

# Some Potential “Pitfalls” in the Construction of Educational Systematic Reviews

John Coverdale<sup>1</sup> · Laura Weiss Roberts<sup>2</sup> · Eugene V. Beresin<sup>3</sup> · Alan K. Louie<sup>2</sup> · Adam M. Brenner<sup>4</sup> · Richard Balon<sup>5</sup>

Received: 16 January 2017 / Accepted: 19 January 2017 / Published online: 10 February 2017  
© Academic Psychiatry 2017

Educational reviews are potentially very valuable resources for educators because they shed light on the priority and methods of educational practices. Reviews can inform on the prevalence of educational problems and types of educational practices and policies, on the advantages and disadvantages of educational assessment tools, and on the effectiveness of methods for teaching and implementing the curriculum and its components. Educational reviews may assist educators in demonstrating how educational theory is implemented for a variety of learner audiences. Educational reviews also can assist in summarizing large bodies of material that may otherwise be challenging to handle. When there is a relative paucity of information on a topic, reviews can identify the importance of developing that literature and serve as a springboard for further research. Reviews can also assist by finding relevant studies that are difficult to obtain as well as by identifying and attempting to resolve divergent views by educators or authors on a particular topic area of interest.

*Academic Psychiatry* publishes systematic and other types of reviews that synthesize important information on a topic of general interest. An excellent example of a high-quality systematic review is published in the April 2017 issue of the journal [1]. Here, Abdool et al. reviewed studies applying simulation-based methodologies in undergraduate psychiatry

education and assessed the depth of learner engagement using simulation methods [1].

Because of the potential value and importance of reviews, notwithstanding the fun in constructing and learning from them, our goals for this editorial are to briefly describe and comment on the different types of educational reviews, their strengths and weaknesses, and the common pitfalls in their construction. We have been advantaged by the opportunity to learn from our authors and from the feedback provided by reviewers about the design and construction of reviews. As editors, we view ourselves as committed to learning about educational study designs. Our wish is to convey some of what we have learned to our readers and to prospective authors in order to assist in developing an understanding of the construction and appraisal of reviews. To this end, we will describe commonly encountered issues in reviews, particularly systematic reviews, submitted to *Academic Psychiatry* that may detract from their quality or the validity of their findings. We will also refer to the example of Abdool et al. [1] for tips on how to do it well.

## Review Architecture

Ordinarily, the typology of reviews includes narrative (traditional) reviews or systematic reviews [2], although these more realistically represent two ends of a continuum. Narrative educational reviews primarily depend on a subjective or idiosyncratic selection of papers. They typically do not identify a specific or focused question and usually make no effort to critically appraise the methodologies of the individual articles. Narrative educational reviews are therefore akin to routine editorials, commentaries, or many book chapters or thought pieces that share a potential to bias which is dependent on the attitudes and opinions of the authors. Narrative

---

✉ John Coverdale  
jhc@bcm.edu

<sup>1</sup> Baylor College of Medicine, Houston, TX, USA

<sup>2</sup> Stanford University, Stanford, CA, USA

<sup>3</sup> Harvard Medical School, Boston, MA, USA

<sup>4</sup> University of Texas Southwestern Medical Center, Dallas, TX, USA

<sup>5</sup> Wayne State University, Detroit, MI, USA

educational reviews are especially valuable, however, in providing a wide range of information on a general educational topic area of interest.

Well-conducted educational systematic reviews follow published standards [3–6]. These standards in turn inform on common pitfalls in their construction. Abdool et al. [1], for example, followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in their construction of the review. Usual standards begin with a focused or clearly formatted question, well-developed search strategies, explicit inclusion and exclusion criteria, and an explicit methodology for critically appraising the strengths and weaknesses of the individual studies incorporated into the review. An explicit description of the details of the methods in conducting the review enables other investigators to replicate and validate the findings. In addition, these processes enable readers to understand the limits of the review and the credibility of the findings. *Academic Psychiatry* understandably has a strong vested interest in encouraging authors to adopt systematic processes in order to reduce the potential for bias in the conduct of a review.

The utility of educational systematic reviews relates closely to an approach in medical education that promotes the best evidence in educational decision-making [7]. While all educational study designs have value so long as they assist in answering the question that was posed [8], the concept of best evidence supposes that there is a preferential ranking of study designs in quantitative educational research. Well-conducted randomized controlled trials are highly ranked in this evidence hierarchy because they tightly control relevant variables and limit potential confounding [9]. Although there is some controversy concerning their utility [10, 11], randomized controlled trials therefore generally outrank controlled non-randomized trials or educational cohort studies because randomization functions give every subject an equal chance to be assigned to the experimental or control group, a design feature that protects against confounding. Case control studies have particular value in the study of uncommon educational events, and cross-sectional studies similarly are especially useful in identifying educational needs and practices. Qualitative research may not be considered as belonging to this evidence hierarchy [9]. Reviews of qualitative research can serve to address an important limitation of that research by enhancing the generalizability of findings.

