ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/337446956

# Systematic Reviews in Educational Research: Methodology, Perspectives and Application

**Chapter** · January 2020

DOI: 10.1007/978-3-658-27602-7_1

CITATIONS
151

READS
1,790

**2 authors**, including:

Mark Newman
University College London

**69** PUBLICATIONS **1,424** CITATIONS

SEE PROFILE

All content following this page was uploaded by Mark Newman on 25 November 2019.

The user has requested enhancement of the downloaded file.

Olaf Zawacki-Richter · Michael Kerres · Svenja Bedenlier · Melissa Bond · Katja Buntins *Eds.*

RESEARCH

# Systematic Reviews in Educational Research

## Methodology, Perspectives and Application

OPEN

Springer VS

# Systematic Reviews in Educational Research

Olaf Zawacki-Richter · Michael Kerres ·
Svenja Bedenlier · Melissa Bond ·
Katja Buntins
Editors

# Systematic Reviews in Educational Research

## Methodology, Perspectives and Application

OPEN

Springer VS

*Editors*
Olaf Zawacki-Richter
Oldenburg, Germany

Svenja Bedenlier
Oldenburg, Germany

Katja Buntins
Essen, Germany

Michael Kerres
Essen, Germany

Melissa Bond
Oldenburg, Germany

ISBN 978-3-658-27601-0 ISBN 978-3-658-27602-7 (eBook)
https://doi.org/10.1007/978-3-658-27602-7

Springer VS
© The Editor(s) (if applicable) and The Author(s) 2020. This book is an open access publication.
**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.
The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer VS imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH part of Springer Nature.
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

# Introduction: Systematic Reviews in Educational Research

## Introduction

In any research field, it is crucial to embed a research topic into the broader framework of research areas in a scholarly discipline, to build upon the body of knowledge in that area and to identify gaps in the literature to provide a rationale for the research question(s) under investigation. All researchers, and especially doctoral students and early career researchers new to a field, have to familiarize themselves with the existing body of literature on a given topic. Conducting a systematic review provides an excellent opportunity for this endeavour.

As educational researchers in the field of learning design, educational technology, and open and distance learning, our scholarship has been informed by quantitative and qualitative methods of empirical inquiry. Some of the colleagues in our editorial team were familiar with methods of content analysis using text-mining tools to map and explore the development and flow of research areas in academic journals in distance education, educational technology and international education (e.g., Zawacki-Richter and Naidu 2016; Bond and Buntins 2018; Bedenlier et al. 2018). However, none of us was a practitioner or scholar of systematic reviews, when we came across a call for research projects by the German Ministry of Education and Research (BMBF) in 2016.

The aim of this research funding program was to support empirical research on digital higher education, and the effectiveness and effects of current approaches and modes of digital delivery in university teaching and learning. Furthermore, it was stated in the call for research projects:

> In addition, research syntheses are eligible for funding. Systematic synthesis of the state of international research should provide the research community and practitioners with knowledge on the effects of certain forms of learning design with

> regard to the research areas and practice in digital higher education described below. Where the research or literature situation permits, systematic reviews in the narrower methodological sense can also be funded (BMBF 2016, p. 2).

What is meant here by "systematic reviews in the narrower methodological sense"? In contrast to traditional or narrative literature reviews, that are criticised as being biased and arbitrary, the aim of a systematic review is to carry out a review that is rigorous and transparent in each step of the review process, to make it reproducible and updateable. "Rather than looking at any study in isolation, we need to look at the body of evidence" (Nordenbo 2009, p. 22) to show systematically that existing primary research results contain arguments to shape and inform practice and policies.

The review question of our systematic review project was concerned with student engagement and educational technology in higher education (see Chap. 7). It became obvious very quickly, that we were dealing with very broad and "fuzzy" concepts here, travelling within an interdisciplinary domain using inconsistent terminology, which made it difficult to develop a very straightforward search strategy. The PICO framework developed in evidence-based medicine and health science (Schardt et al. 2007) to define the population, the intervention, the comparator and outcome or impact of an intervention is less useful in many educational review studies, where there is no clear 'treatment' (e.g. a drug) that leads to a well-defined outcome (e.g. the patient is dead or not dead). Study results are often qualitative in nature or the variance of the different variables at work in an educational setting is too complex to calculate a simple combining effect size across studies in a meta-analysis (see Borenstein et al. 2009).

## The Purpose and Structure of this Book

Given the growing interest in conducting systematic reviews in order to inform policy, and to support research and practice in education (Polanin et al. 2017), the purpose of this volume is to explore the methodology of systematic reviews, as well as the opportunities and challenges of doing a systematic review, in the context of educational research.

Thus, this book is divided into two sections: Authors in the first part provide an overview of various methodological aspects of systematic reviews. We approach this topic from different perspectives. Scholars of the systematic review method (see Chaps. 1, 2 and 3) introduce us to the steps involved in the systematic review process and elaborate on the advantages and disadvantages of

different types of literature reviews, as well as criticisms and ethical dimensions of systematic reviews. One colleague (see Chap. 4) writes about the pedagogy of methodological learning and teaching systematic reviews. And finally, editors of a very prestigious educational research journal (see Chap. 5) write about the benefits of publishing systematic reviews and share some positive examples that can serve as guidelines for researchers new to systematic reviews, in order to get their work published in a peer-reviewed journal.

Reading about a method in a research methods textbook is one thing, actually applying a method and doing the research in a specific context is another. Thus, for the second part of the book, we invited educational researchers coming from educational psychology, educational technology, instructional design and higher education research to share their experiences as worked examples, and to reflect on the promises and pitfalls in each step of the review process. We hope that these examples will be particularly helpful and can serve as a kind of roadmap for colleagues who are conducting a systematic review for the first time.

## Part I: Methodological considerations

In the first Chapter, Mark Newman and David Gough from the University College London Institute of Education, introduce us to the method of systematic review. Depending on the aims of a literature review, they provide an overview of various approaches in review methods. In particular, they explain the differences between an aggregative and configurative synthesis logic that are important for reviews in educational research. We are guided through the steps in the systematic review process that are documented in a review protocol: defining the review question, developing the search strategy, the search string, selecting search sources and databases, selecting inclusion and exclusion criteria, screening and coding of studies, appraising their quality, and finally synthesizing and reporting the results.

Martyn Hammersley from the Open University in the UK offers a critical reflection on the methodological approach of systematic reviews in the second chapter. He begins his introduction with a historical classification of the systematic review method, in particular the evidence-based medicine movement in the 1980s and the role of rigorous Randomised Controlled Trials (RCTs). He emphasises that in the educational sciences RCTs are rare and alternative ways of synthesising research findings are needed, including evidence from qualitative studies. Two main criticisms of systematic reviewing are discussed, one coming from qualitative and the other from realist evaluation researchers. In light of these

criticisms, Hammersley continues to reflect on the methodological features of systematic reviews, in relation to exhaustive searching for relevant material, the transparent methodological assessment of studies, and the synthesis of findings.

In the third chapter, Harsh Suri from Deakin University in Australia elaborates on ethical issues that might be involved in the systematic review process. Systematic reviews are widely read and cited in documents that have an impact on educational policy and practice. Authors have to reveal potential conflicts of interest and reflect on if or how the agenda of a funding source might influence the review process and the synthesis of findings, including various publication and search biases.

Melanie Nind from the Southampton University Education School in the UK, writes about teaching the systematic review method in Chap. 4. Given the growing interest in systematic reviewing, it is necessary to reflect on methodological learning and teaching, especially on the level of postgraduate and doctoral education. She concludes: "Teaching systematic review, as with teaching many social research methods, requires deep knowledge of the method and a willingness to be reflexive and open about its messy realities; to tell of errors that researchers have made and judgements they have formed" (p. 66). We will dig deeper into these "messy realities" of doing a systematic review in the second part of this book.

The first section on methodological considerations finishes with some enlightening reflections from Alicia Dowd and Royel Johnson, both from Pennsylvania State University in the USA, on publishing systematic reviews from an editor's and reader's perspective. Alica is an Associate Editor of the Review of Educational Research (RER), the highest impact factor journal within the SSCI Education & Educational Research category. The overall trend, in terms of the proportion of the total number of reviews published as systematic review articles in RER, has been upward as well. Both authors stress that systematic review authors should 'story' their findings in compelling ways, rather than reporting facts in mechanical and algorithmic terms. Examples of how to do this, are illustrated by selected papers from RER. To get published in a journal like RER, requires of course rigorous application of the review or meta-analysis method. However, the authors of this chapter remind us, that we should try not to get too exhausted by the time- and labour-consuming tasks involved in the systematic review process, and thereby neglect to put much effort into the 'storying' of the synthesis, discussion and implication sections in a review.

## Part II: Examples and Applications

For Chaps. 6, 7, 8 and 9, we invited authors to write about their practical experiences in conducting systematic reviews in educational research. Along the lines of the subsequent steps in the systematic review method, they discuss various challenges, problems and potential solutions:

- The first example in Chap. 6 is provided by Joanna Tai, Rola Ajjawi, Margaret Bearman, and Paul Wiseman from Deakin University in Australia. They carried out a systematic review about the conceptualisations and measures of student engagement.
- Svenja Bedenlier and co-authors from the University of Oldenburg and the University of Duisburg-Essen in Germany also dealt with student engagement, but they were more interested in the role that educational technology can play to support student engagement in higher education (Chap. 7).
- Chung Kwan Lo from the University of Hong Kong reports on a series of five systematic reviews in Chap. 8 that he has worked on in different teams. They were investigating the application of video-based lectures (flipped learning) in K-12 and higher education in various subject areas.
- Naska Goagoses and Ute Koglin from the University of Oldenburg in Germany share the last example in Chap. 9, which addresses the topic of social goals and academic success.

Common challenges of systematic reviews in education that derive from these reports are summarized in the remainder of this introduction.

## Critical Aspects and Common Challenges of Systematic Reviews in Education

As the systematic review examples that are described in this book show, carrying out a systematic review is a labour-intensive exercise. In a study on the time and effort needed to conduct systematic reviews, Borah et al. (2017) analysed 195 systematic reviews of medical interventions. They report a mean time to complete a systematic review project and to publish the review of 67.3 weeks (SD = 31.0; range 6–186 weeks). Interestingly, reviews that reported funding took longer (42 vs 26 weeks) and involved more team members (6.8 vs 4.8 persons) than reviews that reported no funding. The most time-consuming task, after the removal of

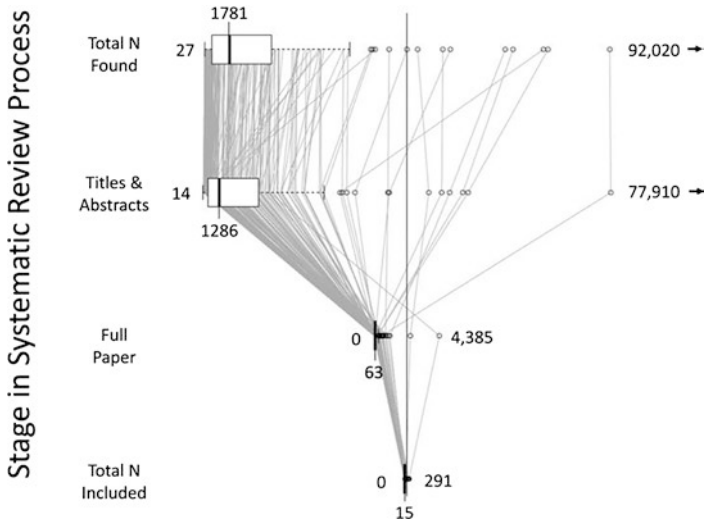



**Fig. 1** Literature filtration process (N = 195; Borah et al. 2017, p. 4)

duplicates from different databases, is the screening and coding of a large number of references, titles, abstracts and full papers. As shown in Fig. 1 from the Borah et al. (2017) study, a large number of full papers (median = 63, maximum = 4385) have to be screened to decide if they meet the final inclusion criteria. The filtering process in each step of the systematic review can be dramatic. Borah et al. report a final average yield rate of below 3%.

The examples of systematic reviews in educational research presented in this volume, echo the results of Borah et al. (2017) from the medical field. Compared to these numbers, the reviews included in this book represent the full range; from smaller systematic reviews that were finished within a couple of months, with only five included studies, to very large systematic reviews that were carried out by an interdisciplinary team over a time period of two years, including over 70,000 initial references and 243 finally included studies. Table 1 provides an overview of the topics, duration, team size, databases searched and the filtration process of the educational systematic review examples presented in Chaps. 6, 7, 8 and 9.

Based on our authors' accounts and reflection on the sometimes thorny path of conducting systematic reviews, some recurrent issues we have to deal with especially in the field of educational research can be identified.

**Table 1** Comparison of educational systematic review examples

| | Tai et al. (Chap. 6) | Bedenlier et al. (Chap. 7) | Lo (Chap. 8) | Goagoses and Koglin (Chap. 9) |
|---|---|---|---|---|
| Topic | Conceptualization and measurement of student engagement | Student engagement and educational technology in higher education | Flipped and video-based learning in various subject areas in higher education | Social goals and academic success |
| Duration | 18 months | 24 months | 1–4 months | 11 months |
| No of team members | 4 authors, 1 research assistant | 5 authors, 2 research assistants | 1–3 authors | 2 authors |
| Initial references | 4,192 | 77,508 | 936–4053 | 2,270 |
| Final references | 186 | 243 | 5–61 | 26 |
| Yield rate | 4.44% | 0.31% | 0.05–1.51% | 1.14% |
| Databases searched | psycINFO, ERIC, Education Source, and Academic Search Complete were accessed via Ebscohost, Scopus, Web of Science | ERIC, Web of Science, psycINFO, and SCOPUS | Academic Search Complete, TOC Premier, and ERIC, PubMed, psycINFO, CINAHL Plus, and British Nursing Index | Web of Science Core Collection, Scopus, and psycINFO |

The perhaps major challenge of conducting systematic reviews in educational research is the 'messiness', which is inherent in domains that use inconsistent terminology and multifaceted concepts like 'student engagement' or 'educational technology' (see Chap. 6 and 7). In such cases, it is crucial to find the right balance between comprehensiveness and relevance, or sensitivity and precision (Brunton et al. 2012), in developing the search strategy. For example, Bedenlier et al. (Chap. 7) decided to leave out any phrase relating to student engagement in the search string and to search for indicators of engagement and disengagement instead. The broadness of this approach made it possible to find research that would have been missed with a more precise search focus on 'engagement': only 26% of the finally included studies in their review explicitly used the term 'student engagement' in the title or abstract.

As already mentioned above, there is often not a clear intervention and outcome in educational research projects in the sense of variable x leads to y. Reviews often deal with questions that begin with how, what, and why: How is student engagement conceptualized? What kind of educational technology can be used to support student interaction? Why and under which conditions do video-based lectures lead to more student attention? Systematic reviews on such questions are therefore more configurative rather than aggregative in nature (see Chap. 1, Sect. 1.2). The exploration of broad concepts requires an open and iterative review approach.

The inclusion of qualitative research is another common challenge in synthesising educational research. Bedenlier et al. (Chap. 7) emphasize the benefits of working in a team with quantitative and qualitative method knowledge, while Goagoses and Koglin (Chap. 9) decided to exclude all qualitative and mixed-method articles "due to the differential methodologies described for systematic reviews of qualitative and quantitative articles and a lack of clear guidance concerning their convergence" (p. 151). However, various methods for the integration of qualitative research have been developed, but again particularly in relation to medicine and health related research (see Hannes and Lockwood 2011). An overview of the various approaches to synthesizing qualitative research and their differences with regard to epistemological assumptions, the extent of iteration during the review process, and quality assessment is provided by Barnett-Page and Thomas (2009). As Martyn Hammersley points out in Chap. 2, systematic reviewing should not downplay the value of qualitative research.

Doing a systematic review is a very fruitful exercise in any research project to gain a solid overview of the relevant body of literature. This makes particular

sense for early career researchers and doctoral students who start to develop their own research topics and agendas. We hope that this introduction to the systematic review method, and the practical hands-on examples presented in this volume, will serve as a useful resource for educational researchers new to the systematic review method.

Olaf Zawacki-Richter

## References

Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology*, *9*(1), 1–11. https://doi.org/10.1186/1471-2288-9-59.

Bedenlier, S., Kondakci, Y., & Zawacki-Richter, O. (2018). Two decades of research into the internationalization of higher education: major themes in the *Journal of Studies in International Education* (1997–2016). *Journal of Studies in International Education*, *22*(2), 108–135. https://doi.org/10.1177/1028315317710093.

BMBF. (2016). *Richtlinie zur Förderung von Forschung zur digitalen Hochschulbildung – Wirksamkeit und Wirkungen aktueller Ansätze und Formate – Trends und neue Paradigmen in Didaktik und Technik*. Retrieved from https://www.bundesanzeiger.de.

Bond, M., & Buntins, K. (2018). An analysis of the Australasian Journal of Educational Technology 2013–2017. *Australasian Journal of Educational Technology, 34*(4), 168–183. https://doi.org/10.14742/ajet.4359.

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*(2), e012545. https://doi.org/10.1136/bmjopen-2016-012545.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, U.K: John Wiley & Sons.

Brunton, G., Stansfield, C., & Thomas, J. (2012). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 107–134). London ; Thousand Oaks, Calif: SAGE.

Hannes, K., & Lockwood, C. (Eds.). (2011). *Synthesizing qualitative research: choosing the right approach*. https://doi.org/10.1002/9781119959847.

Nordenbo, S. E. (2010). Evidence and synthesis: a new paradigm in educational research. In The Research Council of Norway (Ed.), *Rigour and relevance in educational research* (pp. 21–27). St. Hanshaugen, Norway: The Research Council of Norway.

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in Education Research: A Systematic Review and Analysis. *Review of Educational Research*, *87*(1), 172–203. https://doi.org/10.3102/0034654316631117.

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, *7*(1). https://doi.org/10.1186/1472-6947-7-16.

Zawacki-Richter, O., & Naidu, S. (2016). Mapping research trends from 35 years of publications in *Distance Education*. *Distance Education*, *37*(3), 245–269. https://doi.org/10.1080/01587919.2016.1185079.

# Acknowledgment

This research resulted from the ActiveLeaRn project, funded by the Bundesministerium für Bildung und Forschung (BMBF, German Ministry of Education and Research); grant number 16DHL1007.

# Contents

**Part I Methodological Considerations**

**Systematic Reviews in Educational Research: Methodology, Perspectives and Application** . . . . . 3
Mark Newman and David Gough

**Reflections on the Methodological Approach of Systematic Reviews** . . . 23
Martyn Hammersley

**Ethical Considerations of Conducting Systematic Reviews in Educational Research** . . . . . 41
Harsh Suri

**Teaching Systematic Review** . . . . . 55
Melanie Nind

**Why Publish a Systematic Review: An Editor's and Reader's Perspective** . . . . . 69
Alicia C. Dowd and Royel M. Johnson

**Part II Examples and Applications**

**Conceptualizations and Measures of Student Engagement: A Worked Example of Systematic Review** . . . . . 91
Joanna Tai, Rola Ajjawi, Margaret Bearman and Paul Wiseman

**Learning by Doing? Reflections on Conducting a Systematic Review in the Field of Educational Technology** . . . . . 111
Svenja Bedenlier, Melissa Bond, Katja Buntins, Olaf Zawacki-Richter and Michael Kerres

**Systematic Reviews on Flipped Learning in Various Education Contexts** . . . . . 129
Chung Kwan Lo

**The Role of Social Goals in Academic Success: Recounting the Process of Conducting a Systematic Review** . . . . . 145
Naska Goagoses and Ute Koglin

# Contributors

**Dr. Rola Ajjawi,** Associate Professor in Educational Research, Deakin University (Australia), Centre for Research in Assessment and Digital Learning (CRADLE), E-Mail: rola.ajjawi@deakin.edu.au.

**Dr. Margaret Bearman,** Associate Professor, Deakin University (Australia), Centre for Research in Assessment and Digital Learning (CRADLE), E-Mail: margaret.bearman@deakin.edu.au.

**Dr. Svenja Bedenlier,** Research Associate, Carl von Ossietzky University of Oldenburg (Germany), Center for Open Education Research (COER), Institute of Education, E-Mail: svenja.bedenlier@uni-oldenburg.de.

**Melissa Bond,** Research Associate, PhD student, Carl von Ossietzky University of Oldenburg (Germany), Center for Open Education Research (COER), Institute of Education, E-Mail: melissa.bond@uni-oldenburg.de.

**Katja Buntins,** Research Associate, PhD student, University Duisburg-Essen (Germany), Department of Educational Sciences, Learning Lab, E-Mail: katja.buntins@uni-due.de.

**Dr. Alicia C. Dowd,** Professor of Education, The Pennsylvania State University, College of Education, Department of Education Policy Studies, Director of the Center for the Study of Higher Education (CSHE) and the CSHE Academic Leadership Academy, E-Mail: dowd@psu.edu.

**Dr. Naska Goagoses,** Research Associate, Carl von Ossietzky University of Oldenburg (Germany), Department of Special Needs Education & Rehabilitation, E-mail: naska.goagoses@uni-oldenburg.de.

**Dr. David Gough,** Professor of Evidence-informed Policy and Practice, University College London, Institute of Education, Director of the EPPI Centre, E-Mail: david.gough@ucl.ac.uk.

**Dr. Martyn Hammersley,** Professor emeritus, Open University UK, Faculty of Wellbeing, Education & Language Studies, E-Mail: martyn.hammersley@open.ac.uk.

**Dr. Royel M. Johnson,** Assistant Professor, Research Associate, The Pennsylvania State University, College of Education, Department of Education Policy Studies, Center for the Study of Higher Education (CSHE), Department of African American Studies, E-Mail: rmj19@psu.edu.

**Dr. Michael Kerres,** Professor of Educational Science, Chair of Educational Media and Knowledge Management at University Duisburg-Essen (Germany), Department of Educational Sciences, Director of the Learning Lab, E-Mail: michael.kerres@uni-due.de.

**Dr. Ute Koglin,** Professor of Special Needs Education, Carl von Ossietzky University of Oldenburg (Germany), Department of Special Needs Education & Rehabilitation, E-mail: ute.koglin@uni-oldenburg.de.

**Dr. Chung Kwan Lo,** EdD graduate, University of Hong Kong, Faculty of Education, E-Mail: cklohku@gmail.com.

**Dr. Mark Newman,** Reader Evidence-informed Policy and Practice in Education and Social Policy, University College London, Institute of Education, Associate Director of the EPPI Centre, E-Mail: mark.newman@ucl.ac.uk.

**Dr. Melanie Nind,** Professor of Education, University of Southampton, Southampton Education School, Co-Director, ESRC National Centre for Research Methods, E-Mail: M.A.Nind@soton.ac.uk.

**Dr. Harsh Suri,** Senior Lecturer, Learning Futures, Deakin University (Australia), E-Mail: harsh.suri@deakin.edu.au.

**Dr. Joanna Tai,** Research Fellow, Deakin University (Australia), Centre for Research in Assessment and Digital Learning (CRADLE), E-Mail: joanna.tai@deakin.edu.au.

**Paul Wiseman,** PhD student, University of Melbourne (Australia), Melbourne Centre for the Study of Higher Education (CSHE), E-Mail: paul.wiseman@unimelb.edu.au.

**Dr. Olaf Zawacki-Richter,** Professor of Educational Technology and Learning Design, Carl von Ossietzky University of Oldenburg (Germany), Institute of Education, Director of the Center for Lifelong Learning (C3L) and the Center for Open Education Research (COER), E-Mail: olaf.zawacki.richter@uni-oldenburg.de.

# Part I
# Methodological Considerations

# Systematic Reviews in Educational Research: Methodology, Perspectives and Application

Mark Newman and David Gough

## 1 What Are Systematic Reviews?

A literature review is a scholarly paper which provides an overview of current knowledge about a topic. It will typically include substantive findings, as well as theoretical and methodological contributions to a particular topic (Hart 2018, p. xiii). Traditionally in education 'reviewing the literature' and 'doing research' have been viewed as distinct activities. Consider the standard format of research proposals, which usually have some kind of 'review' of existing knowledge presented distinctly from the methods of the proposed new primary research. However, both reviews and research are undertaken in order to find things out. Reviews to find out what is already known from pre-existing research about a phenomena, subject or topic; new primary research to provide answers to questions about which existing research does not provide clear and/or complete answers.

When we use the term research in an academic sense it is widely accepted that we mean a process of asking questions and generating knowledge to answer these questions using rigorous accountable methods. As we have noted, reviews also share the same purposes of generating knowledge but historically we have not paid as much attention to the methods used for reviewing existing literature as we have to the methods used for primary research. Literature reviews can be used for

M. Newman (✉) · D. Gough
Institute of Education, University College London, England, UK
e-mail: mark.newman@ucl.ac.uk

D. Gough
e-mail: david.gough@ucl.ac.uk

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_1

making claims about what we know and do not know about a phenomenon and also about what new research we need to undertake to address questions that are unanswered. Therefore, it seems reasonable to conclude that 'how' we conduct a review of research is important.

The increased focus on the use of research evidence to inform policy and practice decision-making in Evidence Informed Education (Hargreaves 1996; Nelson and Campbell 2017) has increased the attention given to contextual and methodological limitations of research evidence provided by single studies. Reviews of research may help address these concerns when carried on in a systematic, rigorous and transparent manner. Thus, again emphasizing the importance of 'how' reviews are completed.

