemplars by authors who routinely publish in engineering education venues fell into the qualitative category. We believe qualitative syntheses reflect the nature of the field and limited availability of primary studies that can support quantitative meta-analysis (i.e., studies conducted with control or comparison groups). Given this focus of engineering education on more qualitative approaches, and the availability of sources describing statistical meta-analysis methods, we have emphasized qualitative synthesis methods in our Synthesis section above.

Conclusion

Systematic review refers to a collection of evolving research methodologies for synthesizing existing studies to answer a set of research questions formulated by an interdisciplinary team. One purpose of this article was to introduce systematic review methodologies to the field of engineering education and to adapt existing resources on systematic reviews to engineering education and other developing interdisciplinary fields. On the basis of our review of systematic reviews, we suggest the approaches offer considerable promise for advancing the interdisciplinary field of engineering education with respect to several of the criteria identified by Fensham (2004): conceptual and theoretical development, research methodologies, progression, model publications, seminal publications, and implications for practice. Systematic reviews can benefit the field of engineering education by synthesizing prior work, better informing practice, and identifying important new directions for research.

Systematic reviews can be contrasted with narrative reviews, which are currently the most common type of review in engineering education, examples of which include invited articles and chapters in special issues of this Journal (Vol. 94, No. 1; Vol. 97, No. 3; and Vol. 100, No. 1), the Cambridge Handbook of Engineering Education Research (Johri & Olds, in press), and Heywood’s Engineering Education compendium (2005). This article is also primarily a narrative review. Narrative reviews, like other review approaches, attempt to synthesize or at least summarize prior work in a particular area and may stimulate synthesis of innovative conjectures and procedures (Eva, 2008). However, narrative reviews differ from systematic reviews in that the identification and selection criteria for sources are usually implicit; narrative reviews typically do not include methods sections. This lack of implicit methods leaves narrative reviews open to risks of bias (D. A. Cook & West, 2012) and incomplete coverage. Similarly, the analysis in narrative reviews tends to be ad hoc and likely to support the author’s intent, which may be to argue a certain stance. Systematic reviews can be distinguished from
Table 6 Exemplar Systematic Reviews on Engineering Education Topics

<table>
<thead>
<tr>
<th>Article</th>
<th>Type of review</th>
<th>Search approaches</th>
<th>Primary studies (no.)</th>
<th>Methods to be highlighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker and Park (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students’ learning: A preliminary meta-analysis</td>
<td>Quantitative review of quantitative studies</td>
<td>2 databases, internet, dissertations, limited to 1989–2009</td>
<td>28</td>
<td>Informative tables of primary studies and meta-analysis results</td>
</tr>
<tr>
<td>Beddoes, Jesiek, and Borrego (2010). Identifying opportunities for collaborations in international engineering education research on problem- and project-based learning</td>
<td>Qualitative review of qualitative studies</td>
<td>Conferences, journals</td>
<td>105</td>
<td>Expands upon hand searching traditionally used in reviews of engineering education research by consulting a variety of conferences and journals</td>
</tr>
<tr>
<td>Benitti (2012). Exploring the educational potential of robotics in schools: A systematic review</td>
<td>Qualitative review of quantitative studies</td>
<td>6 databases, limited to English only</td>
<td>10</td>
<td>Thorough and replicable search procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Provides rationale for decisions and numbers of included and excluded studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative lists and tables for several phases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Detailed descriptions of primary studies’ experimental designs</td>
</tr>
<tr>
<td>Falchikov and Goldfinch (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks</td>
<td>Quantitative review of quantitative studies</td>
<td>6 databases, citations, requested articles, limited to 1959–1999</td>
<td>48</td>
<td>Discussion of quality appraisal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative tables of primary studies and meta-analysis results</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Relatively strong future work discussion</td>
</tr>
<tr>
<td>Article</td>
<td>Type of review</td>
<td>Search approaches</td>
<td>Primary studies (no.)</td>
<td>Methods to be highlighted</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Li and Ma (2010). A meta-analysis of the effects of computer technology on school students’ mathematics learning</td>
<td>Quantitative review of quantitative studies</td>
<td>6 databases, citations, conferences, dissertations, limited to 1990–2006</td>
<td>46</td>
<td>Clear definition of terms, research questions, and inclusion criteria</td>
</tr>
<tr>
<td>Lou (2004). Understanding process and affective factors in small group versus individual learning with technology</td>
<td>Quantitative review of quantitative studies</td>
<td>5 databases, citations, dissertations</td>
<td>71</td>
<td>Informative tables of moderating variables identified in primary studies</td>
</tr>
<tr>
<td>Meese and McMahon (2012). Knowledge sharing for sustainable development in civil engineering: A systematic review</td>
<td>Qualitative review of qualitative studies</td>
<td>5 databases</td>
<td>87</td>
<td>Detailed table of features coded for mapping purposes</td>
</tr>
<tr>
<td>Monasor, Vizcaíno, Piatinni, and Caballero (2010). Preparing students and engineers for global software development: A systematic review</td>
<td>Qualitative review of qualitative studies</td>
<td>6 databases</td>
<td>38</td>
<td>Primarily a mapping study with limited synthesis</td>
</tr>
</tbody>
</table>