### Avoiding Potential Pitfalls

We found one excellent paper [12] that informed on the practical recommendations for planning an educational literature review and identified common problems areas. These problem areas are many, including outdated citations, too few or too many citations, citing of the most extreme studies, studies

conflicting with the authors' findings not being cited, inappropriately attributed ideas, overcitation of commentaries or editorials or of the authors' own work, lack of adequate synthesis of studies, lack of structure or organization of the review, and an overly long discussion. These authors broadly defined a literature review as a synthetic review and summary of what is known and unknown regarding the topic of a scholarly body of work. Reviews were therefore broadly defined in the context of scholarly articles, whereas systematic reviews were not specifically addressed.

We aim here to build on this helpful list of potential problem areas by reference to our knowledge and experiences in editing systematic reviews submitted to *Academic Psychiatry*. We identified and described pitfalls in thematic groups as presented below and summarized in Table 1.

### Focus of the Educational Question

On occasion, we have identified a lack of precision or focus in the research question. Moreover, elements of the questions may not have been directly incorporated into the search strategy. The educational question is the entryway to the search; the question and search strategy should be closely matched. In turn, the educational question is developed in the context of the available literature and the limitations of that literature which it seeks to address. Too broad a question may impede the specificity of searching and unduly diffuse the applicability of the findings. For example, Abdool et al.'s primary goal was to identify and describe simulation methods that were used in teaching psychiatry to medical students [1].

### Specification of the Inclusion Criteria

On occasion, authors have been overly inclusive in their selection of articles in the review. For example, little can be said about the utility of descriptions of curricula when they have no reported outcomes, and these perhaps should not be included. For another example, lumping widely varying study designs together without justification may overwhelm the review and readers by providing too much material.

Inclusion criteria can be directed or determined by the educational question, the study design, study quality, language (e.g., English or not English), and by the number of high-quality articles. An appreciation of the hierarchy of evidence can assist in deciding what to include. When there are a high number of potentially relevant articles, it may become desirable to include only more rigorous study designs. In this light, randomized controlled trials or controlled non-randomized trials may be selected over all other quantitative designs, because poorer quality designs may unduly bias the findings. The inclusion criteria can therefore be broadened or contracted depending on the volume of material found by a preliminary scoping review that examines the extent, range,

**Table 1** Some potential pitfalls and their remediation in the construction of educational systematic reviews

Lack of focus of the educational question.	The question is developed in context of the available literature. The question helps to develop the search strategy.
Lack of specification of the inclusion criteria.	The inclusion criteria should be justified and should not be too broad or too narrow for the purpose of the search strategy.
Lack of identification of the exclusion criteria.	The exclusion criteria should be justified and should not be too broad or too narrow for the purpose of the search strategy.
Lack of sensitivity of the search strategy.	Search strategies should use methods that aim to find all relevant articles and should specify the search terms and databases searched.
Lack of assessment of the methods of individual articles.	Judging the validity of methods of individual articles helps readers to learn which have the stronger and weaker methods.
Lack of synthesis or organization of the findings.	Presenting the review's findings by giving prominence to the better conducted studies or more important results facilitates readers' interpretation of the results of the review.
Lack of identification of the review's limitations.	Specifying the limitations helps readers to place the overall findings of the review in their limited context.

and nature of research activity in a given field [13]. For example, Abdool et al. specified three inclusion criteria: the studies had to (1) describe an educational intervention using a simulation methodology (2) in the area of psychiatry and (3) apply to undergraduate education [1].

### Identification of the Exclusion Criteria

On occasion, authors have not specified exclusion criteria or, alternatively, were overly exclusive. The inclusion criteria do not necessarily inform the exclusion criteria. Specifying the exclusion criteria helps to focus the search and illustrates to readers how decisions were made to select articles. Restricting the years of the search without justification, for example, may unnecessarily exclude articles of interest. Excluding articles because they are old or dated for the reason that they propagate outdated ideas which potentially falsely inflate study outcomes [12] might be justifiable but may also constitute a bias and unjustly exclude high-quality articles of interest.

Authors have discretion in setting the inclusion and exclusion criteria. As is the case of conducting reviews in public health research [14], to choose to set the bar too high in the selection of study designs may leave little to report. Alternatively, setting the bar too low with inclusion criteria that are too broad and with narrow exclusion criteria will lead to a vast array of heterogeneous evidence that is difficult to assimilate and integrate into a coherent whole. For example, Abdool et al. described that they excluded some duplicate articles and poorly defined interventions and review papers or commentaries [1].