The logic of systematic reviews is that reviews are a form of research and thus can be improved by using appropriate and explicit methods. As the methods of systematic review have been applied to different types of research questions, there has been an increasing plurality of types of systematic review. Thus, the term 'systematic review' is used in this chapter to refer to a family of research approaches that are a form of secondary level analysis (secondary research) that brings together the findings of primary research to answer a research question. Systematic reviews can therefore be defined as "a review of existing research using explicit, accountable rigorous research methods" (Gough et al. 2017, p. 4).

## 2 Variation in Review Methods

Reviews can address a diverse range of research questions. Consequently, as with primary research, there are many different approaches and methods that can be applied. The choices should be dictated by the review questions. These are shaped by reviewers' assumptions about the meaning of a particular research question, the approach and methods that are best used to investigate it. Attempts to classify review approaches and methods risk making hard distinctions between methods and thereby to distract from the common defining logics that these approaches often share. A useful broad distinction is between reviews that follow a broadly configurative synthesis logic and reviews that follow a broadly aggregative synthesis logic (Sandelowski et al. 2012). However, it is important to keep in mind that most reviews have elements of both (Gough et al. 2012).

Reviews that follow a broadly configurative synthesis logic approach usually investigate research questions about meaning and interpretation to explore

and develop theory. They tend to use exploratory and iterative review methods that emerge throughout the process of the review. Studies included in the review are likely to have investigated the phenomena of interest using methods such as interviews and observations, with data in the form of text. Reviewers are usually interested in purposive variety in the identification and selection of studies. Study quality is typically considered in terms of authenticity. Synthesis consists of the deliberative configuring of data by reviewers into patterns to create a richer conceptual understanding of a phenomenon. For example, meta ethnography (Noblit and Hare 1988) uses ethnographic data analysis methods to explore and integrate the findings of previous ethnographies in order to create higher-level conceptual explanations of phenomena. There are many other review approaches that follow a broadly configurative logic (for an overview see Barnett-Page and Thomas 2009); reflecting the variety of methods used in primary research in this tradition.

Reviews that follow a broadly aggregative synthesis logic usually investigate research questions about impacts and effects. For example, systematic reviews that seek to measure the impact of an educational intervention test the hypothesis that an intervention has the impact that has been predicted. Reviews following an aggregative synthesis logic do not tend to develop theory directly; though they can contribute by testing, exploring and refining theory. Reviews following an aggregative synthesis logic tend to specify their methods in advance (a priori) and then apply them without any deviation from a protocol. Reviewers are usually concerned to identify the comprehensive set of studies that address the research question. Studies included in the review will usually seek to determine whether there is a quantitative difference in outcome between groups receiving and not receiving an intervention. Study quality assessment in reviews following an aggregative synthesis logic focusses on the minimisation of bias and thus selection pays particular attention to homogeneity between studies. Synthesis aggregates, i.e. counts and adds together, the outcomes from individual studies using, for example, statistical meta-analysis to provide a pooled summary of effect.

## 3 The Systematic Review Process

Different types of systematic review are discussed in more detail later in this chapter. The majority of systematic review types share a common set of processes. These processes can be divided into distinct but interconnected stages as illustrated in Fig. 1. Systematic reviews need to specify a research question and the methods that will be used to investigate the question. This is often written

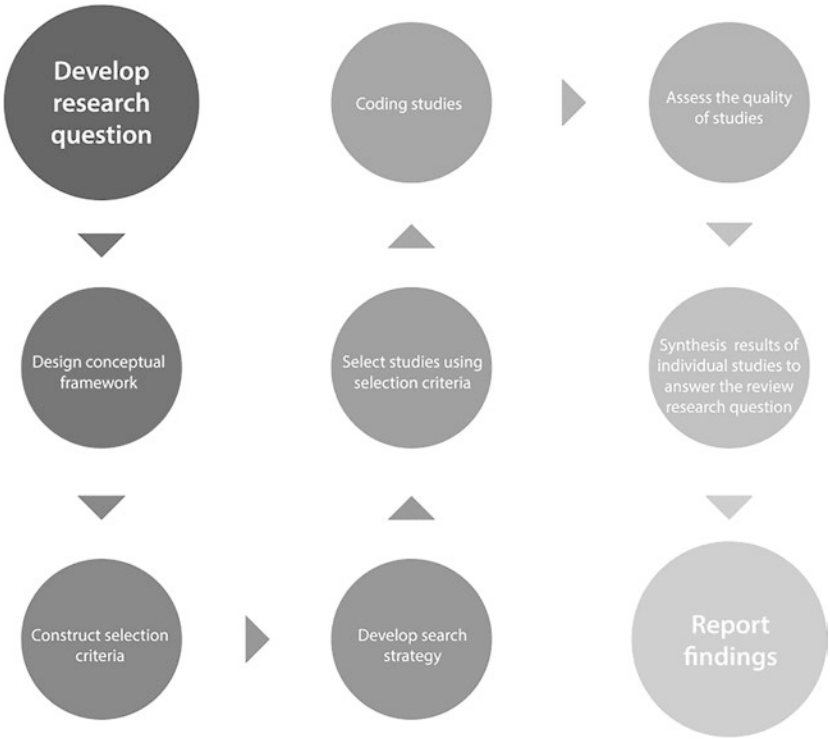



**Fig. 1** The systematic review process

as a 'protocol' prior to undertaking the review. Writing a protocol or plan of the methods at the beginning of a review can be a very useful activity. It helps the review team to gain a shared understanding of the scope of the review and the methods that they will use to answer the review's questions. Different types of systematic reviews will have more or less developed protocols. For example, for systematic reviews investigating research questions about the impact of educational interventions it is argued that a detailed protocol should be fully specified prior to the commencement of the review to reduce the possibility of reviewer bias (Torgerson 2003, p. 26). For other types of systematic review, in which the research question is more exploratory, the protocol may be more flexible and/or developmental in nature.

## 3.1 Systematic Review Questions and the Conceptual Framework

The review question gives each review its particular structure and drives key decisions about what types of studies to include; where to look for them; how to assess their quality; and how to combine their findings. Although a research question may appear to be simple, it will include many assumptions. Whether implicit or explicit, these assumptions will include: epistemological frameworks about knowledge and how we obtain it, theoretical frameworks, whether tentative or firm, about the phenomenon that is the focus of study.

Taken together, these produce a conceptual framework that shapes the research questions, choices about appropriate systematic review approach and methods. The conceptual framework may be viewed as a working hypothesis that can be developed, refined or confirmed during the course of the research. Its purpose is to explain the key issues to be studied, the constructs or variables, and the presumed relationships between them. The framework is a research tool intended to assist a researcher to develop awareness and understanding of the phenomena under scrutiny and to communicate this (Smyth 2004).

A review to investigate the impact of an educational intervention will have a conceptual framework that includes a hypothesis about a causal link between; who the review is about (the people), what the review is about (an intervention and what it is being compared with), and the possible consequences of intervention on the educational outcomes of these people. Such a review would follow a broadly aggregative synthesis logic. This is the shape of reviews of educational interventions carried out for the What Works Clearing House in the USA[1] and the Education Endowment Foundation in England.[2]

A review to investigate meaning or understanding of a phenomenon for the purpose of building or further developing theory will still have some prior assumptions. Thus, an initial conceptual framework will contain theoretical ideas about how the phenomena of interest can be understood and some ideas justifying why a particular population and/or context is of specific interest or relevance. Such a review is likely to follow a broadly configurative logic.

[1] https://ies.ed.gov/ncee/wwc/

[2] https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit

## 3.2 Selection Criteria

Reviewers have to make decisions about which research studies to include in their review. In order to do this systematically and transparently they develop rules about which studies can be selected into the review. Selection criteria (sometimes referred to as inclusion or exclusion criteria) create restrictions on the review. All reviews, whether systematic or not, limit in some way the studies that are considered by the review. Systematic reviews simply make these restrictions transparent and therefore consistent across studies. These selection criteria are shaped by the review question and conceptual framework. For example, a review question about the impact of homework on educational attainment would have selection criteria specifying who had to do the homework; the characteristics of the homework and the outcomes that needed to be measured. Other commonly used selection criteria include study participant characteristics; the country where the study has taken place and the language in which the study is reported. The type of research method(s) may also be used as a selection criterion but this can be controversial given the lack of consensus in education research (Newman 2008), and the inconsistent terminology used to describe education research methods.

## 3.3 Developing the Search Strategy

The search strategy is the plan for how relevant research studies will be identified. The review question and conceptual framework shape the selection criteria. The selection criteria specify the studies to be included in a review and thus are a key driver of the search strategy. A key consideration will be whether the search aims to be exhaustive i.e. aims to try and find all the primary research that has addressed the review question. Where reviews address questions about effectiveness or impact of educational interventions the issue of publication bias is a concern. Publication bias is the phenomena whereby smaller and/or studies with negative findings are less likely to be published and/or be harder to find. We may therefore inadvertently overestimate the positive effects of an educational intervention because we do not find studies with negative or smaller effects (Chow and Eckholm 2018). Where the review question is not of this type then a more specific or purposive search strategy, that may or may not evolve as the review progresses, may be appropriate. This is similar to sampling approaches in primary research. In primary research studies using aggregative approaches, such as quasi-experiments, analysis is based on the study of complete or representative samples.

In primary research studies using configurative approaches, such as ethnography, analysis is based on examining a range of instances of the phenomena in similar or different contexts.

The search strategy will detail the sources to be searched and the way in which the sources will be searched. A list of search source types is given in Box 1 below. An exhaustive search strategy would usually include all of these sources using multiple bibliographic databases. Bibliographic databases usually index academic journals and thus are an important potential source. However, in most fields, including education, relevant research is published in a range of journals which may be indexed in different bibliographic databases and thus it may be important to search multiple bibliographic databases. Furthermore, some research is published in books and an increasing amount of research is not published in academic journals or at least may not be published there first. Thus, it is important to also consider how you will find relevant research in other sources including 'unpublished' or 'grey' literature. The Internet is a valuable resource for this purpose and should be included as a source in any search strategy.

**Box 1: Search Sources**

- The World Wide Web/Internet
  - Google, Specialist Websites, Google Scholar, Microsoft Academic
- Bibliographic Databases
  - Subject specific e.g. Education—ERIC: Education Resources Information Centre
  - Generic e.g. ASSIA: Applied Social Sciences Index and Abstracts
- Handsearching of specialist journals or books
- Contacts with Experts
- Citation Checking

New, federated search engines are being developed, which search multiple sources at the same time, eliminating duplicates automatically (Tsafnat et al. 2013). Technologies, including text mining, are being used to help develop search strategies, by suggesting topics and terms on which to search—terms that reviewers may not have thought of using. Searching is also being aided by technology through the increased use (and automation) of 'citation chasing', where papers that cite, or are cited by, a relevant study are checked in case they too are relevant.

A search strategy will identify the search terms that will be used to search the bibliographic databases. Bibliographic databases usually index records

according to their topic using 'keywords' or 'controlled terms' (categories used by the database to classify papers). A comprehensive search strategy usually involves searching both a freetext search using keywords determined by the reviewers and controlled terms. An example of a bibliographic database search is given in Box 2. This search was used in a review that aimed to find studies that investigated the impact of Youth Work on positive youth outcomes (Dickson et al. 2013). The search is built using terms for the population of interest (Youth), the intervention of interest (Youth Work) and the outcomes of Interest (Positive Development). It used both keywords and controlled terms, 'wildcards' (the *sign in this database) and the Boolean operators 'OR' and 'AND' to combine terms. This example illustrates the potential complexity of bibliographic database search strings, which will usually require a process of iterative development to finalise.

**Box 2: Search string example To identify studies that address the question What is the empirical research evidence on the impact of youth work on the lives of children and young people aged 10-24 years?: CSA ERIC Database**

((TI = (adolescen* or ("young man*") or ("young men")) or TI = (("young woman*") or ("young women") or (Young adult*")) or TI = (("young person*") or ("young people*") or teen*) or AB = (adolescen* or ("young man*") or ("young men")) or AB = (("young woman*") or ("young women") or (Young adult*")) or AB = (("young person*") or ("young people*") or teen*)) or (DE = ("youth" or "adolescents" or "early adolescents" or "late adolescents" or "preadolescents"))) and(((TI = (("positive youth development ") or ("youth development") or ("youth program*")) or TI = (("youth club*") or ("youth work") or ("youth opportunit*")) or TI = (("extended school*") or ("civic engagement") or ("positive peer culture")) or TI = (("informal learning") or multicomponent or ("multi-component ")) or TI = (("multi component") or multidimensional or ("multi-dimensional ")) or TI = (("multi dimensional") or empower* or asset*) or TI = (thriv* or ("positive development") or resilienc*) or TI = (("positive activity") or ("positive activities") or experiential) or TI = (("community based") or "community-based")) or(AB = (("positive youth development ") or ("youth development") or ("youth program*")) or AB = (("youth club*") or ("youth work") or ("youth opportunit*")) or AB = (("extended school*") or ("civic engagement") or ("positive peer culture")) or AB = (("informal learning") or

multicomponent or ("multi-component ")) or AB = (("multi component") or multidimensional or ("multi-dimensional ")) or AB = (("multi dimensional") or empower* or asset*) or AB = (thriv* or ("positive development") or resilienc*) or AB = (("positive activity") or ("positive activities") or experiential) or AB = (("community based") or "community-based"))) or (DE="community education"))

Detailed guidance for finding effectiveness studies is available from the Campbell Collaboration (Kugley et al. 2015). Guidance for finding a broader range of studies has been produced by the EPPI-Centre (Brunton et al. 2017a).

## 3.4 The Study Selection Process

Studies identified by the search are subject to a process of checking (sometimes referred to as screening) to ensure they meet the selection criteria. This is usually done in two stages whereby titles and abstracts are checked first to determine whether the study is likely to be relevant and then a full copy of the paper is acquired to complete the screening exercise. The process of finding studies is not efficient. Searching bibliographic databases, for example, leads to many irrelevant studies being found which then have to be checked manually one by one to find the few relevant studies. There is increasing use of specialised software to support and in some cases, automate the selection process. Text mining, for example, can assist in selecting studies for a review (Brunton et al. 2017b). A typical text mining or machine learning process might involve humans undertaking some screening, the results of which are used to train the computer software to learn the difference between included and excluded studies and thus be able to indicate which of the remaining studies are more likely to be relevant. Such automated support may result in some errors in selection, but this may be less than the human error in manual selection (O'Mara-Eves et al. 2015).

## 3.5 Coding Studies

Once relevant studies have been selected, reviewers need to systematically identify and record the information from the study that will be used to answer the review question. This information includes the characteristics of the studies,

including details of the participants and contexts. The coding describes: (i) details of the studies to enable mapping of what research has been undertaken; (ii) how the research was undertaken to allow assessment of the quality and relevance of the studies in addressing the review question; (iii) the results of each study so that these can be synthesised to answer the review question.

The information is usually coded into a data collection system using some kind of technology that facilitates information storage and analysis (Brunton et al. 2017b) such as the EPPI-Centre's bespoke systematic review software EPPI Reviewer.[3] Decisions about which information to record will be made by the review team based on the review question and conceptual framework. For example, a systematic review about the relationship between school size and student outcomes collected data from the primary studies about each schools funding, students, teachers and school organisational structure as well as about the research methods used in the study (Newman et al. 2006). The information coded about the methods used in the research will vary depending on the type of research included and the approach that will be used to assess the quality and relevance of the studies (see the next section for further discussion of this point).

Similarly, the information recorded as 'results' of the individual studies will vary depending on the type of research that has been included and the approach to synthesis that will be used. Studies investigating the impact of educational interventions using statistical meta-analysis as a synthesis technique will require all of the data necessary to calculate effect sizes to be recorded from each study (see the section on synthesis below for further detail on this point). However, even in this type of study there will be multiple data that can be considered to be 'results' and so which data needs to be recorded from studies will need to be carefully specified so that recording is consistent across studies

## 3.6 Appraising the Quality of Studies

Methods are reinvented every time they are used to accommodate the real world of research practice (Sandelowski et al. 2012). The researcher undertaking a primary research study has attempted to design and execute a study that addresses the research question as rigorously as possible within the parameters of their

[3]https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914

resources, understanding, and context. Given the complexity of this task, the contested views about research methods and the inconsistency of research terminology, reviewers will need to make their own judgements about the quality of the any individual piece of research included in their review. From this perspective, it is evident that using a simple criteria, such as 'published in a peer reviewed journal' as a sole indicator of quality, is not likely to be an adequate basis for considering the quality and relevance of a study for a particular systematic review.

In the context of systematic reviews this assessment of quality is often referred to as Critical Appraisal (Petticrew and Roberts 2005). There is considerable variation in what is done during critical appraisal: which dimensions of study design and methods are considered; the particular issues that are considered under each dimension; the criteria used to make judgements about these issues and the cut off points used for these criteria (Oancea and Furlong 2007). There is also variation in whether the quality assessment judgement is used for excluding studies or weighting them in analysis and when in the process judgements are made.

There are broadly three elements that are considered in critical appraisal: the appropriateness of the study design in the context of the review question, the quality of the execution of the study methods and the study's relevance to the review question (Gough 2007). Distinguishing study design from execution recognises that whilst a particular design may be viewed as more appropriate for a study it also needs to be well executed to achieve the rigour or trustworthiness attributed to the design. Study relevance is achieved by the review selection criteria but assessing the degree of relevance recognises that some studies may be less relevant than others due to differences in, for example, the characteristics of the settings or the ways that variables are measured.

The assessment of study quality is a contested and much debated issue in all research fields. Many published scales are available for assessing study quality. Each incorporates criteria relevant to the research design being evaluated. Quality scales for studies investigating the impact of interventions using (quasi) experimental research designs tend to emphasis establishing descriptive causality through minimising the effects of bias (for detailed discussion of issues associated with assessing study quality in this tradition see Waddington et al. 2017). Quality scales for appraising qualitative research tend to focus on the extent to which the study is authentic in reflecting on the meaning of the data (for detailed discussion of the issues associated with assessing study quality in this tradition see Carroll and Booth 2015).

## 3.7 Synthesis

A synthesis is more than a list of findings from the included studies. It is an attempt to integrate the information from the individual studies to produce a 'better' answer to the review question than is provided by the individual studies. Each stage of the review contributes toward the synthesis and so decisions made in earlier stages of the review shape the possibilities for synthesis. All types of synthesis involve some kind of data transformation that is achieved through common analytic steps: searching for patterns in data; Checking the quality of the synthesis; Integrating data to answer the review question (Thomas et al. 2012). The techniques used to achieve these vary for different types of synthesis and may appear more or less evident as distinct steps.

Statistical meta-analysis is an aggregative synthesis approach in which the outcome results from individual studies are transformed into a standardized, scale free, common metric and combined to produce a single pooled weighted estimate of effect size and direction. There are a number of different metrics of effect size, selection of which is principally determined by the structure of outcome data in the primary studies as either continuous or dichotomous. Outcome data with a dichotomous structure can be transformed into Odds Ratios (OR), Absolute Risk Ratios (ARR) or Relative Risk Ratios (RRR) (for detailed discussion of dichotomous outcome effect sizes see Altman 1991). More commonly seen in education research, outcome data with a continuous structure can be translated into Standardised Mean Differences (SMD) (Fitz-Gibbon 1984). At its most straightforward effect size calculation is simple arithmetic. However given the variety of analysis methods used and the inconsistency of reporting in primary studies it is also possible to calculate effect sizes using more complex transformation formulae (for detailed instructions on calculating effect sizes from a wide variety of data presentations see Lipsey and Wilson 2000).

The combination of individual effect sizes uses statistical procedures in which weighting is given to the effect sizes from the individual studies based on different assumptions about the causes of variance and this requires the use of statistical software. Statistical measures of heterogeneity produced as part of the meta-analysis are used to both explore patterns in the data and to assess the quality of the synthesis (Thomas et al. 2017a).

In configurative synthesis the different kinds of text about individual studies and their results are meshed and linked to produce patterns in the data, explore different configurations of the data and to produce new synthetic accounts of the phenomena under investigation. The results from the individual studies are

translated into and across each other, searching for areas of commonality and refutation. The specific techniques used are derived from the techniques used in primary research in this tradition. They include reading and re-reading, descriptive and analytical coding, the development of themes, constant comparison, negative case analysis and iteration with theory (Thomas et al. 2017b).

## 4 Variation in Review Structures

All research requires time and resources and systematic reviews are no exception. There is always concern to use resources as efficiently as possible. For these reasons there is a continuing interest in how reviews can be carried out more quickly using fewer resources. A key issue is the basis for considering a review to be systematic. Any definitions are clearly open to interpretation. Any review can be argued to be insufficiently rigorous and explicit in method in any part of the review process. To assist reviewers in being rigorous, reporting standards and appraisal tools are being developed to assess what is required in different types of review (Lockwood and Geum Oh 2017) but these are also the subject of debate and disagreement.

In addition to the term 'systematic review' other terms are used to denote the outputs of systematic review processes. Some use the term 'scoping review' for a quick review that does not follow a fully systematic process. This term is also used by others (for example, Arksey and O'Malley 2005) to denote 'systematic maps' that describe the nature of a research field rather than synthesise findings. A 'quick review' type of scoping review may also be used as preliminary work to inform a fuller systematic review. Another term used is 'rapid evidence assessment'. This term is usually used when systematic review needs to be undertaken quickly and in order to do this the methods of review are employed in a more minimal than usual way. For example, by more limited searching. Where such 'shortcuts' are taken there may be some loss of rigour, breadth and/or depth (Abrami et al. 2010; Thomas et al. 2013).

Another development has seen the emergence of the concept of 'living reviews', which do not have a fixed end point but are updated as new relevant primary studies are produced. Many review teams hope that their review will be updated over time, but what is different about living reviews is that it is built into the system from the start as an on-going developmental process. This means that the distribution of review effort is quite different to a standard systematic review, being a continuous lower-level effort spread over a longer time period, rather than the shorter bursts of intensive effort that characterise a review with periodic updates (Elliott et al. 2014).

## 4.1 Systematic Maps and Syntheses

One potentially useful aspect of reviewing the literature systematically is that it is possible to gain an understanding of the breadth, purpose and extent of research activity about a phenomenon. Reviewers can be more informed about how research on the phenomenon has been constructed and focused. This type of reviewing is known as 'mapping' (see for example, Peersman 1996; Gough et al. 2003). The aspects of the studies that are described in a map will depend on what is of most interest to those undertaking the review. This might include information such as topic focus, conceptual approach, method, aims, authors, location and context. The boundaries and purposes of a map are determined by decisions made regarding the breadth and depth of the review, which are informed by and reflected in the review question and selection criteria.

Maps can also be a useful stage in a systematic review where study findings are synthesised as well. Most synthesis reviews implicitly or explicitly include some sort of map in that they describe the nature of the relevant studies that they have identified. An explicit map is likely to be more detailed and can be used to inform the synthesis stage of a review. It can provide more information on the individual and grouped studies and thus also provide insights to help inform choices about the focus and strategy to be used in a subsequent synthesis.

## 4.2 Mixed Methods, Mixed Research Synthesis Reviews

Where studies included in a review consist of more than one type of study design, there may also be different types of data. These different types of studies and data can be analysed together in an integrated design or segregated and analysed separately (Sandelowski et al. 2012). In a segregated design, two or more separate sub-reviews are undertaken simultaneously to address different aspects of the same review question and are then compared with one another.

Such 'mixed methods' and 'multiple component' reviews are usually necessary when there are multiple layers of review question or when one study design alone would be insufficient to answer the question(s) adequately. The reviews are usually required, to have both breadth and depth. In doing so they can investigate a greater extent of the research problem than would be the case in a more focussed single method review. As they are major undertakings, containing what would normally be considered the work of multiple systematic reviews, they are demanding of time and resources and cannot be conducted quickly.

### 4.3 Reviews of Reviews

Systematic reviews of primary research are secondary levels of research analysis. A review of reviews (sometimes called 'overviews' or 'umbrella' reviews) is a tertiary level of analysis. It is a systematic map and/or synthesis of previous reviews. The 'data' for reviews of reviews are previous reviews rather than primary research studies (see for example Newman et al. (2018). Some review of reviews use previous reviews to combine both primary research data and synthesis data. It is also possible to have hybrid review models consisting of a review of reviews and then new systematic reviews of primary studies to fill in gaps in coverage where there is not an existing review (Caird et al. 2015). Reviews of reviews can be an efficient method for examining previous research. However, this approach is still comparatively novel and questions remain about the appropriate methodology. For example, care is required when assessing the way in which the source systematic reviews identified and selected data for inclusion, assessed study quality and to assess the overlap between the individual reviews (Aromataris et al. 2015).

## 5 Other Types of Research Based Review Structures

This chapter so far has presented a process or method that is shared by many different approaches within the family of systematic review approaches, notwithstanding differences in review question and types of study that are included as evidence. This is a helpful heuristic device for designing and reading systematic reviews. However, it is the case that there are some review approaches that also claim to use a research based review approach but that do not claim to be systematic reviews and or do not conform with the description of processes that we have given above at all or in part at least.