**Table 6 (continued)**
<table>
<thead>
<tr>
<th>Article</th>
<th>Type of review</th>
<th>Search approaches</th>
<th>Primary studies (no.)</th>
<th>Methods to be highlighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salleh, Mendes, and Grundy (2011). Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review</td>
<td>Quantitative review of quantitative studies (also some qualitative discussion of a broader group)</td>
<td>10 databases, citations, conferences</td>
<td>73</td>
<td>Good adaptation of meta-analysis procedures to a topic with few randomized control trials</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thorough and replicable search procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Detailed quality appraisal including checklist and scoring information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Used PICO framework to define intervention</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Flowchart of numbers of included and excluded studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative tables to summarize primary studies</td>
</tr>
<tr>
<td>Schmid et al. (2009). Technology’s effect on achievement in higher education: A Stage I meta-analysis of classroom applications</td>
<td>Quantitative review of quantitative studies</td>
<td>14 databases, citations, internet, dissertations, journals</td>
<td>231</td>
<td>Detailed procedures of how multiple researchers worked together</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clear definition of terms, research questions, and inclusion criteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Discussion of quality appraisal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative plots of meta-analysis results</td>
</tr>
<tr>
<td>Springer, Stanne, and Donovan (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis</td>
<td>Quantitative review of quantitative studies</td>
<td>4 databases, conferences, citations, dissertations</td>
<td>39</td>
<td>Informative plots of meta-analysis results that highlight moderating variables identified in primary studies</td>
</tr>
<tr>
<td>Article</td>
<td>Type of review</td>
<td>Search approaches</td>
<td>Primary studies (no.)</td>
<td>Methods to be highlighted</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Tamim, Bernard, Borokhovski, Abrami, and Schmid (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study</td>
<td>Quantitative review of quantitative studies</td>
<td>11 databases, dissertations, internet</td>
<td>25</td>
<td>Second order meta-analysis, i.e., meta-analysis of other meta-analyses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thorough and replicable search procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Extensive appendix of coding variables (for extraction) from meta-analyses in particular</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative table to summarize primary studies</td>
</tr>
<tr>
<td>Tondeur et al. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence</td>
<td>Qualitative meta-ethnography</td>
<td>1 database</td>
<td>19</td>
<td>Good description of meta-ethnography approach</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Detailed quality appraisal description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Informative tables to summarize results</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Well-organized results (around themes)</td>
</tr>
</tbody>
</table>

Full citations are included in the Reference list.
narrative reviews by their formalized synthesis procedures. Specific identification and synthesis approaches, matched to the research questions and types of primary studies selected, distinguish different types of systematic reviews.

A second purpose of this article was to provide an overview of steps in conducting a systematic review. As mentioned, systematic review refers to an evolving collection of methodologies (Tricco, Tetzlaff, & Moher, 2011) within which are ones for identifying the set of studies that will be analyzed as well as separate methodologies for synthesizing the set of studies. Within identifying methodologies, consideration must be given to the types of primary studies, e.g., quantitative, qualitative, mixed methods, and specific subtypes, that will be included in the collection. Often these decisions depend heavily on the research questions. Synthesizing methodologies draw from a wide variety of quantitative (e.g., statistical meta-analysis, network meta-analysis [Tricco et al., 2011]), qualitative (e.g., meta-ethnography, content analysis), and mixed method methodologies. In the past, the synthesizing methodology has been used to refer to the type of systematic review (e.g., meta-analysis), but as systematic review methodologies continue to evolve, it is becoming more widely recognized that the type of systematic review might be more accurately described as the choice of the synthesis methodology. Systematic procedures such as those described in this article ensure the quality and potential contributions of the review.

Third, we applied systematic review procedures to identify high-quality systematic reviews on engineering education topics. These 14 engineering education systematic reviews offer models for the systematic review process, including decisions at various steps.

Engineering education researchers should consider including systematic reviews in their repertoire of methodologies. We hope this article offers an overview, guidelines, and resources that will promote adoption of the methodologies.

Acknowledgments

This research was supported by the U.S. National Science Foundation while Maura Borrego was working for the Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


**Authors**

Maura Borrego is an associate dean and associate professor of engineering education at Virginia Tech, 660 McBryde Hall (0218), Blacksburg, VA 24061; mborrego@vt.edu.

Margaret J. Foster is an assistant professor in the Medical Sciences Library at Texas A&M University, College Station, TX 77843-4462; margaretfoster@tamu.edu.

Jeffrey E. Froyd is a TEES research professor at Texas A&M University, College Station, Texas 77843-3127; froyd@tamu.edu.