### Sensitivity of the Search Strategy

On occasion, authors have limited the databases searched to one or have limited the number of search terms and their combinations. The search strategy is a key component of scientific reporting. Databases selected might include medical, educational, psychological, and general databases [15]; a less than comprehensive strategy risks missing articles of interest. This challenge is not an issue confined to *Academic Psychiatry* alone but has been found to occur in other leading educational journals. In one search of reviews in *Academic Medicine*, *Teaching and Learning in Medicine*, and *Medical Education*, only a minority of identified reviews explicitly described searching nonmedical databases or listed Medical Subject Headings (MeSH) and Boolean operators, and none included reproducible searches [16]. Of course, the sensitivity of the search strategy has to be balanced with its specificity so that there is not an overwhelming number of articles to screen.

Identifying all relevant articles, although very difficult to achieve, is critical to developing a valid review. Electronic searching may be supplemented by hand searching the journals most likely to include relevant articles, as well as citation searching. It is desirable to adopt a process that allows for an independent assessment of the validity of decision-making by having more than one judge and a system for resolving conflicts that arise regarding what should be included in the final tally of articles. Covidence is a screening and data extraction tool that can assist reviewers in these processes [17]. Identifying the details of the methods in the publication facilitates replication and enhances the justification of the recommendations that follow [18]. For example, Abdool et al. specified the keywords used for database searching in three

databases (MEDLINE, PsycINFO, and ERIC). They also presented a flow diagram describing the process of selection of the included articles [1].

### Judging the Validity of Methods in Individual Articles

On occasion, the validity of individual findings has not been judged, or when the findings are judged with the use of a tool, the tool may not have been an optimal fit. Judging the validity of the methods helps the readers to know which of the included articles deserves the most attention. One straightforward way of judging the methods is to state briefly their strengths and weaknesses, perhaps in a table that identifies the key study characteristics. Alternatively, authors could apply a well-validated tool relevant to the appraisal of the studies of interest. For educational controlled trials, for example, such a tool might incorporate an assessment of the presence of randomization, the presence of concealment of allocation, the presence or absence of differences between groups at baseline, the presence of blinding, and the presence or absence of an assessment of the psychometric properties of the outcome measures. For cross-sectional studies, the tool might assess the representativeness of the sample, including whether it was a local or national sample, the size of the sample, and the response rate, as well as the validity of the questions related to the educational practice or issue. Because these may not be easy judgments, it may help to have more than one person or team member rating the validity of individual studies and a process in place to resolve any disagreements.

It should be appreciated that methodological issues should not all count the same. For example, the presence of randomization may count as being more important than an assessment of the reliability of outcome measures even though both may be assigned a weight of 1. Moreover, a rating of the presence or absence of randomization or blinding does not, for example, allow an understanding of the adequacy of methods of randomization or blinding, which is a detail that requires a much deeper read. For these reasons, rigorous findings derived from systematic reviews involve thoughtfulness and careful judgment. For example, Abdool et al. more simply identified the study design of included articles. They also examined how the simulation methodology aligned itself with the depth of learner engagement based on a model of adult learning theory [1].

### Organizing and Synthesizing the Findings

On occasion, the results comment on findings without citing the articles within the compendium of included studies that directly support those findings. Each statement concerning the findings should be justified by citing the relevant articles that support it. Moreover, on occasion, authors have not well synthesized the findings of the review. Because the validity of any

individual study's conclusions are understood in relation to the adequacy and rigor of the methods for answering the question that was posed, a hierarchy of evidence can be constructed. One method is to first describe the better studies in the results section of the manuscript. Similarly, these better studies can be more thoroughly described on the assumption that these are the ones which readers will be most interested in learning about and emulating. Such an approach should counter the provision of a “laundry list” [12] of studies and reduce the weight given to the findings of studies of lesser quality. For example, Abdool et al. [1] gave prominence to those studies that satisfied the most criteria for more deeply engaging learners beyond a limited concrete experience of learning.

### Identifying the Review's Limitations

On occasion, the findings are not placed in context of the limitations of the review that supported them. These limitations commonly pertain to the adequacy of the search strategy to find all relevant articles and the adequacy of a tool and the related scoring system in placing the strengths and weaknesses of individual articles in an appropriate context. For example, Abdool et al. [1] identified several limitations, including that their review did not identify publications in the gray literature, such as industry reports.