### 5.1 Realist Synthesis Reviews

Realist synthesis is a member of the theory-based school of evaluation (Pawson 2002). This means that it is underpinned by a 'generative' understanding of causation, which holds that, to infer a causal outcome/relationship between an intervention (e.g. a training programme) and an outcome (O) of interest (e.g. unemployment), one needs to understand the underlying mechanisms (M) that connect them and the context (C) in which the relationship occurs (e.g. the characteristics of both the subjects and the programme locality). The interest of this

approach (and also of other theory driven reviews) is not simply which interventions work, but which mechanisms work in which context. Rather than identifying replications of the same intervention, the reviews adopt an investigative stance and identify different contexts within which the same underlying mechanism is operating.

Realist synthesis is concerned with hypothesising, testing and refining such context-mechanism-outcome (CMO) configurations. Based on the premise that programmes work in limited circumstances, the discovery of these conditions becomes the main task of realist synthesis. The overall intention is to first create an abstract model (based on the CMO configurations) of how and why programmes work and then to test this empirically against the research evidence. Thus, the unit of analysis in a realist synthesis is the programme mechanism, and this mechanism is the basis of the search. This means that a realist synthesis aims to identify different situations in which the same programme mechanism has been attempted. Integrative Reviewing, which is aligned to the Critical Realist tradition, follows a similar approach and methods (Jones-Devitt et al. 2017).

## 5.2 Critical Interpretive Synthesis (CIS)

Critical Interpretive Synthesis (CIS) (Dixon-Woods et al. 2006) takes a position that there is an explicit role for the 'authorial' (reviewer's) voice in the review. The approach is derived from a distinctive tradition within qualitative enquiry and draws on some of the tenets of grounded theory in order to support explicitly the process of theory generation. In practice, this is operationalised in its inductive approach to searching and to developing the review question as part of the review process, its rejection of a 'staged' approach to reviewing and embracing the concept of theoretical sampling in order to select studies for inclusion. When assessing the quality of studies CIS prioritises relevance and theoretical contribution over research methods. In particular, a critical approach to reading the literature is fundamental in terms of contextualising findings within an analysis of the research traditions or theoretical assumptions of the studies included.

## 5.3 Meta-Narrative Reviews

Meta-narrative reviews, like critical interpretative synthesis, place centre-stage the importance of understanding the literature critically and understanding differences between research studies as possibly being due to differences between their underlying research traditions (Greenhalgh et al. 2005). This means that

each piece of research is located (and, when appropriate, aggregated) within its own research tradition and the development of knowledge is traced (configured) through time and across paradigms. Rather than the individual study, the 'unit of analysis' is the unfolding 'storyline' of a research tradition over time' (Greenhalgh et al. 2005).

## 6 Conclusions

This chapter has briefly described the methods, application and different perspectives in the family of systematic review approaches. We have emphasized the many ways in which systematic reviews can vary. This variation links to different research aims and review questions. But also to the different assumptions made by reviewers. These assumptions derive from different understandings of research paradigms and methods and from the personal, political perspectives they bring to their research practice. Although there are a variety of possible types of systematic reviews, a distinction in the extent that reviews follow an aggregative or configuring synthesis logic is useful for understanding variations in review approaches and methods. It can help clarify the ways in which reviews vary in the nature of their questions, concepts, procedures, inference and impact. Systematic review approaches continue to evolve alongside critical debate about the merits of various review approaches (systematic or otherwise). So there are many ways in which educational researchers can use and engage with systematic review methods to increase knowledge and understanding in the field of education.

## References

Abrami, P. C. Borokhovski, E. Bernard, R. M. Wade, CA. Tamim, R. Persson, T. Bethel, E. C. Hanz, K. & Surkes, M. A. (2010). Issues in conducting and disseminating brief reviews of evidence. *Evidence & Policy*, *6*(3), 371–389.

Altman, D.G. (1991) *Practical statistics for medical research*. London: Chapman and Hall.

Arksey, H. & O'Malley, L. (2005). Scoping studies: towards a methodological framework, *International Journal of Social Research Methodology*, *8*(1), 19–32,

Aromataris, E. Fernandez, R. Godfrey, C. Holly, C. Khalil, H. Tungpunkom, P. (2015). Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare*, *13*.

Barnett-Page, E. & Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review, *BMC Medical Research Methodology, 9*(59), https://doi.org/10.1186/1471-2288-9-59.

Brunton, G., Stansfield, C., Caird, J. & Thomas, J. (2017a). Finding relevant studies. In D. Gough, S. Oliver & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd edition, pp. 93–132). London: Sage.

Brunton, J., Graziosi, S., & Thomas, J. (2017b). Tools and techniques for information management. In D. Gough, S. Oliver & J. Thomas (Eds.), An introduction to systematic reviews (2nd edition, pp. 154–180), London: Sage.

Carroll, C. & Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research Synthesis Methods 6*(2), 149–154.

Caird, J. Sutcliffe, K. Kwan, I. Dickson, K. & Thomas, J. (2015). Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach? *Evidence & Policy, 11*(1), 81–97.

Chow, J. & Eckholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review 30*(3), 727–744.

Dickson, K., Vigurs, C. & Newman, M. (2013). *Youth work a systematic map of the literature*. Dublin: Dept of Government Affairs.

Dixon-Woods, M., Cavers, D., Agarwa, S., Annandale, E. Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology* 6:35 https://doi.org/10.1186/1471-2288-6-35.

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C. & Gruen, R. L. (2014). Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine, 11*(2): e1001603. http://doi.org/10.1371/journal.pmed.1001603.

Fitz-Gibbon, C.T. (1984) Meta-analysis: an explication. *British Educational Research Journal*, *10*(2), 135–144.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228.

Gough, D., Thomas, J. & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews, 1(*28).

Gough, D., Kiwan, D., Sutcliffe, K., Simpson, D. & Houghton, N. (2003). *A systematic map and synthesis review of the effectiveness of personal development planning for improving student learning.* In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Gough, D., Oliver, S. & Thomas, J. (2017). Introducing systematic reviews. In D. Gough, S. Oliver & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd edition, pp. 1–18). London: Sage.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. & Peacock, R. (2005). Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. *Social Science & Medicine, 61*(2), 417–430.

Hargreaves, D. (1996). *Teaching as a research based profession: possibilities and prospects*. Teacher Training Agency Annual Lecture. Retrieved from https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2082.

Hart, C. (2018). *Doing a literature review: releasing the research imagination*. London. SAGE.

Jones-Devitt, S. Austen, L. & Parkin H. J. (2017). Integrative reviewing for exploring complex phenomena. *Social Research Update*. Issue 66.

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Klint Jørgensen, A. M., Hammerstrøm, K., & Sathe, N. (2015). *Searching for studies: A guide to information retrieval for Campbell Systematic Reviews*. Campbell Method Guides 2016:1 (http://www.campbellcollaboration.org/images/Campbell_Methods_Guides_Information_Retrieval.pdf).

Lipsey, M.W., & Wilson, D. B. (2000). *Practical meta-analysis.* Thousand Oaks: Sage.

Lockwood, C. & Geum Oh, E. (2017). Systematic reviews: guidelines, tools and checklists for authors. *Nursing & Health Sciences, 19,* 273–277.

Nelson, J. & Campbell, C. (2017). Evidence-informed practice in education: meanings and applications. *Educational Research, 59*(2), 127–135.

Newman, M., Garrett, Z., Elbourne, D., Bradley, S., Nodenc, P., Taylor, J. & West, A. (2006). Does secondary school size make a difference? A systematic review. *Educational Research Review, 1*(1), 41–60.

Newman, M. (2008). High quality randomized experimental research evidence: Necessary but not sufficient for effective education policy. *Psychology of Education Review, 32*(2), 14–16.

Newman, M., Reeves, S. & Fletcher, S. (2018). A critical analysis of evidence about the impacts of faculty development in systematic reviews: a systematic rapid evidence assessment. *Journal of Continuing Education in the Health Professions, 38*(2), 137–144.

Noblit, G. W. & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies.* Newbury Park, CA: Sage.

Oancea, A. & Furlong, J. (2007). Expressions of excellence and the assessment of applied and practice-based research. *Research Papers in Education 22*.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews 4*(1): 5. https://doi.org/10.1186/2046-4053-4-5.

Pawson, R. (2002). Evidence-based policy: the promise of "Realist Synthesis". *Evaluation, 8*(3), 340–358.

Peersman, G. (1996). *A descriptive mapping of health promotion studies in young people, EPPI Research Report.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Petticrew, M. & Roberts, H. (2005). *Systematic reviews in the social sciences: a practical guide*. London: Wiley.

Sandelowski, M., Voils, C. I., Leeman, J. & Crandell, J. L. (2012). Mapping the mixed methods-mixed research synthesis terrain. *Journal of Mixed Methods Research, 6*(4), 317–331.

Smyth, R. (2004). Exploring the usefulness of a conceptual framework as a research tool: A researcher's reflections. *Issues in Educational Research, 14*.

Thomas, J., Harden, A. & Newman, M. (2012). Synthesis: combining results systematically and appropriately. In D. Gough, S. Oliver & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 66–82). London: Sage.

Thomas, J., Newman, M. & Oliver, S. (2013). Rapid evidence assessments of research to inform social policy: taking stock and moving forward. *Evidence and Policy, 9*(1), 5–27.

Thomas, J., O'Mara-Eves, A., Kneale, D. & Shemilt, I. (2017a). Synthesis methods for combining and configuring quantitative data. In D. Gough, S. Oliver & J. Thomas (Eds.), *An Introduction to Systematic Reviews* (2nd edition, pp. 211–250). London: Sage.

Thomas, J., O'Mara-Eves, A., Harden, A. & Newman, M. (2017b). Synthesis methods for combining and configuring textual or mixed methods data. In D. Gough, S. Oliver & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd edition, pp. 181–211), London: Sage.

Tsafnat, G., Dunn, A. G., Glasziou, P. & Coiera, E. (2013). The automation of systematic reviews. *British Medical Journal, 345*(7891), doi.org/10.1136/bmj.f139.

Torgerson, C. (2003). *Systematic reviews*. London. Continuum.

Waddington, H., Aloe, A. M., Becker, B. J., Djimeu, E. W., Hombrados, J. G., Tugwell, P., Wells, G. & Reeves, B. (2017). Quasi-experimental study designs series—paper 6: risk of bias assessment. *Journal of Clinical Epidemiology, 89*, 43–52.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Reflections on the Methodological Approach of Systematic Reviews

Martyn Hammersley

## 1 Introduction

The concept of systematic reviewing of research literatures became influential in the second half of the 20th century, in the context of the longstanding, and challenging, issue of how to 'translate' research findings into reliable guidance for practical decision-making—to determine which policies, programs, and strategies should (and should not) be adopted (Hammersley 2014; Nisbet and Broadfoot 1980). The idea that research can make a significant contribution in assessing the effectiveness of policies and practices was hardly new, but it was greatly bolstered around this time by the emergence of the evidence-based medicine movement. This identified a problem with the effectiveness of many medical treatments: it was argued that research showed that some commonly used ones were ineffective, or even damaging, and that the value of a great many had never been scientifically tested; despite the fact that such testing, in the rigorous form of Randomised Controlled Trials (RCTs), was feasible. Subsequently, the idea that practice must be based on research evidence about effectiveness spread from medicine to other areas, including education.

In some countries, notably the UK, this coincided with increasing political criticism of the education system for failing to produce the levels of educational achievement required by the 'knowledge economy' and by 'international competition'. Such criticism was closely related to the rise of the 'new public management' in the 1980s, which focused on increasing the 'accountability' of

M. Hammersley (✉)
Open University UK, England, UK
e-mail: martyn.hammersley@open.ac.uk

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_2

public sector workers, including teachers, through setting targets, highlighting 'best practice', and monitoring performance (Hammersley 2000, 2013; Lane 2000). This was held to be the most effective way of 'driving up standards', and thereby improving national economic performance. In this context, it was complained not just that there was insufficient educational research of high quality relevant to key educational issues (Hargreaves 1996/2007; see also Hammersley 1997a), but also that the findings available had not been synthesised systematically so as to provide the practical guidance required. In an attempt to remedy this, not only were funds directed into increasing the amount of policy- and practice-relevant research on teaching and learning, but also into producing systematic reviews of findings relating to a wide range of educational issues (Davies 2000; Oakley et al. 2005).

In the context of medicine, systematic reviewing was usually conceived as summarising results from RCTs, via meta-analysis; and, as already noted, such trials were often regarded as the gold standard for investigations designed to determine the effectiveness of any kind of 'treatment'. However, in the 1990s relatively few RCTs had been carried out in education and therefore many of the systematic reviews produced had to rely on evidence from a wider range of research methods. One effect of this was to encourage the use of alternative ways of synthesising research findings, including ones that could be applied to findings from qualitative studies (see Barnett-Page and Thomas 2009; Dixon-Woods et al. 2005; Hammersley 2013, Chap. 11; Hannes and Macaitis 2012; Pope et al. 2007; Thomas et al. 2017). Furthermore, qualitative research began to be seen as providing a useful supplement to quantitative findings: it was believed that, while the latter indicated whether a policy or practice is effective in principle, these other kinds of evidence could offer useful contextual information, including about how the policy or practice is perceived and responded to by the people involved, which could moderate judgments about its likely effectiveness 'in the real world'. Subsequently, along with a shift towards giving a role to representatives of potential users of reviews in designing them, there was also recognition that some aspects of systematic reviews are not appropriate in relation to qualitative research, so that there came to be recognition of the need for 'integrative reviews' (Victor 2008) or 'configurative reviews' (Gough et al. 2013) as a variant of or complement to them.

Of course, 'systematic' is a laudatory label, so anything that is not systematic would generally be regarded as inadequate. Indeed, advocacy of systematic reviews often involved sharp criticism of 'traditional' or 'narrative' reviews, these being dismissed as "subjective" (Cooper 1998, p. xi), as involving "haphazard" (Slavin 1986, p. 6) or "arbitrary" (p. 10) selection procedures, as frequently summarising "highly unrepresentative samples of studies in an unsystematic and

uncritical fashion" (Petticrew and Roberts 2006, p. 5), or (even more colourfully) as amounting to "selective, opinionated and discursive rampages through the literature which the reviewer happens to know about or can easily lay his or her hands on" (Oakley 2001/2007, p. 96).[1] Given this, it is perhaps not surprising that the concept of systematic review was itself subjected to criticism by many social scientists, for example being treated as reflecting an outdated positivism (Hammersley 2013, Chap. 8; MacLure 2005; Torrance 2004). And discussions between the two sides often generated more heat than light (see, for instance, Chalmers 2003, 2005; Hammersley 2005, 2008a; Oakley 2006).

There are several problems involved in evaluating the methodological arguments for and against systematic reviews. These include the fact that, as just noted, the concept of systematic review became implicated in debates over qualitative versus quantitative method, and the philosophical assumptions these involve. Furthermore, like any other research strategy, systematic reviewing can have disadvantages, or associated dangers, as well as benefits. Equally important, it is an ensemble of components, and it is possible to accept the value of some of these without accepting the value of all. Finally, reviews can serve different functions and what is good for one may be less so for others.[2]

## 2 Criticism of Systematic Reviews

Because systematic review was associated with the evidence-based practice movement, the debates around it were closely linked with wider social and political issues. For instance, the idea that medical decisions should be determined by the results of clinical trials was challenged (not least, by advocates of 'personalised medicine'), and there was even more reaction in other fields against the notion that good professional practice is a matter of 'implementing' proven 'treatments', as against exercising professional expertise to evaluate what would be best in particular circumstances. As Torrance (2004) remarks: "Systematic reviewing can thus be seen as part of a larger discourse of distrust, of professionals and of expertise, and the increasing procedurisation of decision-making processes in risk-averse organisations" (p. 3).

[1]At other times, systematic reviewing is presented as simply one form among others, each serving different purposes (see Petticrew and Roberts 2006, p. 10).

[2]For practical guides to the production of systematic reviews, see Petticrew and Roberts (2006) and Gough et al. (2017).

It was also argued that an emphasis on 'what works' obscures the value issues involved in determining what is good policy or practice, often implicitly taking certain values as primary. Arguments about the need for evidence-based, or evidence-informed, practice resulted, it was claimed, in education being treated as the successful acquisition of some institutionally-defined body of knowledge or skill, as measured by examination or test results; whereas critics argued that it ought to be regarded as a much broader process, whether of a cognitive kind (for instance, 'learning to learn' or 'independent learning') or moral/political/religious in character' (learning to understand one's place in the world and act accordingly, to be a 'good citizen', etc.). Sometimes this sort of criticism operated at a more fundamental level, challenging the assumption that teaching is an instrumental activity (see Elliott 2004, p. 170–176). The argument was, instead, that it is a process in which values, including cognitive learning, are realised intrinsically: that they are internal goods rather than external goals. Along these lines, it was claimed that educational research of the kind assumed by systematic reviewing tends necessarily to focus on the acquisition of superficial learning, since this is what is easily measurable. In this respect systematic reviews, along with the evidence-based practice movement more generally, were criticised for helping to promote a misconceived form of education, or indeed as anti-educational.

There was also opposition to the idea, implicit in much criticism of educational research at the time when systematic reviewing was being promoted, that the primary task of this research is to evaluate the effectiveness of policies and practices. Some insisted that the main function of social and educational research is socio-political critique, while others defended a more academic conception of research on educational institutions and practices. Here, too, discussion of systematic reviewing became caught up in wider debates, this time about the proper functions of social research and, more broadly, about the role of universities in society.

While this broader background is relevant, I will focus here primarily on the specific criticisms made of systematic reviewing. These tended to come from two main sources: as already noted, one was qualitative researchers; the other was advocates of realist evaluation and synthesis (Pawson et al. 2004; Pawson 2006b). Realists argued that what is essential in evaluating any policy or practice is to identify the causal mechanism on which it is assumed to rely, and to determine whether this mechanism actually operates in the world, and if so under what conditions. Given this, the task of reviewing is not to find all relevant literature about the effects of some policy, but rather to search for studies that illuminate the causal processes assumed to be involved (Pawson 2006a; Wong 2018). Furthermore, what is important, often, is not so much the validity of the evidence but

its fruitfulness in generating and developing theoretical ideas about causal mechanisms. Indeed, while realists recognise that the validity of evidence is important when it comes to testing theories, they emphasise the partial and fallible character of all evidence, and that the search for effective causal mechanisms is an ongoing process that must take account of variation in context, since some differences in context can be crucial for whether or not a causal mechanism operates and for what it produces. As a result, realists do not recommend exhaustive searches for relevant material, or the adoption of a fixed hierarchy of evidential quality. Nor are they primarily concerned with aggregating findings, but rather with using these to develop and test hypotheses deriving from theories about particular types of policy-relevant causal process. What we have here is a completely different conception of what the purpose of reviews is from that built into most systematic reviewing.

Criticism of systematic reviewing by qualitative researchers took a rather different form. It was two-pronged. First, it was argued that systematic reviewing downplays the value of qualitative research, since the latter cannot supply what meta-analysis requires: measurements providing estimates of effect sizes. As a result, at best, it was argued, qualitative findings tend to be accorded a subordinate role in systematic reviews. A second line of criticism concerned what was taken to be the positivistic character of this type of review. One aspect of this was the demand that systematic reviewers must employ *explicit procedures* in selecting and evaluating studies. The implication is that reviews must not rely on current judgments by researchers in the field about what are the key studies, or about what is well-established knowledge; nor must they depend upon reviewers' own background expertise and judgment. Rather, a technical procedure is to be employed—one that is held to provide 'objective' evidence about the current state of knowledge. It was noted that this reflects a commitment to procedural objectivity (Newell 1986; Eisner 1992), characteristic of positivism, which assumes that subjectivity is a source of bias, and that its role can and must be minimised. Generally speaking, qualitative researchers have rejected this notion of objectivity. The contrast in orientation is perhaps most clearly indicated in the advocacy, by some, of 'interpretive reviews' (for instance, Eisenhart 1998; see Hammersley 2013, Chap. 10).

In the remainder of this chapter, I will review the distinctive methodological features of systematic reviews and evaluate them in light of these sources of criticism. I take these features to be: exhaustive searching for relevant literature; explicit selection criteria regarding relevance and validity; and synthesis of relevant findings.

## 3 Exhaustive Searching for Relevant Material

One of the criticisms that advocates of systematic review directed at traditional reviews was that they were selective in their identification of relevant literature, rather than being the product of an exhaustive search. They argued not just that, as a result, some relevant literature was not taken into account, but also that this selectivity introduced bias, analogous to sampling bias. This argument relies on a parallel being drawn with social surveys of people (see Petticrew and Roberts 2006, p. 15; Shadish 2006, p. vii).

There is not, or should not be, any disagreement about the need to make good use of previous studies in summarising existing knowledge. And this clearly requires that an effective search is carried out (see Hart 2001). Furthermore, while there is a danger of comparing systematic review as an ideal type with relatively poor actual examples of traditional reviews,[3] there is certainly a difference between the two types of review in the degree to which the search for relevant literature aims to be exhaustive. It is also probably true that the searches carried out in producing many traditional reviews missed relevant literature. Nevertheless, the demand for *exhaustive* searches is problematic.

A first point is that any simple contrast between exhaustive coverage and a biased sample is misleading, since the parallel with social surveys is open to question. At its simplest, the aim of a systematic review is to determine whether a particular type of treatment produces a particular type of effect, and this is a different enterprise from seeking to estimate the distribution of features within some population. The set of studies identified by an exhaustive search may still be a biased sample of the set of studies that could have been done, which would be the appropriate population according to this statistical line of thinking.[4] Furthermore, pooling the results from all the studies that have been done will not give us sound knowledge unless our judgments about the likely validity of the findings from each study are accurate. Increasing the size of the pool from which studies are selected does not, in itself, guarantee any increase in the likely validity of a review's findings.

[3]There are, inevitably, often failings in how systematic reviews are carried out, even in their own terms (see Petticrew and Roberts 2006, p. 270; Thompson 2015).

[4]Indeed, they may not even be a representative sample of the studies that have actually been done, as a result of publication bias: the tendency for studies that find no relationship between the variables investigated to be much less likely to be published than those that produce positive findings.

There are also practical reasons for questioning the ideal of exhaustive searching. Searching for relevant literature usually reaches a point where the value of what is still to be discovered is likely to be marginal. This is not to deny that, because of the patchiness of literatures, it is possible that material of high relevance may be found late on in a search, or missed entirely. But the point is that any attempt to eliminate this risk is not cost-free. Given finite resources, whatever time and effort are devoted to searching for relevant literature will be unavailable for other aspects of the reviewing process. For example, one criticism of systematic reviewing is that it results in superficial reading of the material found: with reviewers simply scanning for relevance, and 'extracting' the relevant information so as to assess likely validity on the basis of a checklist of criteria (MacLure 2005).[5] By contrast, qualitative researchers emphasise the need for careful reading and assessment, insisting that this is a hermeneutic task.[6] The key point here is that, as in research generally, trade-off decisions must be made regarding the time and resources allocated among the various sub-tasks of reviewing research literatures. So, rather than an insistence on maximising coverage, judgments should be made about what is the most effective allocation of time and energy to the task of searching, as against others.

There are also some questions surrounding the notion of relevance, as this is built into how a search is carried out. Where, as with many systematic reviews, the task is to find literature about the effects of a specific policy or practice, there may be a relatively well-defined boundary around what would count as relevant. By contrast, in reviews serving other functions, such as those designed to summarise the current state of knowledge in a field, this is not always the case. Here, relevance may not be a single dimension: potentially relevant material could extend in multiple directions. Furthermore, it is often far from clear where the limit of relevance lies in any of these directions. The principle of exhaustiveness is hard to apply in such contexts, even as an ideal; though, of course, the need to attain sufficient coverage of relevant literature for the purposes of the review remains. Despite these reservations, the systematic review movement has served a useful general function in giving emphasis to the importance of active searching for relevant literature, rather than relying primarily upon existing knowledge in a field.

[5]For an example of one such checklist, from the health field, see https://www.gla.ac.uk/media/media_64047_en.pdf (last accessed: 20.02.19).

[6]For an account of what is involved in understanding and assessing one particular type of research, see Hammersley (1997b).

## 4 Transparent Methodological Assessment of Studies

The second key feature of systematic reviewing is that explicit criteria should be adopted, both in determining which studies found in a search are sufficiently relevant to be included, and in assessing the likely validity of research findings. As regards relevance, clarity about how this was determined is surely a virtue in reviews. Furthermore, it is true that many traditional reviews are insufficiently clear not just about how they carried out the search for relevant literature but also about how they determined relevance. At the same time, as I noted, in some kinds of review the boundaries around relevance are complex and hard to determine, so that it may be difficult to give a very clear indication of how relevance was decided. We should also note the pragmatic constraints on providing information about this and other matters in reviews, these probably varying according to audience. As Grice (1989) pointed out in relation to communication generally, the quantity or amount of detail provided must be neither too little *nor too much*. A happy medium as regards how much information about how the review was carried out should be the aim, tailored to audience; especially given that complete 'transparency' is an unattainable ideal.

These points also apply to providing information about how the validity of research findings was assessed for the purposes of a review. But there are additional problems here. These stem partly from pressure to find a relatively quick and 'transparent' means of assessing the validity of findings, resulting in attempts to do this by identifying standard features of studies that can be treated as indicating the validity of the findings. Early on, the focus was on overall research design, and a clear hierarchy was adopted, with RCTs at the top and qualitative studies near the bottom. This was partly because, as noted earlier, qualitative studies do not produce the sort of findings required by systematic reviews; or, at least, those that employ meta-analysis. However, liberalisation of the requirements, and an increasing tendency to treat meta-analysis as only one option for synthesising findings, opened up more scope for qualitative and other non-experimental findings to be included in systematic reviews (see, for instance, Petticrew and Roberts 2006). But the issue of how the validity of these was to be assessed remained. And the tendency has been to insist that what is required is a list of specified design features that must be present if findings are to be treated as valid.

This raised particular problems for qualitative research. There have been multiple attempts to identify criteria for assessing such work that parallel those

generally held to provide a basis for assessing quantitative studies, such as internal and external validity, reliability, construct validity, and so on. But not only has there been some variation in the qualitative criteria produced, there have also been significant challenges to the very idea that assessment depends upon criteria identifying specific features of research studies (see Hammersley 2009; Smith 2004). This is not the place to rehearse the history of debates over this (see Spencer et al. 2003). The key point is that there is little consensus amongst qualitative researchers about how their work should be assessed; indeed, there is considerable variation even in judgments made about particular studies. This clearly poses a significant problem for incorporating qualitative findings into systematic reviews; though there have been attempts to do this (see Petticrew and Roberts 2006, Chap. 6), or even to produce qualitative systematic reviews (Butler et al. 2016), as well as forms of qualitative synthesis some of which parallel meta-analysis in key respects (Dixon-Woods et al. 2005; Hannes and Macaitis 2012; Sandelowski and Barroso 2007).

An underlying problem in this context is that qualitative research does not employ formalised techniques. Qualitative researchers sometimes refer to what may appear to be standard methods, such as 'thick description', 'grounded theorising', 'triangulation', and so on. However, on closer inspection, none of these terms refers to a single, standardised practice, but instead to a range of only broadly defined practices. The lack of formalisation has of course been one of the criticisms made of qualitative research. However, it is important to recognise, first of all, that what is involved here is a difference from quantitative research *in degree*, not a dichotomy. Qualitative research follows loose guidelines, albeit flexibly. And quantitative research rarely involves the mere application of standard techniques: to one degree or another, these techniques have to be adapted to the particular features of the research project concerned.

Moreover, there are good reasons why qualitative research is resistant to formalisation. The most important one is that such research relies on unstructured data, data not allocated to analytic categories at the point of collection, and is aimed at *developing* analytic categories not testing pre-determined hypotheses. It therefore tends to produce sets of categories that fall short of the requirements of mutual exclusivity and exhaustiveness required for calculating the frequencies with which data fall into one category rather than another—which are the requirements that govern many of the standard techniques used by quantitative researchers, aside from those that depend upon measurement. The looser form of categorisation employed by qualitative researchers facilitates the development of analytic ideas, and is often held to capture better the complexity of the social world. Central here is an emphasis on the role of people's interpretations and

actions in producing outcomes in contingent ways, rather than these being produced by deterministic mechanisms. It is argued that causal laws are not available, and therefore, rather than reliable predictions, the best that research can offer is enlightenment about the complex processes involved, in such a manner as to enable practitioners and policymakers themselves to draw conclusions about the situations they face and make decisions about what policies or practices it would be best to adopt. Qualitative researchers have also questioned whether the phenomena of interest in the field of education are open to counting or measurement, for example proposing thick description instead. These ideas have underpinned competing forms of educational evaluation that have long existed (for instance 'illuminative evaluation', 'qualitative evaluation' or 'case study') whose character is sharply at odds with quantitative studies (see, for instance, Parlett and Hamilton 1977). In fact, the problems with RCTs, and quantitative evaluations more generally, had already been highlighted in the late 1960s and early 1970s.

A closely related issue is the methodological diversity of qualitative research in the field of education, as elsewhere: rather than being a single enterprise, its practitioners are sharply divided not just over methods but sometimes in what they see as the very goal or product of their work. While much qualitative inquiry shares with quantitative work the aim of producing sound knowledge in answer to a set of research questions, some qualitative researchers aim at practical or political goals—improving educational practices or challenging (what are seen as) injustices—or at literary or artistic products—such as poetry, fiction, or performances of some sort (Leavy 2018). Clearly, the criteria of assessment relevant to these various enterprises are likely to differ substantially (Hammersley 2008b, 2009).

Aside from these problems specific to qualitative research, there is a more general issue regarding how research reviews can be produced for lay audiences in such a way as to enable them to evaluate and trust the findings. The ideal built into the concept of systematic review is assessment criteria that *anyone* could use successfully to determine the validity of research findings, simply by looking at the research report. However, it is doubtful that this ideal could ever be approximated, even in the case of quantitative research. For example, if a study reports random allocation to treatment and control groups, this does not tell us how successfully randomisation was achieved in practice. Similarly, while it may be reported that there was double blinding, neither participants nor researcher knowing who had been allocated to treatment and control groups, we do not know how effectively this was achieved in practice. Equally significant, neither randomisation nor blinding eliminate all threats to the validity of research

findings. My point is not to argue against the value of these techniques, simply to point out that, even in these relatively straightforward cases, statements by researchers about what methods were used do not give readers all the information needed to make sound assessments of the likely validity of a study's findings. And this problem is compounded when it comes to lay reading of reviews. Assessing the likely validity of the findings of studies and of reviews is necessarily a matter of *judgment* that will rely upon background knowledge—including about the nature of research of the relevant kinds and reviewing processes—that lay audiences may not have, and may be unable or unwilling to acquire. This is true whether the intended users of reviews are children and parents or policymakers and politicians.

That there is a problem about how to convey research findings to lay audiences is undoubtedly true. But systematic reviewing does not solve it. And, as I have indicated, there may be significant costs involved in the attempt to make reviewers' methodological assessment of findings transparent through seeking to specify explicit criteria relating to the use of standardised techniques.

## 5 Synthesis of Findings

It is important to be clear about exactly what 'synthesis' means, and also to recognise the distinction between the character or purpose of synthesis and the means employed to carry it out. At the most basic level, synthesis involves putting together findings from different studies; and, in this broad sense, many traditional as well as systematic reviews engage in this process, to some degree. However, what is involved in most systematic reviews is a very particular kind of synthesis: the production of a summary measure of the likely effect size of some intervention, based on the estimates produced by the studies reviewed. The assumption is that this is more likely to be accurate than the findings of any of individual studies because the number of cases from which data come is greater. Another significant feature of systematic reviews is that a formal and explicit method is employed, such as meta-analysis. These differences between traditional and systematic reviews raise a couple of issues.

One concerns the assumption that what is to be reviewed is a set of studies aimed at identifying the effects of a 'treatment' of some kind. Much reviewing of literature in the field of education, and in the social sciences more generally, does not deal exclusively with studies of this kind. In short, there are differences between systematic and other kinds of review as regards what is being

synthesised and for what purpose. Traditional reviews often cover a range of types of study, these not only using different methods but also aiming at different types of product. Their findings cannot be added together, but may complement one another in other ways—for example relating to different aspects of some problem, organisation, or institution. Furthermore, the aim, often, is to identify key landmarks in a field, in theoretical and/or methodological terms, or to highlight significant gaps in the literature, or questions to be addressed, rather than to determine the answer to a specific research question. Interestingly, some forms of qualitative synthesis are close to systematic review in purpose and character, while others—such as meta-ethnography—are concerned with theory development (see Noblit and Hare 1988; Toye et al. 2014).

What kind of synthesis or integration is appropriate depends upon the purpose(s) of, and audience(s) for, the particular review. As I have hinted, one of the problems with the notion of systematic reviewing is that it tends to adopt a standard model. It may well be true that for some purposes and audiences the traditional review does not engage in sufficient synthesis of findings, but this is a matter of judgment, as is what kind of synthesis is appropriate. As we saw earlier, realist evaluators argue that meta-analysis, and forms of synthesis modelled on it, may not be the most appropriate method even where the aim is to address lay audiences about what are the most effective policies or practices. They also argue that this kind of synthesis largely fails to answer more specific questions about what works for whom, when, and where—though there is, perhaps, no reason in principle why systematic reviews cannot address these questions. For realists what is required is not the synthesis of findings through a process of aggregation but rather to use previous studies in a process of theory building aimed at identifying the key causal mechanisms operating in the domain with which policymakers or practitioners are concerned. This seems to me to be a reasonable goal, and one that has scientific warrant.

Meanwhile, as noted earlier, some qualitative researchers have adopted an even more radical stance, denying the possibility of useful generalisations about sets of cases. Instead, they argue that inference should be from one (thickly described) case to another, with careful attention to the dimensions of similarity and difference, and the implications of these for what the consequences of different courses of action would be. However, while this is certainly a legitimate form of inference in which we often engage, it seems to me that it involves implicit reliance on ideas about what is likely to be generally true. It is, therefore, no substitute for generalisation.

A second issue concerns, once again, the advantages and disadvantages of standardisation or formalisation.[7] Traditional reviews tend to adopt a less standardised, and often less explicit, approach to synthesis; though the development of qualitative synthesis has involved a move towards more formal specification. Here, as with the methodological assessment of findings, it is important to recognise that exhaustive and fully transparent specification of the reviewing process is an ideal that is hard to realise, since judgment is always involved in the synthesis process. Furthermore, there are disadvantages to pursuing this ideal of formalisation *very* far, since it downgrades the important role of imagination and creativity, as well as of background knowledge and scientific sensibility. Here, as elsewhere, some assessment has to be made about the relative advantages and disadvantages of formalisation, necessarily trading these off against one another, in order to find an appropriate balance. A blanket insistence that 'the more the better', in this area as in others, is not helpful.

## 6 Conclusion

In this chapter I have outlined some of the main criticisms that have been made of systematic reviews, and looked in more specific terms at issues surrounding their key components: exhaustive searching; the use of explicit criteria to identify relevant studies and to assess the validity of findings; and synthesis of those findings. It is important to recognise just how contentious the promotion of such reviews has been, partly because of the way that this has often been done through excessive criticism of other kinds of review, and because the effect has been seen as downgrading some kinds of research, notably qualitative inquiry, at the expense of others. But systematic reviews have also been criticised because of the assumptions on which they rely, and here the criticism has come not just from qualitative researchers but also from realist evaluators.

It is important not to see these criticisms as grounds for dismissing the value of systematic reviews, even if this is the way they have sometimes been formulated. For instance, most researchers would agree that in any review an adequate search of the literature must be carried out, so that what is relevant is identified as clearly as possible; that the studies should be properly assessed in methodological

[7]For an account of the drive for standardisation, and thereby for formalisation, in the field of health care, and of many of the issues involved, see Timmermans and Berg (2003).

terms; and that this ought to be done, as far as possible, in a manner that is intelligible to readers. They might also agree that many traditional reviews in the past were not well executed. But many would insist, with Torrance (2004, p. 3), that 'perfectly reasonable arguments about the transparency of research reviews and especially criteria for inclusion/exclusion of studies, have been taken to absurd and counterproductive lengths'. Thus, disagreement remains about what constitutes adequate search for relevant literature, how studies should be assessed, what information can and ought to be provided about how a review was carried out, and what degree and kind of synthesis should take place.

The main point I have made is that reviews of research literatures serve a variety of functions and audiences, and that the form they need to take, in order to do this effectively, also varies. While being 'systematic', in the tendentious sense promoted by advocates of systematic reviewing, may serve some functions and audiences well, this will not be true of others. Certainly, any idea that there is a single standard form of review that can serve all purposes and audiences is a misconception. So, too, is any dichotomy, with exhaustiveness and transparency on one side, bias and opacity on the other. Nevertheless, advocacy of systematic reviews has had benefits. Perhaps its most important message, still largely ignored across much of social science, is that findings from single studies are likely to be misleading, and that research knowledge should be communicated to lay audiences via reviews of all the relevant literature. While I agree strongly with this, I demur from the conclusion that these reviews should always be 'systematic'.

## References

Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: a critical review, *BMC Medical Research Methodology*, 9:59. https://doi.org/10.1186/1471-2288-9-59.

Butler, A., Hall, H., & Copnell, B. (2016). A guide to writing a qualitative systematic review protocol to enhance evidence-based practice in nursing and health care. *Worldviews on Evidence-based Nursing, 13*(3), 241–249. https://doi.org/10.1111/wvn.12134.

Chalmers, I. (2003). Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations. *Annals of the American Academy of Political and Social Science*, 589, 22–40. https://doi.org/10.1177/0002716203254762.

Chalmers, I. (2005). If evidence-informed policy works in practice, does it matter if it doesn't work in theory? *Evidence and Policy*, *1*(2), 227–42. https://doi.org/10.1332/1744264053730806.

Cooper. H. (1998). *Research synthesis and meta-analysis*, Thousand Oaks CA: Sage.

Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26(3/4), 365–78. https://doi.org/10.1080/713688543.

Dixon-Woods, M., Agarwall, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Service Research and Policy*, *10*(1), 45–53. https://doi.org/10.1177/135581960501000110.

Eisenhart, M. (1998). On the subject of interpretive reviews. *Review of Educational Research*, *68*(4), 391–399. https://www.jstor.org/stable/1170731.

Eisner, E. (1992). Objectivity in educational research. *Curriculum Inquiry*, *22*(1), 9–15. https://doi.org/10.1080/03626784.1992.11075389.

Elliott, J. (2004). Making evidence-based practice educational. In G. Thomas & R. Pring (Eds.), *Evidence-Based Practice in Education* (pp. 164–186), Maidenhead: Open University Press.

Gough, D., Oliver, S., & Thomas, J. (2013). *Learning from research: systematic reviews for informing policy decisions: A quick guide*. London: Alliance for Useful Knowledge.

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd edition). London: Sage.

Grice, P. (1989). *Studies in the way of words*. Cambridge MA: Harvard University Press.

Hammersley, M. (1997a). Educational research and teaching: a response to David Hargreaves' TTA lecture. *British Educational Research Journal*, *23*(2), 141–161. https://doi.org/10.1080/0141192970230203.

Hammersley, M. (1997b). *Reading Ethnographic Research* (2nd edition). London: Longman.

Hammersley, M. (2000). Evidence-based practice in education and the contribution of educational research. In S. Reynolds & E. Trinder (Eds.), *Evidence-based practice* (pp. 163–183), Oxford: Blackwell.

Hammersley, M. (2005). Is the evidence-based practice movement doing more good than harm? *Evidence and Policy*, *1*(1), 1–16. https://doi.org/10.1332/1744264052703203.

Hammersley, M. (2008a). Paradigm war revived? On the diagnosis of resistance to randomised controlled trials and systematic review in education. *International Journal of Research and Method in Education*, *31*(1), 3–10. https://doi.org/10.1080/17437270801919826.

Hammersley, M. (2008b). Troubling criteria: a critical commentary on Furlong and Oancea's framework for assessing educational research. *British Educational Research Journal*, *34*(6), 747–762. https://doi.org/10.1080/01411920802031468.

Hammersley, M. (2009). Challenging relativism: the problem of assessment criteria. *Qualitative Inquiry*, *15*(1), 3–29. https://doi.org/10.1177/1077800408325325.

Hammersley, M. (2013). *The myth of research-based policy and practice*. London: Sage.

Hammersley, M. (2014). Translating research findings into educational policy or practice: the virtues and vices of a metaphor. *Nouveaux c@hiers de la recherche en éducation*, *17*(1), 2014, 54–74, and *Revue suisse des sciences de l'éducation*, *36*(2), 213–228.

Hannes K. & Macaitis K. (2012). A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qualitative Research 12*(4), 402–442. https://doi.org/10.1177/1468794111432992.

Hargreaves, D. H. (2007). Teaching as a research-based profession: Possibilities and prospects', Annual Lecture. In M. Hammersley (Ed.), *Educational research and evidence-based practice* (pp. 3–17). London, Sage. (Original work published 1996)

Hart, C. (2001) *Doing a literature search: A comprehensive guide for the social sciences*. London: Sage.

Lane, J. E. (2000) *New Public Management*. London: Routledge.

Leavy, P. (Ed.) (2018). *Handbook of arts-based research*. New York: Guilford Press.

MacLure, M. (2005). Clarity bordering on stupidity: where's the quality in systematic review? *Journal of Education Policy*, *20*(4), 393–416. https://doi.org/10.1080/02680930500131801.

Newell, R. W. (1986). *Objectivity, empiricism and truth*. London: Routledge and Kegan Paul.

Nisbet, J., & Broadfoot, P. (1980). *The impact of research on policy and practice in education*. Aberdeen: Aberdeen University Press.

Noblit G., & Hare R. (1988). *Meta-ethnography: synthesising qualitative studies.* Newbury Park CA: Sage.

Oakley, A. (2006). Resistances to "new" technologies of evaluation: education research in the UK as a case study. *Evidence and Policy*, *2*(1), 63–87. https://doi.org/10.1332/174426406775249741.

Oakley, A. (2007). Evidence-informed policy and practice: challenges for social science. In M. Hammersley (Ed.), *Educational research and evidence-based practice* (pp. 91–105). London: Sage. (Original work published 2001)

Oakley A., Gough, D., Oliver, S., & Thomas, J. (2005). The politics of evidence and methodology: lessons from the EPPI-Centre. *Evidence and Policy*, *1*(1), 5–31. https://doi.org/10.1332/1744264052703168.

Parlett, M.R., & Hamilton, D. (1977). Evaluation in illumination: a new approach to the study of innovative programmes. In D. Hamilton et al. (Eds.), *Beyond the numbers game* (pp. 6–22). London: Macmillan.

Pawson, R. (2006a). Digging for nuggets: how "bad" research can yield "good" evidence. *International Journal of Social Research Methodology*, *9*(2), 127–42. https://doi.org/10.1080/13645570600595314.

Pawson, R. (2006b). *Evidence-based policy: a realist perspective*. London: Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: an introduction.* Retrieved from https://www.researchgate.net/publication/228855827_Realist_Synthesis_An_Introduction (last accessed October 17, 2018).

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.

Pope, C., Mays, N., & Popay, J. (2007). Synthesising qualitative and quantitative health evidence: a guide to methods. Maidenhead: Open University Press.

Sandelowski M., & Barroso J. (2007). *Handbook for synthesising qualitative research*. New York: Springer.

Shadish, W. (2006). Foreword to Petticrew and Roberts. In M. Petticrew, & H. Roberts, *Systematic reviews in the social sciences: A practical guide* (pp. vi–ix). Oxford: Blackwell.

Slavin, R. (1986). Best-evidence synthesis: an alternative to meta-analysis and traditional reviews'. *Educational Researcher*, 15(9), 5–11.

Smith, J. K. (2004). Learning to live with relativism. In H. Piper, & I. Stronach (Eds.), *Educational research: difference and diversity* (pp. 45–58). Aldershot, UK: Ashgate.

Spencer, L., Richie, J., Lewis, J., & Dillon, L. (2003). Quality in qualitative evaluation: a framework for assessing research evidence. Prepared by the National Centre for Social

Research on behalf of the Cabinet Office. Retrieved from https://www.heacademy.ac.uk/system/files/166_policy_hub_a_quality_framework.pdf (last accessed October 17, 2018).

Thomas, J., O'Mara-Eves, A., Harden, A., & Newman, M. (2017). Synthesis methods for combining and configuring textual and mixed methods data. In D. Gough, S. Oliver, S. & J. Thomas (Eds.), *An introducation to systematic reviews* (2nd edition, pp. 181–210). London: Sage.

Thompson, C. (2015). Sins of omission and commission in systematic reviews in nursing: a commentary on McRae et al. (2015). *International Journal of Nursing Studies*, *52*(7), 1277–1278. https://doi.org/10.1016/j.ijnurstu.2015.03.003.

Timmermans, S., & Berg, M. (2003). *The gold standard: the challenge of evidence-based medicine and standardization in health care*. Philadelphia PA: Temple University Press.

Torrance, H. (2004, June). Systematic reviewing—the 'call centre' version of research synthesis. Time for a more flexible approach'. Invited presentation to Economic and Social Research Council/Research Capacity Building Network seminar on Systematic Reviewing, University of Sheffield, UK. Retrieved from http://www.esri.mmu.ac.uk/respapers/papers-pdf/seminar-systematicreviewing.pdf (last accessed October 17, 2018).

Toye F., Seers K., Allcock N, Briggs, M., Carr, E., & Barker, K. (2014). Meta-ethnography 25 years on: challenges and insights for synthesising a large number of qualitative studies. *BMC Medical Research Methodology* 14:80. https://doi.org/10.1186/1471-2288-14-80.

Victor, L. (2008). Systematic reviewing. *Social Research Update*, *54*, Retrieved from http://sru.soc.surrey.ac.uk/SRU54.pdf (last accessed October 17, 2018).

Wong, G. (2018). Data gathering in realist reviews: looking for needles in haystacks. In N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, & S. Dalkin (Eds.) *Doing realist research* (pp. 131–146). London: Sage.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Ethical Considerations of Conducting Systematic Reviews in Educational Research

Harsh Suri

Ethical considerations of conducting systematic reviews in educational research are not typically discussed explicitly. As an illustration, 'ethics' is not listed as a term in the index of the second edition of 'An Introduction to Systematic Reviews' (Gough et al. 2017). This chapter draws from my earlier in-depth discussion of this topic in the *Qualitative Research Journal* (Suri 2008) along with more recent publications by colleagues in the field of research ethics and methods of research synthesis.

Unlike primary researchers, systematic reviewers do not collect deeply personal, sensitive or confidential information from participants. Systematic reviewers use publicly accessible documents as evidence and are seldom required to seek an institutional ethics approval before commencing a systematic review. Institutional Review Boards for ethical conduct of research do not typically include guidelines for systematic reviews. Nonetheless, in the past four decades systematic reviews have evolved to become more methodologically inclusive and play a powerful role in influencing policy, practice, further research and public perception. Hence, ethical considerations of how interests of different stakeholders are represented in a research review have become critical (Franklin 1999; Hammersley 2003; Harlen and Crick 2004; Popkewitz 1999).

Educational researchers often draw upon the philosophical traditions of consequentialism, deontology or virtue ethics to situate their ethical decision-making. Consequentialism or utilitarianism focuses on maximising benefit and minimising harm by undertaking a cost-benefit analysis of potential positive

H. Suri (✉)
Deakin University, Melbourne, Australia
e-mail: harsh.suri@deakin.edu.au

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_3

and negative impacts of research on all stakeholders. Deontology or universalism stems from Immanuel Kant's logic that certain actions are inherently right or wrong and hence ends cannot justify the means. A deontological viewpoint is underpinned by rights-based theories that emphasise universal adherence to the principles of beneficence (do good), non-maleficence (prevent harm), justice, honesty and gratitude. While both consequentialism and deontology focus on actions and behaviour, virtue ethics focuses on being virtuous, especially in relationships with various stakeholders. There are several overlaps, as well as tensions, between and across these philosophical traditions (Brooks et al. 2014; Cohen et al. 2018).

Recognising the inherently situated nature of ethical decision-making, I am selectively eclectic in drawing from each of these traditions. I discuss a variety of ethical considerations of conducting systematic reviews informed by rights-based theories, ethics of care and Foucauldian ethics. Rights-based theories underpin deontology and consequentialism. Most regulatory research ethics guidelines, such as those offered by British Educational Research Association (BERA 2018) and American Educational Research Association are premised on rights-based theories that emphasises basic human rights, such as liberty, equality and dignity. Ethics of care prioritises attentiveness, responsibility, competence and responsiveness (Tronto 2005). Foucauldian ethics highlights the relationship of power and knowledge (Ball 2013).

In my earlier publications, I have identified the following three guiding principles for a quality research synthesis (Suri 2018; Suri and Clarke 2009):

- Informed subjectivity and reflexivity
- Purposefully informed selective inclusivity
- Audience-appropriate transparency

In the rest of this chapter, I will discuss how these guiding principles can support ethical decision making in systematic reviews in each of the following six phases of systematic reviews as identified in my earlier publications (Suri 2014):

1. identifying an appropriate epistemological orientation
2. identifying an appropriate purpose
3. searching for relevant literature
4. evaluating, interpreting and distilling evidence from selected reports
5. constructing connected understandings
6. communicating with an audience

To promote ethical production and use of systematic reviews through this chapter, I have used questioning as a strategic tool with the purpose of raising awareness about a variety of ethical considerations among systematic reviewers and their audience

## 1 Identifying an Appropriate Epistemological Orientation

*What philosophical traditions are amenable for guiding ethical decision-making in systematic reviews positioned along distinct epistemologies?*

Practising informed subjectivity and reflexivity, all systematic reviewers must identify an appropriate epistemological orientation, such as post-positivist, interpretive, participatory and/or critical, that is aligned with their review purpose and research competence (Suri 2013, 2018).

Deontological ethics is more relevant to post-positivist reviewers who focus on explaining, predicting or describing educational phenomena as generalisable laws expressed through relationships between measurable constructs and variables. The ethical focus of post-positivist systematic reviews tends to be on minimising threats to internal validity, external validity, internal reliability and external reliability of review findings. This is typically achieve by using a priori synthesis protocols, defining all key constructs conceptually and operationally in behavioural terms, employing exhaustive sampling strategies and employing variable oriented statistical analyses (Matt and Cook 2009; Petticrew and Roberts 2006).

Teleological ethics is more relevant to interpretive systematic reviews aiming to construct a holistic understanding of the educational phenomena that takes into account subjective experiences of diverse groups in varied contexts. Ethical decision making in interpretive systematic reviews lays an emphasis on authentically representing experiences and perceptions of diverse groups, especially those whose viewpoints tend to be less represented in the literature, to the extent that is permissible from the published literature. Maintaining a questioning gaze and a genuine engagement with diverse viewpoints, interpretive systematic reviewers focus on how individual accounts of a phenomenon reinforce, refute or augment each other (Eisenhart 1998; Noblit and Hare 1988).