### Overviews

Overviews, also known as umbrella reviews or meta-reviews, are essentially systematic reviews of systematic reviews. Their rationale is therefore similar to those for educational systematic reviews in that they identify a focused question justified by the importance of the topic. The importance of a topic may be justified in part by an identification of conflicting findings, which the overview may seek to resolve. In contrast to traditional reviews, their construction is also a systematic and disciplined process. Overviews should describe their inclusion and exclusion criteria, the search strategy, and the appraisal tool for judging the methodology of individual reviews. We have not come across any overviews published in *Academic Psychiatry* to date, although we hope to see these as the field grows and evidence accumulates. The concept of an overview underscores the critical point that educational systematic reviews are not by themselves definitive or determinative of educational practices. Instead, conclusions are shaped and molded as new information becomes available.

### Closing Thoughts

All types of educational reviews are potentially valuable. Systematic reviews, which begin with a focused question

and define their search strategies, inclusion and exclusion criteria, and critical appraisal strategies, are an especially important educational resource. Occasional pitfalls in the construction of educational systematic reviews include lack of focus in the educational question, lack of specification in the inclusion and exclusion criteria, limitations in the search strategies, limitations in the methods for judging the validity of findings of individual articles, lack of synthesis of the findings, and lack of identification of the review's limitations.

Educational systematic reviews are hard to do well, and they rely, to some extent, on the maturity of the field as well as the rigor of the review methodology. We at *Academic Psychiatry* understand these challenges. The use of the word *pitfalls* is not intended to be pejorative or discouraging to our authors. One of our goals in preparing this editorial, however, was to contribute to our efforts to publish the “best journal” [19] we can. On the assumption that anything worth doing is worth doing well, we hope to assist and encourage our prospective authors and readers to be cognizant of possible pitfalls. This knowledge should enhance the quality of the work in all of our respective roles for *Academic Psychiatry* and contribute to our success in publishing useful educational resources.

#### Compliance with Ethical Standards

**Disclosure** On behalf of all authors, the corresponding author states that there is no conflict of interest.

#### References

1. Abdool PS, Nirula L, Bonato S, Rajji TK, Silver IL. Simulation in undergraduate psychiatry: exploring the depth of learner engagement. *Acad Psychiatry*. 2017. doi:10.1007/s40596-016-0633-9.
2. McGaghie WC. Varieties of integrative scholarship: why rules of evidence, criteria, and standards matter. *Acad Med*. 2015;90:294–302.
3. Cook DA. Conducting systematic reviews in medical education; a stepwise approach. *Med Educ*. 2012;46:943–52.
4. Hammick M, Dornan T, Steinert Y. Conducting a best evidence systematic review. Part 1: from idea to data coding. BEME Guide No. 13. *Med Teach*. 2010;32:3–15.
5. PRISMA. The preferred reporting items for systematic reviews and meta-analyses. Available at <http://www.prisma-statement.org/>. Last accessed 19 January 2017.
6. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65–94.
7. BEME. Best-evidence medical and health professional education. Available at <http://www.bemecollaboration.org>. Last accessed 19 January 2017.
8. Horn C, Plazas Snyder B, Coverdale J, et al. Educational research questions and study design. *Acad Psychiatry*. 2009;33:261–67.
9. Coverdale J, Balon R, Beresin EV, et al. An argument for conducting methodologically strong, randomized controlled trials in educational research. *Acad Psychiatry*. 2013;37:145–9.
10. Norman G. RCT = results confounded and trivial; the perils of grand educational experiments. *Med Educ*. 2003;37:582–4.
11. Thistlewaite J, Davies H, Dornan T, et al. What is evidence? Reflections on the AMEE symposium, Vienna, August 2011. *Med Teach*. 2012;34:454–7.
12. Maggio LA, Sewell JL, Artino AR. The literature review: a foundation for high quality medical education research. *J Grad Med Educ*. 2016;8:297–303.
13. Thomas A, Lubarsky S, Durning SJ, Young ME. Knowledge syntheses in medical education: demystifying scoping reviews. *Acad Med*. 2016. doi:10.1097/ACM.0000000000001452.
14. McCall J, Connor J. Systematic reviews in public health research. *Aust NZ J Public Health*. 2010;34:343–4.
15. Sewell JL, Maggio LA, Artino AR. Planning the literature review. *Acad Med*. 2016. doi:10.1097/ACM.0000000000001296.
16. Maggio LA, Tannery NH, Kanter SL. Reproducibility of literature search reporting in medical education reviews. *Acad Med*. 2011;86:1049–54.
17. Covidence. Accelerate your systematic review. Available at <https://www.covidence.org>. Last accessed 19 January 2017.
18. Coverdale J, Roberts LW, Louie A, Beresin E. Writing the methods. *Acad Psychiatry*. 2006;30:361–4.
19. World Association of Medical Editors. WAME professionalism code of conduct. <http://www.wame.org/wame-professionalism-code-of-conduct>. Last accessed 19 January 2017.