Ethics of care is amenable to participatory systematic reviews that are designed to improve participant reviewers' local world experientially through critical engagement with the relevant research. Ethical decision making in participatory systematic reviews promotes building teams of practitioners with the

purpose of co-reviewing research that can transform their own practices and representations of their lived experiences. Participant co-reviewers exercise greater control throughout the review process to ensure that the review remains relevant to generating actionable knowledge for transforming their practice (Bassett and McGibbon 2013).

Foucauldian ethics is aligned with critical systematic reviews that contest dominant discourse by problematizing the prevalent metanarratives. Ethical decision making in critical systematic reviews focuses on problematizing 'what we might take for granted' (Schwandt 1998, p. 410) in a field of research by raising 'important questions about how narratives get constructed, what they mean, how they regulate particular forms of moral and social experiences, and how they presuppose and embody particular epistemological and political views of the world' (Aronowitz and Giroux 1991, pp. 80–81).

## 2 Identifying an Appropriate Purpose

*What are key ethical considerations associated with identifying an appropriate purpose for a systematic review?*

In this age of information explosion, systematic reviews require substantial resources. Guided by teleological ethics, systematic reviewers must conduct a cost-benefit analysis with a critical consideration of the purpose and scope of the review and its potential benefits to various groups of stakeholders.

If we consider the number of views or downloads as a proxy measure of impact, then we can gain useful insights by examining the teleological underpinnings of some of the highly read systematic reviews. *Review of Educational Research (RER)* tends to be regarded as the premiere educational research review journal internationally. Let us examine the scope and purpose of the three 'most read' articles in *RER,* as listed on 26 September 2018. Given the finite amount of resources available, an important question for educators is 'what interventions are likely to be most effective, and under what circumstances?'. *The power of feedback* (Hattie and Timperley 2007), with 11463 views and downloads, is a conceptual analysis primarily drawing from the findings of published systematic reviews (largely meta-analyses) conducted to address this important question. In addition to effectively teaching what is deemed important, educators also have an important role of critiquing what is deemed important and why. *The theory and practice of culturally relevant education: A synthesis of research across content areas* (Aronson and Laughter 2016), with 8958 views and downloads, is an example of

such a systematic review. After highlighting the positive outcomes of culturally relevant education, the authors problematise the validity of standardised testing as an unbiased form of a desirable educational outcome for all. As education is essentially a social phenomenon, understanding how different stakeholders perceive various configurations of an educational intervention is critical. *Making sense of assessment feedback in higher education* (Evans 2013), with 5372 views and downloads, is an example of a systematic review that follows such a pursuit. Even though each of these reviews required significant resources and expertise, the cost is justified by the benefits evident from the high number of views and downloads of these articles. Each of these three reviews makes clear recommendations for practitioners and researchers by providing an overview, as well as interrogating, current practices.

All educational researchers are expected to prevent, or disclose and manage, ethical dilemmas arising from any real or perceived conflicts of interest (AERA 2011; BERA 2018). Systematic reviewers should also carefully scrutinise how their personal, professional or financial interests may influence the review findings in a specific direction. As systematic reviews require significant effort and resources, it is logical for systematic reviewers to bid for funding. Recognising the influence of systematic reviews in shaping perceptions of the wider community, many profit and not profit organisations have become open to funding systematic reviews. Before accepting funding for conducting a systematic review, educational researchers must carefully reflect on the following questions:

- How does the agenda of the funding source intersect with the purpose of the review?
- How might this potentially influence the review process and findings? How will this be managed ethically to ensure integrity of the systematic review findings?

In case of sponsored systematic reviews, it is important to consider at the outset how potential ethical issues will be managed if the interest of the funding agency conflicts with the interests of relatively less influential or less represented groups. Systematic reviews funded by a single agency with a vested interest in the findings are particularly vulnerable to ethical dilemmas arising from a conflict of interest (The Methods Coordinating Group of the Campbell Collaboration 2017). One approach could be to seek funding from a combination of agencies representing interests of different stakeholder groups. Exploring the option of crowdfunding is another option that systematic reviewers could use to represent

the interests of marginalised groups whose interests are typically overlooked in the agenda of powerful funding agencies. In participatory synthesis, it is critical that the purpose of the systematic review evolves organically in response to the emerging needs of the practitioner participant reviewers.

## 3 Searching for Relevant Literature

*What are key ethical considerations associated with developing an appropriate strategy for sampling and searching relevant primary research reports to include in a systematic review?*

A number of researchers in education and health sciences have found that studies with certain methodological orientations or types of findings are more likely to be funded, published, cited and retrieved through common search channels (Petticrew and Roberts 2006). Serious ethical implications arise when systematic reviews of biased research are drawn upon to make policy decisions with an assumption that review findings are representative of the larger population. In designing an appropriate sampling and search strategy, systematic reviewers should carefully consider the impact of potential publication biases and search biases.

Funding bias, methodological bias, outcome bias and confirmatory bias are common forms of publication bias in educational research. For instance, studies with large sample-sizes are more likely to attract research funding, being submitted for publishing and getting published in reputable journals (Finfgeld-Connett and Johnson 2012). Research that reports significantly positive effects of an innovative intervention is more likely to be submitted for publishing by primary researchers and being accepted for publishing by journal editors (Dixon-Woods 2011; Rothstein et al. 2004). Rather than reporting on all the comparisons made in a study, often authors report on only those comparisons that are significant (Sutton 2009). As a result, the effectiveness of innovative educational interventions gets spuriously inflated in published literature. Often, when an educational intervention is piloted, additional resources are allocated for staff capacity building. However, in real life when the same intervention is rolled out at scale, the same degree of support is not provided to teachers whose practice is impacted by the intervention (Schoenfeld 2006).

Even after getting published, certain types of studies are more likely to be cited and retrieved through common search channels, such as key databases and professional networks (Petticrew and Roberts 2006). Systematic reviewers must carefully consider common forms of search biases, such as database bias, citation

bias, availability bias, language bias, country bias, familiarity bias and multiple publication bias. The term 'grey literature' is sometimes used to refer to published and unpublished reports, such as government reports, that are not typically included in common research indexes and databases (Rothstein and Hopewell 2009). Several scholars recommend inclusion of grey literature to minimise potential impact of publication bias and search bias (Glass 2000) and to be inclusive of key policy documents and government reports (Godin et al. 2015). On the other hand, several other scholars argue that systematic reviewers should include only published research that has undergone the peer-review process of academic community to include only high-quality research and to minimise the potential impact of multiple publications based on the same dataset (La Paro and Pianta 2000).

With the ease of internet publishing and searching, the distinction between published and unpublished research has become blurred and the term grey literature has varied connotations. While most systematic reviews employ exhaustive sampling, in recent years there has been an increasing uptake of purposeful sampling in systematic reviews as evident from more than 1055 Google Scholar citations of a publication on this topic: *Purposeful sampling in qualitative research synthesis* (Suri 2011).

Aligned with the review's epistemological and teleological positioning, all systematic reviewers must prudently design a sampling strategy and search plan, with complementary sources, that will give them access to most relevant primary research from a variety of high-quality sources that is inclusive of diverse viewpoints. They must ethically consider positioning of the research studies included in their sample in relation to the diverse contextual configurations and viewpoints commonly observed in practical settings.

## 4 Evaluating, Interpreting and Distilling Evidence from the Selected Research Reports

*What are key ethical considerations associated with evaluating, interpreting and distilling evidence from the selected research reports in a systematic review?*

Systematic reviewers typically do not have direct access to participants of primary research studies included in their review. The information they analyse is inevitably refracted through the subjective lens of authors of individual studies. It is important for systematic reviewers to critically reflect upon contextual position of the authors of primary research studies included in the review, their methodological and pedagogical orientations, assumptions they are making, and how they

might have influenced the findings of the original studies. This becomes particularly important with global access to information where critical contextual information, that is common practice in a particular context but not necessarily in other contexts, may be taken-for-granted by the authors of the primary research report and hence may not get explicitly mentioned.

Systematic reviewers must ethically consider the quality and relevance of evidence reported in primary research reports with respect to the review purpose (Major and Savin-Baden 2010). In evaluating quality of evidence in individual reports, it is important to use the evaluation criteria that are commensurate with the epistemological positioning of the author of the study. Cook and Campbell's (1979) constructs of internal validity, construct validity, external validity and statistical conclusion are amenable for evaluating postpositivist research. Valentine (2009) provides a comprehensive discussion of criteria suitable for evaluating research employing a wide range of postpositivist methods. Lincoln and Guba's (1985) constructs of credibility, transferability, dependability and confirmability are suitable for evaluating interpretive research. The Centre for Reviews and Dissemination (CRD 2009) provides a useful comparison of common qualitative research appraisal tools in Chap. 6 of its open access guidelines for systematic reviews. Herons and Reason's (1997) constructs of critical subjectivity, epistemic participation and political participation emphasising a congruence of experiential, presentational, propositional, and practical knowings are appropriate for evaluating participatory research studies. Validity of transgression, rather than correspondence, is suitable for evaluating critically oriented research reports using Lather's constructs of ironic validity, paralogical validity, rhizomatic validity and voluptuous validity (Lather 1993). Rather than seeking perfect studies, systematic reviewers must ethically evaluate the extent to which findings reported in individual studies are grounded in the reported evidence.

While interpreting evidence from individual research reports, systematic reviewers should be cognisant of the quality criteria that are commensurate with the epistemological positioning of the original study. It is important to ethically reflect on plausible reasons for critical information that may be missing from individual reports and how might that influence the report findings (Dunkin 1996). Through purposefully informed selective inclusivity, systematic reviewers must distil information that is most relevant for addressing the synthesis purpose.

Often a two-stage approach is appropriate for evaluating, interpreting and distilling evidence from individual studies. For example, in their review that won the *American Educational Research Association's Review of the Year Award*, Wideen et al. (1998) first evaluated individual studies using the criteria aligned with the methodological orientation of individual studies. Then, they distilled information that was most relevant for addressing their review purpose. In this

phase, systematic reviewers must ethically pay particular attention to the quality criteria that are aligned with the overarching methodological orientation of their review, including some of the following criteria: reducing any potential biases, honouring representations of the participants of primary research studies, enriching praxis of participant reviewers or constructing a critically reflexive account of how certain discourses of an educational phenomenon have become more powerful than others. The overarching orientation and purpose of the systematic review should influence the extent to which evidence from individual primary research studies is drawn upon in a systematic review to shape the review findings (Major and Savin-Baden 2010; Suri 2018).

## 5 Constructing Connected Understandings

*What are key ethical considerations associated with constructing connected understandings in a systematic review?*

Through informed subjectivity and reflexivity, systematic reviewers must ethically consider how their own contextual positioning is influencing the connected understandings they are constructing from the distilled evidence. A variety of systematic techniques can be used to minimise unacknowledged biases, such as content analysis, statistical techniques, historical methods, visual displays, narrative methods, critical sensibilities and computer-based techniques. Common strategies for enhancing quality of all systematic reviews include 'reflexivity; collaborative sense-making; eliciting feedback from key stakeholders; identifying disconfirming cases and exploring rival connections; sensitivity analyses and using multiple lenses' (Suri 2014, p. 144).

In addition, systematic reviewers must pay specific attention to ethical considerations particularly relevant to their review's epistemological orientation. For instance, all post-positivist systematic reviewers should be wary of the following types of common errors: unexplained selectivity, not discriminating between evidence of varying quality, inaccurate coding of contextual factors, overstating claims made in the review beyond what can be justified by the evidence reported in primary studies and not paying adequate attention to the findings that are at odds with the generalisations made in the review (Dunkin 1996). Interpretive systematic reviews should focus on ensuring authentic representation of the viewpoints of the participants of the original studies as expressed through the interpretive lens of the authors of those studies. Rather than aiming for generalisability of the findings, they should aim at transferability by focusing on how the findings of individual studies intersect with their methodological and contextual configurations. Ethical considerations in participatory systematic reviews should pay attention to

the extent to which practitioner co-reviewers feel empowered to drive the agenda of the review to address their own questions, change their own practices through the learning afforded by participating in the experience of the synthesis and have practitioner voices heard through the review (Suri 2014). Critically oriented systematic reviews should highlight how certain representations silence or privilege some discourses over the others and how they intersect with the interests of various stakeholder groups (Baker 1999; Lather 1999; Livingston 1999).

## 6 Communicating with an Audience

*What are key ethical considerations associated with communicating findings of a systematic review to diverse audiences?*

All educational researchers are expected to adhere to the highest standards of quality and rigour (AERA 2011; BERA 2018). The PRISMA-P group have identified a list of 'Preferred reporting items for systematic review and meta-analysis protocols' (Moher et al. 2015) which are useful guidelines to improve the transparency of the process in systematic reviews. Like all educational researchers, systematic reviewers also have an obligation to disclose any sources of funding and potential conflicts of interest that could have influenced their findings.

All researchers should reflexively engage with issues that may impact on individuals participating in the research as well as the wider groups whose interests are intended to be addressed through their research (Greenwood 2016; Pullman and Wang 2001; Tolich and Fitzgerald 2006). Systematic reviewers should also critically consider the potential impact of the review findings on the participants of original studies and the wider groups whose practices or experiences are likely to be impacted by the review findings. They should carefully articulate the domain of applicability of a review to deter the extrapolation of the review findings beyond their intended use. Contextual configurations of typical primary research studies included in the review must be comprehensively and succinctly described in a way that contextual configurations missing from their sample of studies become visible.

## 7 Summary

Like primary researchers, systematic reviewers should reflexively engage with a variety of ethical issues associated that potential conflicts of interest and issues of voice and representation. Systematic reviews are frequently read and cited in

documents that influence educational policy and practice. Hence, ethical issues associated with what and how systematic reviews are produced and used have serious implications. Systematic reviewers must pay careful attention to how perspectives of authors and research participants of original studies are represented in a way that makes the missing perspectives visible. Domain of applicability of systematic reviews should be scrutinised to deter unintended extrapolation of review findings to contexts where they are not applicable. This necessitates that they systematically reflect upon how various publication biases and search biases may influence the synthesis findings. Throughout the review process, they must remain reflexive about how their own subjective positioning is influencing, and being influenced, by the review findings. Purposefully informed selective inclusivity should guide critical decisions in the review process. In communicating the insights gained through the review, they must ensure audience-appropriate transparency to maximise an ethical impact of the review findings.

## References

AERA. (2011). Code of ethics (Approved by Amercian Educational Research Association Council February 2011). *Educational Researcher, 40*(3), 145–156.

Aronowitz, S., & Giroux, H. A. (1991). *Postmodern education: Politics, culture, and social criticism*. Minneapolis: University of Minnesota Press.

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research, 86*(1), 163–206.

Baker, B. (1999). What is voice? Issues of identity and representation in the framing of reviews. *Review of Educational Research, 69*(4), 365–383.

Ball, S. (2013). *Foucault, power and education*. NY: Routledge.

Bassett, R., & McGibbon, E. (2013). A critical participatory and collaborative method for scoping the literature. *Quality and quantity, 47*(6), 3249–3259.

BERA. (2018). *British Educational Research Associations' ethical guidelines for educational research*. Retrieved 4 April 2018 from https://www.bera.ac.uk/wp-content/uploads/2018/06/BERA-Ethical-Guidelines-for-Educational-Research_4thEdn_2018.pdf?noredirect=1.

Brooks, R., Kitty, t. R., & Maguire, M. (2014). *Ethics and Education Research*. London: Sage.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.). Abingdon: Routledge.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.

CRD. (2009). Systematic reviews: CRD's guidance for undertaking reviews in health care. Retrieved April 10, 2019 from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.

Dixon-woods, M. (2011). Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine, 9*(3), 39–40.

Dunkin, M. J. (1996). Types of errors in synthesizing research in education. *Review of Educational Research, 66*(2), 87–97.

Eisenhart, M. (1998). On the subject of interpretive reviews. *Review of Educational Research, 68*(4), 391–399.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120. https://doi.org/10.3102/0034654312474350.

Finfgeld-Connett, D., & Johnson, E. D. (2012). Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews. *Journal of Advanced Nursing, 69*(1), 194–203. https://doi.org/10.1111/j.1365-2648.2012.06037.x.

Franklin, B. M. (1999). Discourse, rationality and educational research: A historical perspective of RER. *Review of Educational Research, 69*(4), 347–363.

Glass, G. V. (2000, January). Meta-analysis at 25. Retrieved April 10, 2019, from http://www.gvglass.info/papers/meta25.html.

Godin, K., Stapleton, J., Kirkpatrick, S. I., Hanning, R. M., & Leatherdale, S. T. (2015). Applying systematic review search methods to the grey literature: A case study examining guidelines for school-based breakfast programs in Canada. *Systematic Reviews, 4*(1), 138. https://doi.org/10.1186/s13643-015-0125.

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2017). *An introduction to systematic reviews* (2nd ed.). London: Sage.

Greenwood, M. (2016). Approving or improving research ethics in management journals. *Journal of Business Ethics, 137*, 507–520.

Hammersley, M. (2003). Systematic or unsystematic, is that the question? Some reflections on the science, art, and politics of reviewing research. Paper presented at the Department of Epidemiology and Public Health, University of Leicester.

Harlen, W., & Crick, R. D. (2004). Opportunities and challenges of using systematic reviews of research for evidence-based policy in education. *Evaluation and Research in Education, 18*(1–2), 54–71.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–113.

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry, 3*(3), 274–294.

La Paro, K., & Pianta, R. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*(4), 443–484.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *The Sociological Quaterly, 34*(4), 673–693.

Lather, P. (1999). To be of use: The work of reviewing. *Review of Educational Research, 69*(1), 2–7.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Livingston, G. (1999). Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research, 69*(1), 9–19.

Major, C. H., & Savin-Baden, M. (2010). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*. London: Routledge.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 537–560). New York: Sage.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1–9).

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park: Sage.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.

Popkewitz, T. S. (1999). Reviewing reviews: *RER*, research and the politics of educational knowledge. *Review of Educational Research, 69*(4), 397–404.

Pullman, D., & Wang, A. T. (2001). Adaptive designs, informed consent, and the ethics of research. *Controlled Clinical Trials, 22*(3), 203–210.

Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–125). New York: Sage.

Rothstein, H. R., College, B., Turner, H. M., & Lavenberg, J. G. (2004). *The Campbell Collaboration information retrieval policy brief*. Retrieved 2006, June 05, from http://www.campbellcollaboration.org/MG/IRMGPolicyBriefRevised.pdf.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 13–21.

Schwandt, T. A. (1998). The interpretive review of educational matters: Is there any other kind? *Review of Educational Research, 68*(4), 409–412.

Suri, H. (2008). Ethical considerations in synthesising research: Whose representations? *Qualitative Research Journal, 8*(1), 62–73.

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal, 11*(2), 63–75.

Suri, H. (2013). Epistemological pluralism in qualitative research synthesis. *International Journal of Qualitative Studies in Education 26*(7), 889–911.

Suri, H. (2014). *Towards methodologically inclusive research synthesis*. UK: Routledge.

Suri, H. (2018). 'Meta-analysis, systematic reviews and research syntheses' In L. Cohen, L. Manion & K. R. B. Morrison *Research Methods in Education* (8th ed., pp. 427–439). Abingdon: Routledge.

Suri, H., & Clarke, D. J. (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research, 79*(1), 395–430.

Sutton, A. J. (2009). Publication bias. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435–452). New York: Sage.

The Methods Coordinating Group of the Campbell Collaboration. (2017). Methodological expectations of Campbell Collaboration intervention reviews: Reporting standards. Retrieved April 21, 2019, from https://www.campbellcollaboration.org/library/campbell-methods-reporting-standards.html.

Tolich, M., & Fitzgerald, M. (2006). If ethics committees were designed for ethnography. Journal of Empirical Research on Human Research Ethics. *Journal of Empirical Research on Human Research Ethics, 12*(2), 71–78.

Tronto, J. C. (2005). An ethic of care. In A. E. Cudd & R. O. Andreasen (Eds.), *Feminist theory: A philosophical anthology* (pp. 251–263). Oxford, UK Malden, Massachusetts: Blackwell.

Valentine, J. C. (2009). Judging the quality of primary research. In H. M. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 129–146). New York: Sage.

Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research, 68*(2), 130–178.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Teaching Systematic Review

Melanie Nind

## 1 Introduction

I last wrote about systematic review more than a decade ago when, having been immersed in conducting three systematic reviews for the Teacher Training Agency in England, I felt the need to reflect on the process. Writing a reflexive narrative (Nind 2006) was a mechanism for me to think through the value of getting involved in systematic review in education when there were huge questions being asked of the relevance of evidence-based practice (EBP) for education (e.g. Hammersley 2004; Pring 2004). Additionally, critics of systematic review from education were making important contributions to the debate about the method itself, with Hammersley (2001) questioning its positivist assumptions and MacLure (2005) focusing on what she proposed was the inherent reduction of complexity to simplicity involved, the degrading of reading and interpreting into something quite different to disinter "tiny dead bodies of knowledge" (p. 394). I concluded then that while the privileging of certain kinds of studies within systematic review could be problematic, systematic reviews themselves produced certain kinds of knowledge which had value. My view was that the things systematic reviews were accused of—over-simplicity, failing to look openly or deeply—were not inevitable. My defence of the method lay not just in my experience of using it, but in the way in which I was taught about it—and how to conduct it—and led to a longer term interest in the teaching of research methods.

M. Nind (✉)
Southampton Education School, University of Southampton, Southampton, UK
e-mail: M.A.Nind@soton.ac.uk

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_4

At the time of writing this chapter I have just concluded a study of the *Pedagogy of Methodological Learning* for the *National Centre for Research Methods* in the UK (see http://pedagogy.ncrm.ac.uk/). This explored in some depth how research methods are taught and learned and teased out the pedagogical content knowledge (Shulman 1987) held by methods teachers and their often implicit craft knowledge (Brown and MacIntyre 1993). In the study we sought to engage teachers and learners as stakeholders in the process of building capability and capacity in the co-construction of understandings of what is important in teaching and learning advanced social science research methods (Nind and Lewthwaite 2018a). This included among others, teachers and learners from the discipline of education and teachers and learners of the method of systematic review.

This chapter about teaching systematic review combines and builds on insights from these two sets of research experiences. To clarify, any guidance included here is not the product of systematic review but of deep engagement with systematic review and with the teaching of social research methods including systematic review. To conduct a systematic review on this topic in order to transparently assemble, critically appraise and synthesise the available studies would necessitate there being a body of research in the area to systematically trawl through, which there is not. This is partly because, as colleagues and I have argued elsewhere (Kilburn et al. 2014; Lewthwaite and Nind 2016), the pedagogic culture around research methods is under-developed, and partly because EBP is not as dominant in education as it is in medicine and health professions. If we are teaching systematic review to education researchers we do not have the option of identifying best evidence to bring to bear on the specific challenge. However, the pedagogy of research methods is a nascent field; interest in it is gathering momentum, stimulated in part by reviews of the literature that I discuss next, identification of the need for pedagogic research to inform capacity building strategy (Nind et al. 2015) and new research purposefully designed to develop the pedagogic culture (Lewthwaite and Nind 2016; Nind and Lewthwaite 2018a, 2018b).

## 2 Contribution of Systematic Reviews and Other Literature Reviews

Wagner et al. (2011) took a broad look at the topics covered in the literature on teaching social science research methods, reviewing 195 journal articles from the decade 1997–2007. These were identified through:

> a database search of the Social Sciences Citation Index, ScienceDirect, Academic Search Premier, EBSCOhost, PsycINFO, Swetswise and Google Scholar. The keywords research, teaching, training, methodology, methods, pedagogy, social sciences, higher education and curriculum were used in various combinations to search the databases … [plus] examining the reference lists of the accumulated material for additional sources, until a point of saturation had been reached (Wagner et al. 2011, p. 76).

No papers on teaching systematic review were identified. Their review "proceeded according to Thody's (2006) five steps: recording, summarising, integrating, analysing and criticising sources" (Wagner et al. 2011, p. 78). From this they concluded that when it comes to teaching research methods there has been little debate in the literature, little cross-citation and limited empirical research.

Cooper et al. (2012) conducted a meta-study with a related focus, looking at thirty years of primary research on the experiences of students learning qualitative research methods. Their concerns were with learning from the past, not just about the students' experience but about the research methods used to study them. Hence, their meta-study included:

> a meta-method analysis of the methodologies and procedures used in the previous published primary research sources; a meta-theory analysis of the theoretical frameworks and conceptualization utilized in the previous published primary research sources; and a meta-synthesis of the results from the meta-data-analysis, meta-method analysis, and the meta-theory analysis to determine patterns between the results produced, the methodologies employed, and the theoretical orientations engaged (Cooper et al. 2012, p. 2).

While retaining a qualitative constructivist grounded theory approach in the analysis, the authors were influenced by the observation by Littell et al. (2008) of the increasing use of systematic review in education (and other social science) research. Their search focused on the Teaching and Learning Qualitative Research and Qualitative Research Design Resources database, ProQuest, ERIC, and Google Scholar with some hand-searching. This led them to identify 25 published articles providing the student perspective. Papers were appraised using a modification of the Primary Research Appraisal Tool (Paterson et al. 2001). They conclude "that the student experience of learning qualitative research is made up of three central dimensions—experiential, affective, and cognitive—which combine to form an experience of active learning necessary to understand and practice qualitative research" (pp. 6–7).

Next up, Earley (2014) undertook a synthesis of 89 studies (1987 to 2012) pertaining to social science research methods education (search terms and databases unspecified), asking

(1) What does the current literature tell us about teaching research methods?
(2) What does the current literature tell us about the learning of research methods?
(3) What gaps are there in the current literature related to teaching and learning research methods?
(4) What suggestions for further research can be identified through an exploration of the current literature on teaching and learning research methods? (Earley 2014, p. 243)

He followed Cooper's (1998) five stages for conducting a research synthesis (problem formulation, literature search, assessment of the quality and applicability of the studies, analysis and interpretation, and presenting the results). Earley (2014) was able to show patterns in the research in how learners are characterised (largely unmotivated and nervous), teaching techniques covered (active learning, problem-based learning, cooperative learning, service learning, experiential learning and online learning), and teacher objectives (concerned with educating consumers or producers of research). More importantly perhaps, he identified problems that have been ongoing and that our *Pedagogy of Methodological Learning* study sought to address: unfulfilled need to establish what student learning of social research methods looks like and the literature being dominated by teacher reflections on their own classrooms rather than studies that cross contextual boundaries or look from the outside in.

As a bridge between previous reviews and new empirical work, my colleagues and I conducted a new literature review (Kilburn et al. 2014), purposefully constructed in terms of deep reading of the literature as opposed to systematic review. We engaged in thematic qualitative exploration of insights into how methods teachers approach their craft. We sought to identify all peer-reviewed outputs on the teaching and learning of social research methods, focusing on the endpoint for the Wagner et al. synthesis in 2007 through to 2013. We searched the ISI Web of Knowledge database and for the 'high sensitivity' search (Barnett-Page and Thomas 2009) used the search terms: "research methods" OR "methodology" OR "qualitative" OR "quantitative" OR "mixed methods" AND "teaching" OR "learning" OR "education" OR "training" OR "capacity building". This led to sifting over 800 titles, moving to a potential pool of 66 papers and examination of 24 papers. As with Earley (2014), we found that most of the papers reported on teachers' reflections on their practice and there was an emphasis on active and experiential learning. However, we also found greater "cause for optimism

regarding the state of pedagogical practice and enquiry relating to social science research methods" in that "considerable attention is being paid to the ways in which teaching and learning is structured, delivered and facilitated" and "methods teachers are innovating and experimenting" in response to identified limitations in pedagogic practice and "developing conceptually or theoretically useful frames of reference" (p. 204).

The state of the research literature indicates a willingness among methods teachers to systematically reflect on their own practice, thereby making some connection with pedagogic theory, but that there is limited engagement with the practice of other methods teachers working in other disciplines or with other methods. It is noteworthy that none of the above searches turned up papers about teaching systematic review specifically. This situation may be indicative of the way in which education (and certainly not higher education (Bearman et al. 2012)) is not an evidence-based profession in the way that Hargreaves (1996) and Goldacre (2013) have argued it should be. If teachers of methods are relying on their own professional judgement (or trial-and-error as Earley (2014) argues), the knowledge of the team and feedback from their students, it may be that they do not feel the need to draw on a pool of wider evidence. They may be rejecting the "calls for more scientific research" and "reliable evidence regarding efficacy in education systems and practices" that Thomas (2012, p. 26) discusses when he argues that in education, "Our landscape of inquiry exists not at the level of these big 'what works' questions but at the level of personalized questions posed locally. It exists in the dynamic of teachers' work, in everyday judgments" (p. 41). This disjuncture with systematic review principles poses real and distinctive challenges for teachers of systematic review method in education, as I shall go on to show.

Before moving on from the contribution of systematic reviews to our understanding of how to teach them we should note the systematic reviews conducted pertaining to educating medicine and health professionals about evidence-based practice. Coomarasamy and Khan (2004) synthesised 23 studies, including four randomised trials, looking at the outcome measures of knowledge, critical appraisal skills, attitudes, and behaviour in medicine students taught EBP. They concluded that standalone teaching improved knowledge but not skills, attitudes, or behaviour, whereas clinically integrated teaching improved knowledge, skills, attitudes and behaviour. This led them to recommend that the "teaching of evidence based medicine should be moved from classrooms to clinical practice to achieve improvements in substantial outcomes" (p. 1). Kyriakoulis et al. (2016) similarly used systematic review to find the best teaching strategies for teaching EBP to undergraduate health students. The studies included in their review evalu-

ated pedagogical formats for their impact on EBP skills. They found "little robust evidence" (p. 8) to guide them, only that multiple interventions combining lectures, computer sessions, small group discussions, journal clubs, and assignments were more likely to improve knowledge, skills, and attitude than single interventions or no interventions. This and other meta-studies serve to highlight the need for new research and, I argue, more work at the open, exploratory stage to understand pedagogy in action.

## 3 The Pedagogy of Methodological Learning Study

The *Pedagogy of Methodological Learning* study was in large part my response to a policy demand for methods training to build capacity among social science researchers that was not yet recognising the contribution that pedagogic research could make, and to the limitations in the scope of the research to date. It was designed to find and share the pedagogical content knowledge of social science research methods teachers and to be conducted in a collaborative, non-judgemental spirit so that together we could better understand and develop our pedagogic practices. The study comprised a series of connected parts:

- an international expert panel to explore—both individually and collectively—the practices and pedagogical content knowledge of methods teachers with extensive teaching experience, followed up with seven focus groups with methods teachers in the UK to further the insights and check the resonance of core themes from analysis of the experts' data;
- video stimulated recall, reflection and dialogue between teachers and learners of various social research methods in a series of focus groups to reflect on pedagogical decision-making and experience of research methods pedagogy in action;
- a methods learning diary circle to access and explore together a range of learner perspectives on their methods learning journeys over an extended period;
- in-depth case studies to add nuanced detail and test the emerging typology of pedagogic practice in situ.

The methods are discussed elsewhere, including their role in offering pedagogic leadership (Lewthwaite and Nind 2016) and in supporting pedagogic culture-building and dialogue (Nind and Lewthwaite 2018a). In this chapter I discuss the findings for the light they can shed on the teaching and learning of systematic

review in the field of education. I draw in particular on video stimulated dialogue about the teaching of synthesis methods within systematic review.

The *Pedagogy of Methodological Learning* study has identified that the participating methods teachers have particular pedagogical content knowledge about how to teach with, through and about data, including the affordances of learner data and teacher data, and the value of authentic data, immersion in data and actively doing things with data. Teachers of qualitative methods understand that their work involves conceptually difficult material, which requires them to have deep knowledge of qualitative research and to foster reflexivity in their classrooms. They value and use authentic data and their own and learners' standpoints in their teaching. Teachers of quantitative methods stress the teaching of technical skills, the necessary logic to make sound judgements and the role of actively practising on data. They understand that their work requires an understanding of the difficulty and sequencing of content and they use diverse strategies and tactics including chunking, bootstrapping, backfilling and scaffolding to convey knowledge, build competence and deepen learning (Nind and Lewthwaite 2018b). There is a recurrent narrative about underprepared, fearful, diverse and anxious quantitative methods students leading teachers to develop student-centred approaches that deploy visual or verbal non-technical strategies to support learning. Teachers of mixed methods understand the particularly challenging nature of supporting learners in going back and forth between deductive and inductive thinking and thinking critically as well as pragmatically.

Some participating methods experts and teachers struggled to articulate their pedagogic approach, some readily identified with a known, named pedagogic approach, and some articulated and named their own unique approach. They described using active learning, experiential learning, student-centred learning, peer/interactive/collaborative/dialogical learning, problem-based learning and independent learning approaches. The teaching of qualitative methods was associated with experiential learning approaches and the teaching of quantitative methods had a notable lack of collaborative approaches. Teachers in the study identified a range of conscious pedagogic strategies for structuring content, organizing the classroom and engaging students, often using data or drawing on their own experiences as pedagogic hooks. Within their classrooms they had tactics for supporting active learning, including generating effective exercises and creating space and scaffolds for reflection. They had tactics for being student-centred, including finding out about their students, attuning, empathizing, and connecting with students' interests. They had tactics for connecting the techniques of research methods with real life research problems, including narrating stories and going behind the scenes of their own research work.

Through the various components of the study the participants and researchers probed together what makes teaching research methods challenging and distinctive, their responses to the challenges, and the pedagogical choices made. One of the first challenges is about getting a good fit between the methods course and the needs of the methods learner and a repeated refrain from learners and teachers was that mismatches were common. When writing this chapter I came upon this informative course description:

> This course is designed for health care professionals and researchers seeking to consolidate their understanding and ability in contextualising, carrying out, and applying systematic reviews appropriately in health care settings. Core modules will introduce the students to the principles of evidence-based health care, as well as the core skills and methods needed for research design and conduct. Further modules will provide students with specific skills in conducting basic systematic reviews, meta-analysis, and more complex reviews, such as realist reviews, reviews of clinical study reports and diagnostic accuracy reviews.

We see here how embedded systematic review has become in health care and medicine as evidence-based professions. The equivalent would be unlikely in education where one could imagine something like:

> This course is designed for education professionals and researchers seeking skills in contextualising, carrying out, and applying systematic reviews appropriately in education settings where there is considerable skepticism about such methods. Core modules will introduce the students to doing systematic review when the idea of evidence-based education is hugely controversial. …

I am being facetious here only in part, as this is an aspect of the challenge facing teachers of systematic review in education. Fortunately perhaps, advanced courses in systematic review are often multi-disciplinary and attitudes to systematic review are likely to be diverse. Diversity in the preparedness and background of research methods learners was a frequently discussed challenge among teachers in the study, but learners invariably welcomed diverse peers from whom they could learn.

In the study's video stimulated dialogue about teaching and learning systematic review, in a focus group immediately following a short course on synthesis hosted in an education department, teachers immediately responded to an opening question about the challenges of teaching this material by focusing on the need to understand the diverse group. They expressed the need to find out about the background knowledge of course participants so as to avoid making errant assump-

tions and to follow brief introductions with ongoing questioning and monitoring of knowledge and of emotional states. As one participating teacher explained,

> you really need to understand research in order to get what's going on … we don't want to have to assume too much, but on the other hand if you go right back to explaining basic research methods, then you don't have time to get onto the synthesis bit, that which most people come for. So sometimes it's a challenge knowing exactly where, how much sort of background to cover.

Participating students were equally aware of the challenge, one commenting on the usefulness of having an "overview of everything, because obviously everyone has come from slightly different arenas" and acknowledging

> We didn't do super-technical things, but I think that's important because otherwise you get people that don't understand and then you lose half the group, so it's important that the tasks are feasible for everybody, but that they give you the technique so you can go home and do it yourself.

In this course, the disciplinary backgrounds of the students varied somewhat. The teachers managed this, in the way of many of the teachers in the study, by working out—and working with—the varied standpoints in the room. One of the teachers celebrated the pedagogical potential of having "people from different perspectives and different disciplines talking to one another". This was the view of the students too, arguing that "the diversity, speaking to all the different people is, I think, is key in methods, and teaching in particular, because we're all doing similar things, just in different topics". The reasoning was clear too with the reflection that "if you've only got people who have exactly the same positionality, then how do you ever critique your own work and … reflect back and think why are we doing this".

The focus group included a lively discussion about a point in the course when one student, as she put it, "disagreed very strongly with what was being said", explaining that this "was because of disciplinary differences, because I don't have a disciplinary allegiance to that sort of health promotion initiative". The different disciplinary backgrounds supported debate about how synthesised data get reduced with students recognising that the "friction and tension … makes it so much more interesting to kind of discuss".

This should help teachers of systematic review not to fear diversity among students; a standpoint, peer collaborative learning approach can be used to address different attitudes (Nind and Lewthwaite 2018b) and an active learning approach

can address the differences in knowledge. The systematic review teachers in the *Pedagogy of Methodological Learning* study spoke of their tried and tested "slides and then practice, slides and practice", using exercises developed and honed over time. Again the students liked the mix of input with opportunities to practice; "the quantitative stuff came really easily … And if it was applied, then I was really engaged … I could try those [calculations] myself and make sense for myself". This student continued,

> The qualitative exercises in particular I really liked, but I wouldn't have naturally been drawn to them, but I found they were really interesting and found some strength I didn't know I had in doing them, whereas I would have just crunched numbers instead, happily, you know without ever trying to break it into themes.

The focus group discussion turned from the welcome role of the exercises following the underpinning theoretical concepts to the welcome role of discussion between themselves in that "people came with quite a lot of resources in terms of their knowledge and experience and skills". They concurred that they would have liked more time discussing, "to really work out what [quality criterion] was". While the teacher spoke of concerns about the risks of leaving chunks of time in the hands of the students, the students reassured, "by that point we kind of knew each other well enough that it was really helpful doing this group work". They noted that "it's so much nicer talking in peer groups rather than just asking direct questions all the time, because … [for] little bits that you need clarification on, it's easy to do with the person sitting next to you". The complexity of the material and the need for active engagement was recognised by the students:

S3: I think that was quite a hard session to teach.
S2: Yeah.
S3: Because what you wanted to do was to bring out different approaches to judging quality, they're actually similar in many ways, so I think it could have, perhaps it could have been a bit more us sort of just experiencing these quality issues

…

T3: my main objective wasn't actually to do with the exercises. It was just to get you to realise how hard it was, and then realise that that's right, it's hard, that's fine, now work together to make it manageable, and that can be done as well, you can begin to start to do that, and that's really all I wanted out of the session

We are also able to learn from the video stimulated dialogue of this group about the way that pedagogic hooks work to connect students to the learning being targeted. We prompted discussion about a point in the day when everyone was laughing. Reviewing the video excerpt of that moment we were able to see how the methodological learning was being pinned to the substantive finding regarding a point about the sensitivity of the tool and the impact on the message that came from the synthesis. The group were enraptured by the finding that 'two bites of the apple' made a difference, which led them into appreciating how some findings made "a really good soundbite that you could disseminate … in a press release". They appreciated "the point of doing good, methodologically sound studies is so we don't have a soundbite like that based on crappy evidence … that's why systematic reviews are so good". This was important learning and the data provided the pedagogic hook.

Another successful strategy was to use the pedagogic hook of going behind the scenes of the teachers' own research (echoed throughout our study). The teachers talked of liking to teach using their own systematic reviews as examples:

> It's a lot easier. I think because you know whatever it is backwards, … I mean that review, the two bites of an apple was done in 2003, so I don't feel all that familiar with the studies anymore, but if you know something as well as that, it's much easier to talk about it

Reflexivity played an important role in this practice too with another teacher in the team reflecting, "I found it easier to be critical about my own work, partly because I know it so well and partly because then I'm also freed up". He spoke of becoming increasingly interested in the limitations of the work, "not in a self-defeating kind of way, but more just I find them genuinely interesting and challenging … what are we going to do? These limitations are there, how do we proceed from here?". In teaching, he said, "I'm able to say more and be more genuinely reflexive, reflective about the work, just because I did it." The students respected the value they gained from this, one likening it to "going to a really good GP" with knowledge of a broad range of problems. While the teachers valued their own experiences as a teaching resource because "we know the difficulties that we had doing them and we know the mistakes that we have made doing them", the students valued the accompanying depth and credibility, "the answers that you could give to questions having done it, are much more complete and believed". Systematic review as a method has been criticised for being overly formulaic, but this was not my experience in learning from reflective practitioners of the method and these teachers stressed this too, "it's not a nice neat clean process,

[whereby you] turn the wheel on a machine and out comes the review at the end, and it can look like that if you read some of the textbooks". Even the students stressed, "you know the flowcharts and stuff, but actually there's a lot more to consider".

## 4 Conclusion

I first reflected on the politics of doing systematic review when, as Lather (2006) summarised, the "contemporary scene [was] of a resurgent positivism and governmental incursion into the space of research methods" (p. 35). This could equally be said of today and this makes it especially important that when we are teaching the method of systematic review we do some from a position in which teachers and students understand and discuss the standpoint from which it has developed and from which they choose to operate. Like Lather (2006), and many of the teachers in the *Pedagogy of Methodological Learning* study, I advocate teaching systematic review, like research methods more widely, "in such a way that students develop an ability to locate themselves in the tensions that characterize fields of knowledge" (Lather 2006, p. 47). Moreover, when teaching systematic review there are lessons that we can draw from pedagogic research and from other practitioners and students who provide windows into their insights. These enable us to follow the advice of Biesta (2007) and to reflect on research findings to consider "what has been possible" (p. 16) and to use them to make our "problem solving more intelligent" (pp. 20–21). I have found particular value in bringing people together in pedagogic dialogue, where they co-produce clarity about their previously somewhat tacit *know-how* (Ryle 1949), generating a synthesis of another kind to that generated in systematic review. However we elicit it, teachers have craft knowledge that others can, with careful professional judgement, draw upon and apply. Those of us teaching systematic review benefit from this kind of practical reflection and from appreciating the resources that students offer us and each other. Teaching systematic review, as with teaching many social research methods, requires deep knowledge of the method and a willingness to be reflexive and open about its messy realities; to tell of errors that researchers have made and judgements they have formed. It is when we scrutinize pedagogic and methodological decision-making, and teach systematic review so as to avoid a rigid, unquestioning mentality, that we can feel comfortable with the kind of educational researchers we are trying to foster.

**Acknowledgements** I am grateful to my fellow researchers Daniel Kilburn, Sarah Lewthwaite and Rose Wiles and to the teachers and learners of research methods who have contributed to the study and to my thinking.

## References

Barnett-Page, E. & Thomas, J. (2009). Methods for the synthesis of qualitative research: a critical review. Unpublished National Centre for Research Methods (NCRM) Working Paper, NCRM Hub, University of Southampton.

Bearman, M., Smith, C. D., Carbone, A. Slade, C., Baik, C., Hughes-Warrington, M., & Neumann, D. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 625–640.

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory, 57*(1), 1–22.

Brown, S. & McIntyre, D. (1993). *Making Sense of Teaching*. Buckingham: Open University Press.

Coomarasamy, A. & Khan, K. S. (2004). What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *British Medical Journal,* 329, 1–5.

Cooper R., Chenail, R. J., & Fleming, S. (2012). A grounded theory of inductive qualitative research education: Results of a meta-data-analysis. *The Qualitative Report, 17*(52), 1–26.

Earley, M. (2014). A synthesis of the literature on research methods education. *Teaching in Higher Education*, *19*(3), 242–253.

Goldacre, B. (2013). Teachers! What would evidence based practice look like? Retrieved August 30, 2018 from https://www.badscience.net/2013/03/heres-my-paper-on-evidence-and-teaching-for-the-education-minister/.

Hammersley, M. (2001). On 'systematic' reviews of research literatures: a 'narrative' response to Evans & Benefield. *British Educational Research Journal, 27*(5), 543–554.

Hammersley, M. (2004). Some questions about evidence-based practice in education. In G. Thomas & R. Pring (Eds.), *Evidence-based Practice in Education* (pp. 133–149). Maidenhead: Open University.

Hargreaves, D. (1996). Teaching as a research-based profession: Possibilities and prospects. The Teacher Training Agency Annual Lecture, London, TTA.

Kilburn, D., Nind, M., & Wiles, R. (2014). Learning as researchers and teachers: the development of a pedagogical culture for social science research methods? *British Journal of Educational Studies*, *62*(2), 191–207.

Kyriakoulis, K., Patelarou, A., Laliotis, A., Wan, A.C., Matalliotakis, M., Tsiou, C., & Patelarou, E. (2016). Educational strategies for teaching evidence-based practice to undergraduate health students: systematic review. *Journal of Educational Evaluation for Health Professionals*, *13*(34), 1–10.

Lather, P. (2006). Paradigm proliferation as a good thing to think with: teaching research in education as a wild profusion. *International Journal of Qualitative Studies in Education*, *19*(1), 35–57.

Lewthwaite, S., & Nind, M. (2016). Teaching research methods in the social sciences: Expert perspectives on pedagogy and practice. *British Journal of Educational Studies,* 64, 413–430.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic Reviews and Meta-analysis*. New York, NY: Oxford University Press.

MacLure, M. (2005). 'Clarity bordering on stupidity': where's the quality in systematic review? *Journal of Education Policy. 20*(4), 393–416.

Nind, M. (2006). Conducting systematic review in education: a reflexive narrative. *London Review of Education*, *4*(2), 183–95.

Nind, M., Kilburn, D., & Luff, R. (2015). The teaching and learning of social research methods: developments in pedagogical knowledge. *International Journal of Social Research Methodology*, *18*(5), 455–461.

Nind, M. & Lewthwaite. S. (2018a). Methods that teach: developing pedagogic research methods, developing pedagogy. *International Journal of Research & Method in Education, 41*(4), 398–410.

Nind, M. & Lewthwaite, S. (2018b). Hard to teach: inclusive pedagogy in social science research methods education. *International Journal of Inclusive Education*, *22*(1), 74–88.

Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of Qualitative Health Research: A Practical Guide to Meta-analysis and Meta-synthesis.* Thousand Oaks, CA: Sage.

Pring, R. (2004). Conclusion: Evidence-based policy and practice, In G. Thomas & R. Pring (Eds.), *Evidence-based Practice in Education* (pp. 201–212). Maidenhead: Open University.

Ryle, G. (1949). *The Concept of Mind*. New York: Hutchinson.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23.

Thomas, G. (2012). Changing our landscape of inquiry for a new science of education, *Harvard Educational Review, 82*(1), 26–51.

Wagner, C., Garner, M., & Kawulich, B. (2011). The state of the art of teaching research methods in the social sciences: Towards a pedagogical culture. *Studies in Higher Education*, *36*(1), 75–88.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Why Publish a Systematic Review: An Editor's and Reader's Perspective

Alicia C. Dowd and Royel M. Johnson

"Stylish" academic writers write "with passion, with courage, with craft, and with style" (Sword 2011, p. 11). By these standards, Mark Petticrew and Helen Roberts can well be characterized as writers with style. Their much cited book *Systematic Reviews in the Social Sciences: A Practical Guide* (2006) has the hallmarks of passion (for the methods they promulgate), courage (anticipating and effectively countering the concerns of naysayers who would dismiss their methods), and, most of all, craft (the craft of writing clear, accessible, and compelling text). Readers do not have to venture far into Petticrew and Roberts' *Practical Guide* before encountering engaging examples, a diverse array of topics, and notable characters (Lao-Tze, Confucius, and former US Secretary of Defense Donald Rumsfeld, prominently among them). Metaphors draw readers in at every turn and offer persuasive reasons to follow the authors' lead. Systematic reviews, we learn early on, "provide a redress to the natural tendency of readers and researchers to be swayed by [biases], and … fulfill an essential role as a sort of scientific gyroscope, with an in-built self-righting mechanism" (p. 6). Who among us, in our professional and personal lives, would not benefit from a gyroscope or some other "in-built self-righting mechanism"? This has a clear appeal.

A. C. Dowd (✉)
College of Education, Department of Education Policy Studies,
Center for the Study of Higher Education (CSHE),
The Pennsylvania State University, Pennsylvania, USA
e-mail: dowd@psu.edu

R. M. Johnson
College of Education, Department of Education Policy Studies,
Center for the Study of Higher Education (CSHE),
Department of African American Studies, The Pennsylvania State University,
Pennsylvania, USAe-mail: rmj19@psu.edu

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_5

It is no wonder that the *Practical Guide* has been highly influential in shaping the work of researchers who conduct systematic reviews.

Petticrew and Roberts' (2006) influential text reminds us that, for maximum benefit and impact, researchers should "story" their systematic reviews with people and places and an orientation to readers as audience members. If, as William Zinsser says at the outset of *On Writing Well* (1998), "Writing is like talking to someone else on paper" (p. x), then the audience should always matter to the author and the author's voice always matters to readers. This is not the same as saying 'put the audience first' or 'write for your audience' or 'lose your scientific voice.' Research and writing in the social and health sciences is carried out to produce knowledge to address social and humanistic problems, not to please readers (or editors). In answer to the central question of this chapter, "Why publish systematic reviews?", the task of communicating findings and recommendations must be accomplished, otherwise study findings will languish unread and uncited. Authors seek to publish their work to have their ideas heard and for others to take up their study findings in consequential ways. In comparison with conference presentations, meetings with policymakers, and other forms of in-person dissemination, text-based presentations of findings reach a wider audience and remain available as an enduring reference.

As an editor (Dowd),[1] researcher (Johnson), and diligent readers (both of us) of systematic reviews over the past several years, we have read many manuscripts and published journal articles that diligently follow the steps of Petticrew and Roberts' (2006) prescribed methods, but do not even attempt to emulate the capacity of these two maestros for stylistic presentation. Authors of systematic reviews often present a mechanical accounting of their results—full of lists, counts, tables, and classifications—with little pause for considering the people, places, and problems that were the concern of the authors of the primary studies. The task of communicating "nuanced" findings that are in need of "careful interpretation" (Petticrew and Roberts 2006, p. 248) is often neglected or superficially engaged. Like Petticrew and Roberts, we observe two recurring flaws of systematic review studies (in manuscript and published form): a "lack of any systematic

[1]Alicia Dowd began a term in July 2016 as an associate editor of the Review of Educational Research (RER), an international journal published by the American Educational Research Association. All statements, interpretations, and views in this chapter are hers alone and are not to be read as a formal statement or shared opinion of RER editors or editorial board members.

critical appraisal of the included studies" and a "lack of exploration of heterogeneity among the studies" (p. 271).

In addition, many scholars struggle to extract compelling recommendations from their reviews, even when the literature incorporated within it is extensive. To be influential in communicating the results of systematic reviews, researchers must consider how they will go about "selling the story" and "making sure key messages are heard" (Petticrew and Roberts 2006, p. 248). However, as these leading practitioners and teachers of systematic review methods have observed, researchers often lack or fail to engage the necessary storytelling skills. Although research is produced and read by "by people in specific times and places, with lives as well as careers," as sociologist Robert R. Alford pointed out in *The Craft of Inquiry: Theories, Methods, Evidence* (1998, p. 7),[2] the prescriptions of systematic reviews (in our readings) are often not well variegated by the who, what, where, and why of the research enterprise.

## 1 Starting Points and Standpoints

We believe the remedy to this problem is for authors to story and inhabit systematic review articles with the variety of compelling people and places that the primary research study authors deemed worthy of investigation. Inhabiting systematic review reports can be accomplished without further privileging the most dominant researchers—and thus upholding an important goal of systematic review, the "democratization of knowledge" (Petticrew and Roberts 2006, p. 7)—by being sure to discuss characteristics that have elevated some studies over others as well as characteristics that *should* warrant greater attention, for reasons the systematic review author must articulate. A study might be compelling to the author because it is highly cited, incorporates a new theoretical perspective, represents the vanguard of an emerging strand of scholarship, or any number of reasons that the researcher can explain, transparently revealing epistemological, political-economic, and professional allegiances in the process. This can be achieved without diminishing the scientific character of the systematic review findings.

[2]Although Alford was referring specifically to sociological research, we believe this applies to all social science research and highlight the relevance of his words in this broader context.

Alford (1998) argues for the integrated use of multiple paradigms of research (which he groups broadly for purposes of explication as multivariate, interpretive, and historical) and for explanations that engage the contradictions of findings produced from a variety of standpoints and epistemologies. This approach, which informs our own scholarship, allows researchers to acknowledge, from a postmodern standpoint, that "knowledge is historically contingent and shaped by human interests and social values, rather than external to us, completely objective, and eternal, as the extreme positive view would have it" (p. 3). At the same time, researchers can nevertheless embrace the "usefulness of a positivist epistemology," which "lies in the pragmatic assumption that there is a real world out there, whose characteristics can be observed, sometimes measured, and then generalized about in a way that comes close to the truth" (p. 3). To manage multiple perspectives such as these, Alford encourages researchers to foreground one type of paradigmatic approach (e.g., multivariate) while drawing on the assumptions of other paradigms that continue to operate in the background (e.g., the assumption that the variables and models selected for multivariate analysis have a historical context and are value laden).

## 2 An Editor's Perspective

During my years of service to date (2016–2019) as an associate editor of the *Review of Educational Research* (RER), a broad-interest educational research journal published by the American Educational Research Association (AERA) for an international readership, I (Dowd) reviewed dozens of manuscripts each year, a great number of which were systematic reviews. Approximately one-third to one-half of the manuscripts in my editor's queue at any given time were specifically described by the authors as involving systematic review methods (often but not always including meta-analysis). Although many other authors who submitted manuscripts did not specifically describe systematic review as their methodology, they did describe a comprehensive approach to the literature review that involved notable hallmarks of systematic review methods (such as structured data base searches, well defined inclusion and exclusion criteria, and precisely detailed analytical procedures).

Given the nature of the supply of manuscripts submitted for review, it is not surprising that a large and growing proportion of articles published in RER in recent years have involved systematic review methods. Table 1 summarizes this publication trend, by categorizing the articles published in RER from September

**Table 1** Articles Published in the Review of Educational Research, Sept. 2019–Dec. 2018, by Review Type

| Year[a] | Systematic Review | Systematic Review with Meta-Analysis | Meta-Analysis without Systematic Review | Comprehensive[b] | Other[c] | Total | Proportion Systematic Review[d] |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 2009 | 0 | 0 | 6 | 0 | 3 | 9 | 0.00 |
| 2010 | 2 | 3 | 4 | 4 | 5 | 18 | 0.29 |
| 2011 | 0 | 1 | 5 | 5 | 6 | 17 | 0.06 |
| 2012 | 0 | 0 | 3 | 5 | 5 | 13 | 0.00 |
| 2013 | 4 | 0 | 4 | 4 | 4 | 16 | 0.25 |
| 2014 | 3 | 0 | 4 | 6 | 5 | 18 | 0.17 |
| 2015 | 4 | 1 | 7 | 5 | 3 | 20 | 0.25 |
| 2016 | 6 | 4 | 13 | 7 | 4 | 34 | 0.29 |
| 2017 | 6 | 7 | 6 | 9 | 4 | 32 | 0.41 |
| 2018 | 4 | 6 | 5 | 6 | 2 | 23 | 0.43 |

*Note.* Authors' calculations based on review of titles, abstracts, text, and references of all articles published in RER from Sept. 2009 to Dec. 2018, obtained from http://journals.sagepub.com/home/rer. An article's assignment to a category reflects the methodological descriptions presented by the authors

[a]The count for 2009 is partial for the year, including only those published during Gaea Leinhardt's editorship (Volume 79, issues 3 and 4). The remaining years cover the editorships of Zeus Leonardo and Frank Worrell (co-editors-in-chief, 2012–2014), Frank Worrell (2015–2016), and P. Karen Murphy (2017–2018, including articles prepublished in Online-First in 2018 that later appeared in print in 2019).

[b]Refers to articles that do not explicitly designate "systematic review" as the review method, but do include transparent methods such as a list of searched data bases, specific keywords, date ranges, inclusion/exclusion criteria, quality assessments, and consistent coding schema

[c]Reviews that do not demonstrate characteristics of either systematic review or meta-analysis, including expert reviews, methodological guidance reports, and theory or model development

[d]Includes both Systematic Review (col. 2) and Systematic Review with Meta-Analysis (col. 3) categories as proportion of the total

of 2009 to December of 2018 by their use of systematic review methods. Two categories, called "systematic review" (col. 2) and "systematic review with meta-analysis" (col. 3), include those articles where the authors demonstrated their use

of systematic review methods and specifically described their study methodology as involving systematic review.

The share of systematic review articles (with or without meta-analysis) as a proportion of the total published is shown in column 8. The proportion has fluctuated, but the overall trend has been upward. In all except one of the past five years, systematic review articles have contributed one-quarter or greater of the total published. From 2013 to 2016 the share ranged from 17% to 29% and then increased in 2017 and 2018 to 41% and 43% respectively. Keeping in mind that even those meta-analyses that were not described by the authors as involving systematic review (col. 4) and all of the articles we chose to categorize as "comprehensive" (col. 5) have hallmarks of the systematic review method, it is clear that we and other RER editors and readers have been well exposed to systematic review methods in recent years.

The summary data in Table 1 indicate that systematic reviews are finding a home in RER, which is a highly cited journal, typically ranking (as measured by impact factor) near the top of educational research journals.[3] However, for each article published in RER, there were many more submitted works that were reviewed by the editorial team and peer reviewers and not accepted for publication.[4] In my role as editor, I was struck by the high number of authors of submitted manuscripts using systematic review methods who reported their findings in algorithmic terms. The methods had swallowed the authors, it seemed, who felt compelled to enumerate in their text all of the counts, proportions, lists, and categories used to taxonomize the results of these reviews.

Even where thematic findings had been generated, many authors still led their presentation with an "x of y studies examined [X topic]…" formulation, rather than advancing their synthesis using integrative topic sentences. While counting and enumeration are often appropriate forms of summary, when this sentence structure recurs repeatedly as the first sentence of the multiple paragraphs and pages of a results section, it is easy to lose interest. I encountered authors who were using this enumeration approach alongside useful and extensive summary tables and figures designed to convey the same information. Such

[3]Based on the Journal Citation Reports, 2018 release; Scopus, 2018 release; and Google Scholar, RER's two-year impact factor was 8.24 and the journal was ranked first out of 239 journals in the category of Education and Educational Research (see https://journals.sagepub.com/metrics/rer).

[4]The RER publication acceptance rate varies annually but has consistently been less than 10% of submitted manuscripts.

recounting left me as a reader and editor without a toehold or a compass to enter what was often a vast landscape of scholarship, sometimes spanning decades and continents and very often focused on topics that were new to me. Peer reviewers, too, would comment on the challenge of investing themselves into the findings of studies that the authors had not shaped through meaningful synthesis for their readers.

In such a landscape, an editor must lean on the author as the "intelligent provider" of the research synthesis (Petticrew and Roberts 2006, p. 272, citing Davies 2004). The intelligent provider of the results of systematic review is scientific in their approach—this is clearly a primary value of the systematic review research community—but is also a guide who invites readers into the reviewed literature to achieve the goals of the review. The goal of systematic review is not merely to be "comprehensive"; the objective also is to "answer a specific question," "reduce bias in the selection and inclusion of studies," "appraise the quality of the included studies," and "summarize them objectively, with transparency in the methods employed" (Petticrew and Roberts 2006, p. 266).

It struck me that the majority of researchers within my (non-random and not necessarily representative) sample of RER manuscript submissions who had utilized systematic review methods had gone to tremendous lengths to conduct extensive data base searches, winnow down the often voluminous "hits" using clearly defined inclusion and exclusion criteria, and then analyze a subset of the literature using procedures well-documented at each step. These rigorous and time-consuming aspects of the systematic review method had perhaps exhausted the researchers, I felt, because the quality of the discussion and implications sections often paled in comparison to the quality of the methods. This was understandable to me and I was motivated to write this chapter for scholars who might benefit from encouragement and advice from an editor's perspective as they tackled these last stages of the research and publication process.

My scholarship has involved multivariate and interpretive methods of empirical study. As an action researcher I have foregrounded an advocacy stance in my work, which has been focused on issues of equity, particularly racial equity (see e.g., Dowd and Bensimon 2015). Neither a practitioner nor scholar of systematic review, I encountered the method through my editing. To learn about these methods in a more structured manner, in April 2018, I attended an introductory-level AERA professional development workshop focused on this methodology. I also asked my faculty colleague Royel Johnson for his perspective, because I knew he had immersed himself in reading the methodological literature as he embarked on

a systematic review study of the college access and experiences of (former) foster youth in the United States.

## 3 A Reader's (and Researcher's) Perspective

As an educational researcher and social scientist, I (Johnson) have relied on qualitative and quantitative methods and drew, in my work, on multivariate and interpretive epistemologies. These approaches have been useful in exploring complex social phenomena, as well modeling and testing the relationships between variables related to college student success, particularly for vulnerable student populations in higher education. During the summer of 2017, however, I was introduced to a new method—well new to me at least: systematic literature review.

My introduction to the methods of systematic review was quite timely as I had begun expanding my research on students impacted by foster care (e.g., Johnson and Strayhorn, 2019). I was particularly struck by headlines that were popping up at that time in news and popular media outlets in the United States (U.S.) that painted a "doom and gloom" picture about the education trajectory and outcomes of youth formerly in foster care. Equally troubling, it struck me that efforts to improve the college access, experiences, and outcomes of this group of students through evidence-based policy and practice would be poorly informed by the work of my own scholarly field (higher education), which to the best of my knowledge at the time had produced little empirical research on the topic.

Like all good researchers embarking on new studies, it seemed important to me, before reaching any stronger conclusions about the quality of the research base, to first locate and familiarize myself with the broader existing literature on the topic, including studies conducted in other fields of study. What is it that we know from research about the experiences and outcomes of college students formerly in foster care in college? This was my guiding question as I searched the literature of the higher education field and in related areas such as social work and public policy.

Searching the literature to answer this question was initially daunting. There were no apparent comprehensive literature reviews on the topic. And the places where I was *inclined* to look for studies yielded very few returns. Notice my emphasis on "inclined." One of the goals of systematic literature review is to reduce reviewer bias. If not careful, such inclinations can lead to incomplete or partial collections of information or studies, and also result in erroneous (and biased) conclusions about the state of knowledge on a given topic. Winchester and Salji (2016) refer to this as "cherry picking" (p. 310). Indeed, as researchers, we are not empty vessels, nor do we approach our work as such—though some

might suggest otherwise. Our backgrounds, experiences, and perspectives (e.g., about what constitutes as knowledge) all shape the questions we ask, the places we search for answers to those questions, and what we deem as credible.

To avoid "cherry picking" and produce the most comprehensive literature review possible, I set out to learn as much as I could about conducting a systematic literature review. I identified at least a dozen texts and scholarly publications, including Higgins and Green's (2006) widely-cited book. I also reviewed the recommendations incorporated in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols (Moher et al. 2015)—also known as the PRISMA statement. These sources, while instructive, seemed to almost exclusively focus on the more technical aspects of systematic reviews, offering steps, guidance, and recommendations for developing protocols, defining search terms, outlining inclusion/exclusion criteria, and critically appraising studies—the traditional hallmarks of 'rigor' and 'quality' for this method. However, few resources, as mentioned in previous sections, offer recommendations or strategies for "telling the systematic review story" (Petticrew and Roberts 2006, p. 248).

As I worked on turning the results of my systemic review of college students impacted by foster care into a journal article (Johnson, In Press), I read studies published in RER and other journals. I looked for models that would help me to determine how to position myself as a compelling and persuasive storyteller of my study's topic, methods, findings, and recommendations. Of the dozen or so reviews I read, published over the past decade, only a few emerged as exemplars to inform my decisions as a writer. One such study was a review published by three colleagues in the field of higher education, Crisp et al. (2015). Their study focused on identifying factors associated with academic success outcomes for undergraduate Latina/o students. It stood out to me for its clarity of purpose and rationale. Another study by Poon and colleagues (2016), which examined the model minority myth among Asian Americans and Pacific Islanders (AAPI), stood out as well. Notably, the authors offered insight about their motivations for the work while clearly stating their researcher positionalities, describing themselves "as longtime educators and scholars in the fields of higher education and student affairs committed to AAPI communities and social justice" (p. 476). Such statements acknowledging one's relationship or commitment to the subject of study were rare in the systematic review literature I read. This statement resonated with me because, just as Poon et al. were committed to social justice for AAPI communities, I, too, was vested in and committed to improving the material conditions students impacted by foster care experienced in college.

From these starting points and standpoints, and as we move to argue for 'storying' the systematic review, we acknowledge that our epistemological values may not align with those of the leading methodological experts of systematic review

(or of other editors and readers of systematic reviews). Our recommendations may, therefore, be of more interest and value to researchers who are interested in carrying out comprehensive reviews that are systematic (rather than, specifically, "systematic reviews"). This distinction is reflected in our selection of a few published RER articles discussed in the following section, where the topics of the featured studies also reflect our interests in educational policy, equity, and student success. There we include works identified by the authors as a systematic, comprehensive, meta-analytic, or critical review. It is important to note, given the emphasis on unbiased reporting in the systematic review methodology, however, that we judged all of these works as providing a transparent and detailed description of their purpose and methods. All studies reported search criteria, data bases searched, inclusion and exclusion criteria moving from broader to narrower criteria, and supplementary tables providing a brief methodological summary of every article included in the group of studies selected for focal synthesis.

## 4 Storying the Systematic Review

The five studies discussed in this section were successful in RER's peer-review and editorial process. We selected them as a handful of varied examples to highlight how authors of articles published in RER "story" their findings in consequential and compelling ways. These published works guide us scientifically and persuasively through the literature reviewed. At the same time, the authors acted as an "intelligent provider" (Petticrew and Roberts 2006, p. 272, citing Davies 2004) of information by inhabiting the review with the concerns of particular people in particular places. The problems of study are teased out in complex ways, using multifocal perspectives grounded in theory, history, or geography. Two had an international scope and three were restricted to studies conducted in settings in the U.S. Whether crossing national boundaries or focused on the U.S. only, each review engaged variations in the places where the focal policies and practices were carried out. Further, all of the reviews we discuss in this section story their analyses with variation in the characteristics of learners and in the educational practices and policies being examined through the review.

### 4.1 Theoretical Propositions as Multifocal Lenses: Storying Reviews with Ideas

Østby et al. (2019) of "Does Education Lead to Pacification: A Systematic Review of Quantitative Studies on Education and Political Violence" capture the

attention of the non-specialist RER reader by citing Steven Pinker's acclaimed book *The Better Angels of Our Nature*. They highlight Pinker's metaphor characterizing education as an "escalator of reason," an escalator that has the power to act globally as a "pacifying" force (p. 46). Noting wide acceptance of the idea that societies with a higher level of education will experience lower levels of political violence and armed conflict, the authors quickly shake this assumption. Recent studies have shown, for example, that terrorists and genocide perpetrators have had higher than average levels of education relative to others in their societies. Further, the story of the relationship between increases in educational attainments and political violence in a society unfolds in a more complicated manner when factors such as initial baselines of education in the population, gender disparities in access to elementary and secondary schools, and inequalities among socio-economic groups are taken into account.

Østby et al. (2019) organize their review of 42 quantitative studies of education and political violence around theoretical propositions that add complexity to the notion that education is a pacifying force. From an economic perspective, there are several reasons why education should lead to a decrease in political violence and social unrest. Those with more education typically have higher earnings and may be deterred from engaging in social unrest because they may lose their jobs, a consideration of less consequence to the unemployed or those with marginal labor force status. Alternatively, a political explanation for a positive impact can be found in the fact that those who are more highly educated are more greatly exposed to and culturally inculcated through the curriculum sanctioned by the government, which may be dominated by nationalistic historical narratives.

In contrast, a sociological explanation based in theories of relative deprivation points in the opposite direction, as the sociologist attends to inequality among socio-economic groups. Groups that lack political power and have historically been oppressed or disenfranchised may become more likely to engage in violent political action as they gain in educational attainment yet continue to lag behind dominant social groups. When it comes to the study of the relationship between education and political violence, Østby et al. (2019) show that it is insufficient to characterize a country in terms of the educational attainment of a population without also considering governmental influence in the curriculum, political oppression, and educational inequality.

As Østby et al. (2019) discuss contrasting theoretical propositions for positive and negative associations between education and violence, the reader quickly buys into the premise that the authors' study of this "complex, multi-faceted, and multidirectional" phenomenon is highly consequential (p. 47). More nuanced understandings clearly hold the potential to inform the manner and

degree of governmental and philanthropic investments in education in developed and developing countries around the globe.

Similarly, García and Saavedra (2017), in their examination of the impacts of "conditional cash transfer programs," utilize human capital and household decision making theories to introduce readers to a very precise, yet varied, set of hypotheses that they subsequently use to structure the reporting of their results. These economic hypotheses postulate the potential effects of governmental programs that provide cash rewards to households or individuals to encourage them to who respond to policy incentives in desired ways. They highlight that the direction and strength of effects depend on a range of household inputs such as parental education, sources of income (e.g., formal and informal labor force participation), time use among household members (adults and children), and community characteristics. As other researchers have before them, these authors meta-analyze impact estimates from studies meeting their threshold methodological quality criteria for making causal claims. Their review synthesizes 94 studies of 47 conditional cash transfer (CCT) programs carried out in 31 countries (p. 929, 934). Their work builds on and extends the findings of prior meta-analyses that produced CCT impact estimates by also examining questions of cost-effectiveness.

In García and Saavedra's (2017) study, the examination of effects comprises seven outcomes: "primary school enrollment, primary school attendance, primary school dropout, secondary school attendance, secondary school dropout, and school completion" (p. 933). The authors demonstrate that variations in program characteristics delineated in their review correspond to variations in program effectiveness, both in terms of these various effects and of economic investments in the intervention. An important finding of this study (among many others) is that "all else constant, primary enrollment impact estimates are greater in CCT programs that complement cash transfers with supply-side interventions such as school grants" (p. 923). The finding is consequential to future policy design because less than 10% of the CCT programs studied had a design component that attempted to incentivize changes in schooling practices at the same time they were providing incentives for greater household investments in education.

## 4.2 Engaging Interactions: Storying Reviews with People, Policies, and Practices

The capacity to model, measure, and attend to dynamic interactions among governmental policies, educational institutions or settings, and the behavior of

individuals is a hallmark of the quality of this small set of exemplar RER articles. Each of the studies we reviewed in this chapter attend to differences among students in their experiences of schools and educational interventions with varying characteristics. For some this involves differences in national contexts and for others differences among demographic groups.

Welsh and Little (2018), for example, motivate their comprehensive review through synthesis of a large body of research that raises concerns about racial inequities in the administration of disciplinary procedures in elementary and secondary schools in the U.S. Prior studies had shown that Black boys in U.S. schools were more likely than girls and peers with other racial characteristics to receive out-of-school suspensions and other forms of sanctions that diminished students opportunities to learn or exposed them to involvement in the criminal justice system. The authors engage readers in the "complexity of the underlying drivers of discipline disparites" (p. 754) by showing that the phenomenon of the unequal administration of discipline cannot be fully accounted for by behavioral differences among students of different racial and gender characteristics.

By incorporating a synthesis of studies that delineate the problems of inequitable disciplinary treatment alongside a synthesis of what is known about programmatic interventions intended to improve school climate and safety, Welsh and Little make a unique contribution to the extant literature. Winnowing down from an initial universe of over 1300 studies yielded through their broad search criteria, they focus our attention on 183 peer-reviewed empirical studies published between 1990 and 2017 (p. 754). Like García and Saavedra (2017), these authors use critical appraisal of the methodological characteristics of the empirical literature they review to place the findings of some studies in the foreground of their analysis and others in the background. Pointing out that many earlier studies used two-level statistical models (e.g. individual and classroom level), they make the case for bringing the findings of multi-level models that incorporate variables measuring student-, classroom-, school-, and neighborhood-level effects to the foreground. Multi-level modeling allows the complexities of interactions among students, teachers, and schools that are enacting particular policies and practices to emerge.

Teasing out the contributors to disciplinary disparities among racial, gender, and income groups, and also highlighting studies that show unequal treatment of students with learning disabilities and lesbian, bisexual, trans*, and queer-identified youth, Welsh and Little (2018) conclude that race "trumps other student characteristics in explaining discipline disparities" (p. 757). This finding contextualizes their deeper examination of factors such as the racial and gender "match" of teachers and students, especially in public schools where the predominantly White,

female teaching force includes very few Black male teachers. Evidence suggests that perceptions, biases, and judgments of teachers and other school personnel (e.g. administrators, security officers) matter in important ways that are not fully addressed by programmatic interventions that have mainly focused on moderating students' behavior. The interventions examined in this review, therefore, run the gamut from those that seek to instill students with greater social and emotional control to those that attempt to establish "restorative justice" procedures (p. 778).

Ultimately Welsh and Little (2018) conclude that "cultural mismatches play a key role in explaining the discipline disparities" but "there is no 'smoking gun' or evidence of bias and discrimination on the part of teachers and school leaders" (p. 780). By presenting a highly nuanced portrayal of the complexities of interactions in schools, Welsh and Little create a compelling foundation for the next generation of research. Their conclusion explicates the challenges to modeling causal effects and highlights the power of interdisciplinary theories. They synthesized literature from different fields of study including education, social work, and criminal justice to expand our understanding of the interactions of students and authorities who judge the nature of disciplinary infractions and determine sanctions. Their insights lend credence to their arguments that future analyses should be informed by integrative theories that enable awareness of local school contexts and neighborhood settings.

The importance of engaging differences in student characteristics and the settings in which students go to school or college also emerges strongly in Bjorklund's (2018) study of "undocumented" students enrolled or seeking to enroll in higher education in the United States, where the term undocumented refers to immigrants whose presence in the country is not protected by any legal status such as citizenship, permanent resident, or temporary worker. This study makes a contribution by synthesizing 81 studies, the bulk of which were peer-reviewed journal articles published between 2001 and 2016, while attending to differences in the national origins; racial and ethnic characteristics; language use; and generational status of individuals with unauthorized standing in the U.S. Generational status contrasts adult immigrants with child immigrants, who are referred to as the 1.5 generation and "DACA" students, the latter term deriving from a failed federal legislative attempt, the Deferred Action for Childhood Arrivals (DACA), to allow the children of immigrants who were brought to the country by their parents to have social membership rights such as the right to work and receive college financial aid from governmental sources. DACA also sought to establish a pathway to citizenship for unauthorized residents.

Using the word "undocumented" carries political freight in a highly charged social context in which others, with opposing political views, use terms such as

"illegal aliens" (Bjorklund 2018, p. 631). Bjorklund acknowledges that he is politically situated and that his review has a political point of view by titling his study a "critical review." Rather than claiming a lack of bias with respect to the treatment of undocumented students, the author positions himself within the literature with a clear purpose of generating findings that will inform policy makers and practitioners who would like to support the success of undocumented students. Bjorklund then stories his findings through a review of relevant judicial cases, changes in and attempted changes to federal law, and variations in state laws and policies, the latter of which are highly salient in the U.S., where education is primarily governed at the state level. These accounts are more accurately described as purposeful relative to the goals of the review, rather than unbiased. Nevertheless, in describing historical facts and the specifics of policy design, the author's account proves trustworthy to readers in the sense that these details are transparently referenced with respect to documented legislative actions, proposed and implemented federal and state policies and judicial case law, including Supreme Court rulings.

The extent to which individual legislatures in the 50 U.S. states allow undocumented students to access state benefits (such as reduced college tuition charges for state residents) emerges as an important aspect of this review. Geography matters, too, in the consideration of student characteristics and the design of institutional practices and policies to meet the varied needs of undocumented college students. Some states, cities, and rural areas have a larger proportion of unauthorized immigrants from border countries such as Mexico and countries in Central and South America (which figure prominently in the narratives of those opposing state and federal policies that would provide higher education benefits to undocumented college students), whereas other regions have a larger proportion of immigrants from Asia and Europe.

In addition to reporting salient themes and appraising studies for their intellectual merit, authors of systematic reviews help translate a research purpose for intended audiences and offer a charge for the future. Crisp et al. (2015) accomplish this precisely in their review of literature on undergraduate Latina/o students and factors associated with their academic success. The authors firmly establish the significance of their review, using trend data and statistics showing the growth of the Latina/o population in the U.S. broadly to demonstrate the timeliness of their topic. This growth, the authors note, has also resulted in increases in college enrollment for Latina/o students across the wide variety of postsecondary institutions in the U.S, but institutional policies and practices have not kept up in response to this demographic change. Appreciating the within-group differences of Latina/o students, Crisp and colleagues also acknowledge the varied experiences of Mexican, Peruvian, Colombian, and Salvadoran college students. Such

distinctions and clarifications help frame their review within the full context of the topic for the reader.

Crisp, Taggart, and Nora's (2015) methodological decisions for their systematic review are also clearly informed by the authors' positionalities and commitment to "be inclusive of a broad range of research perspectives and paradigms" (p. 253). They employ a broad set of search terms and inclusion criteria so as to fully capture the diversity that exists among Latina/o students' college experiences. For instance, the authors operationalize the conceptualization and measurement of 'academic success outcomes' broadly. This yields a wider range of studies—employing quantitative, qualitative, and mixed methodological approaches—for inclusion.

Consistent with this approach, prior to describing the methodological steps taken in their review, the authors present a "prereview note." The purpose of this section was to offer additional context about larger and overlapping structural, cultural, and economic conditions influencing Latina/o students broadly. For instance, the authors discuss how "social phenomena such as racism and language stigmas" impact the educational experiences of Latina/o students. They also acknowledge cultural mismatch between students' home culture and school/classroom culture, which "has been linked to academic difficulties among Latina/o students" (Crisp et al. 2015, p. 251). These are just several examples of how the authors help contextualize the topic for readers, especially for those not familiar with the topic or with larger issues impacting Latina/o groups. This is also necessary context for a reader to make sense of the major findings presented in a later section.

Finally, we appreciate the way that Crisp et al. (2015) also make their intended audience of educational researchers clear. They spend the balance of their review, after reporting findings, making connections among various strands of the research they have reviewed, their goal being to "put scholars on a more direct path to developing implications for policy and practice." They direct their charge, specifically, to call on "the attention of equity-minded scholars" (p. 263). As these authors illustrate, knowing your audience allows you to *story* your systematic review in ways that directly speak to the intended benefactor(s).

## 5 Gaining an Audience by Connecting with Readers

This chapter posed the question "Why publish systematic reviews?" and offered an editor's and a reader's response. The reason to publish systematic reviews of educational research is to communicate with people who may or may not be familiar with the topic of study. In the task of "selling the story" and "making

sure key messages are heard" (Petticrew and Roberts 2006, p. 248), authors must generate new ideas and reconfigure existing ideas for readers who hold the potential, informed by the published article, to more capably tackle complex problems of society that involve the thoroughly human endeavors of teaching and learning. More often than not, this will not involve producing "the" answer to a uni-dimensional framing of a problem. For this reason, published works should engage the heterogeneity and dynamism of the educational enterprise, rather than present static taxonomies and categorizations.

Good reviews offer clear and compelling answers to questions related to the "why" and "when" of a study. That is, why is this review important? And why is *now* the right time to do it? They also story and inhabit their text with particular people in particular places to help contextualize the problem or issue being addressed. Rich description and context adds texture to otherwise flat or unidimensional reviews. Inhabited reviews are not only well-focused, presenting a clear and compelling rationale for their work, but they also have a target audience. Petticrew and Roberts (2006, citing "Research to Policy") ask a poignant question: "To whom should the message be delivered" (p. 252). The 'know your audience' adage is highly relevant. We have illustrated a variety of ways that authors story systematic and other types of reviews to extract meaning in ways that are authentic to their purpose as well as situated in histories, policies, and schooling practices that are consequential.

Introducing one's relationship to a topic of study not only lends transparency to the task of communicating findings, it also opens the door to acknowledging variation among readers of a publication. Members of the research community and those who draw on research to inform policy and practice were all raised on some notion of what counts as good and valuable research. Critical appraisals of research based in the scientific principles of systematic review can warrant the quality of the findings. Absent active depictions of the lived experiences and human relationships of people in the sites of study, systematic reviews frequently yield prescriptions directed at a generic audience of academic researchers who are admonished to produce higher quality research. How researchers might respond to a call for higher quality research—and what that will mean to them—will certainly depend on their academic training, epistemology, personal and professional relationships, available resources, and career trajectory.

One way to relate to readers is to explicitly engage multiple paradigms and research traditions with respect, keeping in mind that the flaws as much as the merits of research "illustrate the human character of any contribution to social science" (Alford 1998, p. 7). The inhabited reviews we have in mind will be as systematic as they are humanistic in attending to the variations of people, place,

and audience that characterize "the ways in which people do real research projects in real institutions" (Alford 1998, p. 7). Their authors will keep in mind that the quest to know 'what works' in a generalized sense, which is worthy and essential for the expenditure of public resources, does not diminish a parent's or community's interests in knowing what works for their child or community members. Producers and consumers of research advocate for their ideas all the time. Whether located in the foreground or background of a research project, advocacy is inescapable (Alford 1998)—even if one is advocating for the use of unbiased studies of causal impact and effectiveness.

**Acknowledgements** The authors appreciate the valuable research assistance of Ms. Ali Watts, doctoral student in the higher education program at the Pennsylvania State University and research associate at the Center for the Study of Higher Education (CSHE).

## References

Alford, R. R. (1998). *The craft of inquiry*. New York: Oxford University Press.

Bjorklund, P. (2018). Undocumented students in higher education: A review of the literature, 2001 to 2016. *Review of Educational Research, 88*(5), 631–670. https://doi.org/10.3102/0034654318783018.

Crisp, G., Taggart, A., & Nora, A. (2015). Undergraduate Latina/o students: A systematic review of research identifying factors contributing to academic success outcomes. *Review of Educational Research, 85*(2), 249–274. https://doi.org/10.3102/0034654314551064.

Dowd, A. C., & Bensimon, E. M. (2015). *Engaging the "race question": Accountability and equity in higher education*. New York: Teachers College Press.

García, S., & Saavedra, J. E. (2017). Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: A meta-analysis. *Review of Educational Research, 87*(5), 921–965. https://doi.org/10.3102/0034654317723008.

Higgins, J. P. T., & Green, S. (Eds.). (2006). *Cochrane handbook for systematic reviews of interventions 4.2.6* (Vol. Issue 4). Chichester, UK: Wiley: The Cochrane Library.

Johnson, R.M. & Strayhorn, T.L. (2019). Preparing youth in foster care for college through an early outreach program. *Journal of College Student Development, 60*(5), 612-616

Johnson (In Press). The state of research on undergraduate youth formerly in foster care: A systematic review. *Journal of Diversity in Higher Education*. doi 10.1037/dhe0000150

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews, 4*(1), 1.

Østby, G., Urdal, H., & Dupuy, K. (2019). Does education lead to pacification? A systematic review of statistical studies on education and political violence. *Review of Educational Research, 89*(1), 46–92. https://doi.org/10.3102/0034654318800236.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden: Blackwell Publishing.

Poon, O., Squire, D., Kodama, C., Byrd, A., Chan, J., Manzano, L., . . . Bishundat, D. (2016). A critical review of the model minority myth in selected literature on Asian Americans and Pacific Islanders in higher education. *Review of Educational Research, 86*(2), 469–502. https://doi.org/10.3102/0034654315612205.

Sword, H. (2011). *Stylish academic writing*. Cambridge: Harvard University Press.

Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research, 88*(5), 752–794. https://doi.org/10.3102/0034654318791582.

Winchester, C. L., & Salji, M. (2016). Writing a literature review. *Journal of Clinical Urology, 9*(5), 308–312.

Zinsser, W. (1998). *On writing well* (6th ed.). New York: HarperPerennial.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Part II

# Examples and Applications

# Conceptualizations and Measures of Student Engagement: A Worked Example of Systematic Review

Joanna Tai, Rola Ajjawi, Margaret Bearman and Paul Wiseman

## 1 Introduction

This chapter provides a commentary on the potential choices, processes, and decisions involved in undertaking a systematic review. It does this through using an illustrative case example, which draws on the application of systematic review principles at each stage as it actually happened. To complement the many other pieces of work about educational systematic reviews (Gough 2007; Bearman et al. 2012; Sharma et al. 2015), we reveal some of the particular challenges of undertaking a systematic review in higher education. We describe some of the 'messiness', which is inherent when conducting a systematic review in a domain with inconsistent terminology, measures and conceptualisations. We also describe solutions—ways in which we have overcome these particular challenges, both in this particular systematic review and in our work on other, similar, types of reviews.

J. Tai (✉) · R. Ajjawi · M. Bearman
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Geelong, Australia
e-mail: joanna.tai@deakin.edu.au

R. Ajjawi
e-mail: rola.ajjawi@deakin.edu.au

M. Bearman
e-mail: margaret.bearman@deakin.edu.au

P. Wiseman
Melbourne Centre for the Study of Higher Education (CSHE), University of Melbourne, Melbourne, Australia
e-mail: paul.wiseman@unimelb.edu.au

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_6

The chapter firstly introduces the topic of 'student engagement' and explains why a review was decided appropriate for this topic. The chapter then provides an exploration of the methodological choices and methods we used within the review. Next, the issues of results management and presentation are discussed. Reflections on the process, and key recommendations for undertaking systematic reviews on education topics are made, on the basis of this review, as well as the authors' prior experiences as researchers and authors of review papers. The example sections are bounded by a box.

## 2 First Steps: Identifying the Area for the Systematic Review

Student engagement is a popular area of investigation within higher education, as an indicator of institutional and student success, and as a proxy for student learning (Coates 2005). In initial attempts to understand what was commonly thought of as student engagement within the higher education literature, one of the authors (JT) found both a large number of studies, and a wide variation in the ways of both conceptualising and investigating student engagement. We hypothesised that it was unlikely that studies were focussed on exactly the same concept of student engagement given the variety already noted, and surmised that ways to investigate student engagement must also be differing, dependent upon the conceptualisation held by the researchers conducting the investigation. Our motivations at this stage were to successfully make an advance on the current plethora of publications to identify and outline some directions for future research, which we ourselves might be able to partake in.

Systematic reviews are seen as a means of understanding the literature in a field, particularly for doctoral students and early-career researchers, as a broad familiarity with the literature will be required for research in the area (Pickering and Byrne 2013; Olsson et al. 2014). Systematic reviews are particularly valuable when they create new knowledge or new understandings of an area (Bearman 2016). Furthermore, systematic reviews are less likely to suffer from criticisms faced by narrative or other less rigorous review processes, and are thus likely to doubly serve researchers in their ability to be published. Thus, choosing to do a systematic review on student engagement appeared to be a logical choice, serving two practical purposes: firstly, for the researchers themselves to gain a better understanding of the research being done in the field of student engagement; and secondly, to advance others' understanding through being able to share the results of such a literature review, in a publishable research output. At the time of

writing, we have shared our preliminary findings at a research conference (Tai et al. 2018), and will submit a journal article for publication in the near future.

We justified our choice to commence a broad systematic review on student engagement as follows:

**Overview**

Student engagement is a popular area of investigation within higher education, as an indicator of institutional and student success, and as a proxy for student learning (Coates 2005). In the marketisation of higher education, it is also seen as a way to measure 'customer' satisfaction (Zepke 2014). Student engagement has been conceptualised at a macro, organisational level (e.g. the National Survey of Student Engagement (NSSE) in the United States, and its counterparts the United Kingdom Engagement Survey and the Australasian Survey of Student Engagement) where a student's engagement is with the entire institution and its constituents, through to meso or classroom levels, and micro or task levels which focus more on the granularity of courses, subjects, and learning activities and tasks (Wiseman et al. 2016).

Seminal conceptual works describe student engagement as students "participating in educational practices that are strongly associated with high levels of learning and personal development" (Kuh 2001, p. 12), with three fundamental components: behavioural engagement, emotional engagement, and cognitive engagement (Fredricks et al. 2004). This work has a strong basis within psychological studies, with some scholars relating engagement to the idea of 'flow' (Csikszentmihalyi 1990), where engagement is an absorbed state of mind in the moment. These types of ideas have also been taken up within the work engagement literature (Schaufeli 2006). More recent conceptual work has progressed student engagement to be recognised as a holistic concept encompassing various states of being (Kahu 2013). In this conceptualisation, there are still strong links to student success, but students must be viewed as existing within a social environment encompassing a myriad of contextual factors (Kahu and Nelson 2018). Post-humanist perspectives on student engagement have also been proposed, where students are part of an assemblage or entanglement with their educators, peers, and the surrounding environment, and engagement exists in many ways between many different proponents (Westman and Bergmark 2018).

Though previous review work had been done in the area of student engagement in higher education, these reviews have taken a more selective approach with a view to development of broad conceptual understanding

without any quantification of the variation in the field (Kahu 2013; Azevedo 2015; Mandernach 2015). If we were to selectively sample, even with a view to diversity, we would not be able to say with any certainty that we had captured the full range of ways in which student engagement is researched within higher education. Thus, a systematic review of the literature on student engagement is warranted.

## 3 Determining the Function of the Systematic Review and Formulating Review Questions

Acknowledging these variety of conceptualisations already present within the field, we decided that clarity on conceptions, and also clarity on which types of measures and ways of investigating student engagement would be helpful in understanding what research had already occurred. Secondly, it seemed logical that investigating the alignment between the conceptualisation and measures of engagement might be a good place to devote our efforts, to also understand their relationships to student engagement strategies.

The decision to focus on classroom level measures was made for three reasons. First, this seemed to be the level with most confusion. Second, there seemed to be less stability and consistency in conceptualisations and measures as compared to the institutional level measures (i.e. national surveys of student engagement). Third, we felt that by investigating the classroom level, our findings were most likely to have potential to effect change for student engagement at a level which all students experience (as opposed to out-of-class engagement in social activities).

The review in this example borrows from the approach to synthesis previously used in work on mentoring (Dawson 2014), rubrics (Dawson 2015) and peer assessment (Adachi et al. 2018) to investigate and synthesise the design space of a term which has been used to describe many different practices. This involves reading a wide range of literature to identify diversity and similarity. In the case of this systematic review, there is more known about the conceptualisations but less understanding of the measures of student engagement. This approach to the systematic review search allows for additional understanding of the popularity of conceptions and measurement designs.

Therefore, a broad approach to understanding the field was taken, resulting in the research questions being "open"—i.e. beginning with a "how, what, why" rather than asking "does X lead to Y?"

In this study we aimed to answer the following research questions, in relation to empirical studies of engagement undertaken in classroom situations in higher education:

1. how is student engagement conceptualised?
2. how is student engagement investigated or measured?
3. what is the alignment between espoused conceptualisations of student engagement, and the conceptualisation of measures used?

## 4 Searching, Screening and Data Extraction

A protocol is usually developed for the systematic review: this stems from the clinical origins of systematic review, but is a useful way to set out *a priori* the steps taken within the systematic strategy. The elements we discuss below may need some piloting, calibration, and modification prior to the protocol being finalised. Should the review need to be repeated at any time in the future, the protocol is extremely useful to have as a record of what was previously done. It is also possible to register protocols through databases such as PROSPERO (https://www.crd.york.ac.uk/prospero/), an international prospective register.

### 4.1 Search Strategy

University librarians were consulted regarding both search term and database choice. This was seen as particularly necessary as the review intended to span all disciplines covered within higher education. As such, PsycINFO, ERIC, Education Source, and Academic Search Complete were accessed via Ebscohost simultaneously. This is a helpful time-saving option, to avoid having to input the search terms, and export citations in several independent databases. Separate searches were also conducted via Scopus and Web of Science to cover any additional journals not included within the former four databases.

## 4.2 Search Terms

A commonly used strategy to determine search terms is the PICO framework, taken from evidence-based medicine (Sharma et al., 2015). "P" stands for the people, group or subject of interest; "I" stands for the intervention, "C" for a comparison intervention or group, and "O" for outcome(s), which are of interest. However, in educational reviews, some of these categories are less useful, as a review might be taken to determine the range of outcomes (rather than a particular outcome), and comparison groups are not always used due to the potential inequalities in delivering an educational intervention to one group, and not another. If the systematic review seeks to establish what is known about a topic, then studies without interventions may also be helpful to include.

Prior to determining the final search terms and databases, a significant amount of scoping was undertaken, i.e. trial searches were run to gauge the number and type of citations returned. This was necessary to ensure that the search terms selected captured an appropriate range of data, and that the databases chosen indexed sufficiently different journals, so that the returned citations were not a direct duplicate. A key part of the scoping was ensuring that papers we had independently identified as being eligible for inclusion, were returned within the searches conducted. This made us more confident that we would capture appropriate citations within the searches that we did conduct.

Scoping also demonstrated that 'engagement' was a commonly used term within the higher education literature. Search terms therefore needed to be sufficiently specific to avoid screening excessively large numbers of papers. The first and second search strings focussed on the subject of interest; while the third string specified the types of studies we were interested. We added the fourth search string to ensure we were only capturing studies focussed at the classroom level, rather than institutional measures of engagement, and this was done after using the first three strings yielded a number of citations that was deemed too large for the research team to successfully screen in a reasonable amount of time.

**Search terms used**

1. ("student engagement" or "learner engagement")
   AND
2. ("higher education" or universit* or college* or post secondary or post-secondary)
   AND

3. (measur* or evidenc* or evaluat* or assess* or concept* or experiment*)
   AND
4. (classroom or online or blend* or distanc* or "face to face" or "virtual")

## 4.3 Determining Criteria for Inclusion and Exclusion

In the search databases, returned citations were filtered to only English language. We had set the time period to be from 2000 to 2016, as scoping searches revealed that articles using the word 'engagement' in higher education pre 2000 were not discussing the concept of student engagement. This was congruent with the NSSE coming into being in 2001. Throughout the screening process, the following inclusion and exclusion criteria were applied.

**Inclusion**
Higher education, empirical, educational intervention or correlational study, measuring engagement, online/blended and face-to-face, must be peer reviewed, classroom-academic-level, pertaining to a unit or course (i.e. classroom activity), 2000 and post, English, undergrad and postgrad.
**Exclusion** pre-2000, K-12, not empirical, not relevant to research questions, institutional level measures/macro level, not English, not available full-text, not formally peer reviewed (i.e. conference papers, theses and reports), only measures engagement as part of an instrument which is intended to investigate another construct or phenomenon, is not part of a course or unit which involves classroom teaching (i.e. is co- or extra-curricular in nature).

## 4.4 Revision of Inclusion and Exclusion Criteria

While the inclusion and exclusion criteria are now presented as a final list, there was some initial refinement of inclusion and exclusion criteria according to our big picture idea of what should be included, through testing them with an initial batch of papers as part of the researcher decision calibration process. This refined our descriptions of the criteria so that they fully aligned with what we were including or excluding.

## 5 Citation Management

A combination of tools was used to manage citations across the life of the project. Citation export from the databases was performed to be compatible with EndNote. This allowed for the collation of all citations, and use of the EndNote duplicate identification feature. The compiled EndNote library was then imported into Covidence, a web-based system for systematic review management.

### 5.1 Using Covidence to Manage the Review

Covidence (www.covidence.org) is review management software which was developed to support Cochrane, a non-profit organisation, which organises medical research findings, to provide higher levels of evidence for medical treatments. As such, it takes a default quantitative and medical approach to reviews of the literature, especially at the quality assessment and data extraction stage. However, the templates within Covidence can be altered to suit more qualitative review formats. The system has several benefits: it is web-based, so it can be used anywhere, on any device that has an Internet connection. The interface is simple to use and allows access to full-texts once they are uploaded. This means that institutional barriers to data sharing do not limit researchers. Importantly, Covidence tracks the decisions made for each citation, and automatically allows for double handling at each stage. It tracks the activity of each researcher so individual progress on screening and data extraction can be monitored. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, www.prisma-statement.org) diagram can be generated for the review, demonstrating numbers for each stage of the review. While there is a 'trial' option, which affords access to the system, for full team functionality a subscription is required.

### 5.2 Citation Screening

A total of 4192 citations were identified through the search strategy. Given the large number of citations and the nature of the review, the approach to citation screening focused on establishing up-front consensus and calibrating the decisions of researchers, rather than double-screening all citations at all steps of the process. This pragmatic approach has previously been used, with the key requirement of sensitivity rather than specificity, i.e. papers are included rather than

excluded at each stage (Tai et al. 2016). We built upon this method in a series of pilot screenings for each stage, where all involved researchers brought papers that they were unsure about to review meetings. The reasons for inclusion or exclusion were discussed in order to develop a shared understanding of the criteria, and to come to a joint consensus.

**Overview**

For initial title and abstract screening, two reviewers from the team screened an initial 200 citations, and discrepancies discussed. Minor clarifications were made to the inclusions and exclusion criteria at this stage. A further 250 citations were then screened by two reviewers, where 15 discrepancies between reviewers were identified, which arose from the use of the "maybe" category within Covidence. Based on this relative consensus, it was agreed that individual reviewers could proceed with single screening (with over 10% of the 4192 used as the training sample), where citations for which a decision could not be made based on title and abstract alone, passed on to the next round of screening.

1079 citations were screened at the full-text level. Again, an initial 110 or just over 10% were double reviewed by two of the review team. Discrepancies were discussed and used as training for further consensus building and refinement of exclusion reasons at this level, as Covidence requires a specific reason for each exclusion at this level. 260 citations remained at the conclusion of this stage to commence data extraction.

## 5.3 Determining the Proportion of Citations Used in Calibration

While the initial order of magnitude of citations for this review was not large, we were also cognisant that there would be a substantial number of papers included within the review. At each screening stage, an estimate of the yield for that stage was made based on the initial 10% screening process. Given the overall large numbers, and human reviewers involved, we determined that a 10% proportion for this review would be sufficient to train reviewers on inclusion and exclusion criteria at each stage. For reviews with smaller absolute numbers, a larger proportion for training may be required.

This review also employed a research assistant for the early phases of the review. This was extremely helpful in motivating the review team and keeping

track of processes and steps. The initial searching and screening phases of the review can be time-consuming and so distributing the workload is conducive to progress.

## 6 Data Extraction

Similar to the citation screening, we refined and calibrated our data extraction process on a small subset of papers, firstly to determine appropriate information was being extracted, and secondly to ensure consistency in extraction within the categories.

**Overview**

Data were extracted into an Excel spread sheet. In addition to extracting standard information around study information (country, study context, number of participants, aim of study/research questions, brief summary of results), the information relating to the research questions (conceptualisations and measures) were extracted, and also coded immediately. Codes were based on common conceptualisations of engagement however additional new codes could also be used where necessary. Conceptualisations were coded as follows, with multiple codes used where required:

- behavioural
- cognitive
- emotional
- social
- flow
- physical
- holistic
- multi-dimensional
- unclear
- work engagement
- other
- n/a

Five papers were initially extracted by all reviewers, with good agreement, likely due to all reviewers being asked to copy the relevant text from papers verbatim into the extraction table where possible. Further citations

were then split between the reviewing team for independent data extraction. During the process, an additional number of papers were excluded: while at a screening inspection they appeared to contain relevant information, extraction revealed they did not meet all requirements. The final number of included studies was 186.

### 6.1 Data Extraction Templates

While Covidence now has the ability to extract data into a custom template, at the time of the review, this was more difficult to customise. Therefore, a Microsoft Excel spread sheet was used instead. This method also came with the advantages of being able to sort and filter studies on characteristics where categorical or numerical data was input, e.g. study size, year of study, or type of conceptualisation. This aids with initial analysis steps. Conditional functions and pivot tables/pivot charts may also be helpful to understand the content of the review.

## 7 Data Analysis

Analysis methods are dependent on the data extracted from the papers; in our case, since we extracted largely qualitative information, much of the analysis was aimed to describe the data in a qualitative manner.

**Overview**

Simple demographic information (study year, country, and subject area) was tabulated and graphed using Excel functionalities. A comparison of study and measure conceptualisations was achieved through using the conditional (IF) function in Excel; this was also tabulated using the PivotChart function.

Study conceptualisations of engagement were further read to identify references used. A group of conceptualisations had been coded as "unclear"; these were read more closely to determine if they could be reassigned to a particular conceptualisation. For those conceptualisations that this was not possible for, their content was inductively coded. Content analysis was also applied to the information extracted on measures used

within studies to compile the range of measures used across all studies, and descriptions generated for each category of measure.

## 8 Reporting Results

Some decisions need to be made about which data are presented in a write-up of a review, and how they are presented. Demographic data about the country and discipline in which the study was conducted was useful in our review to contrast the areas from which student engagement research originated. Providing an overview of studies by year may also give an indication of the overall emergence or decline of a particular field.

There was a noticeable increase in papers published from 2011 onwards, with multiple papers from the USA (101), Canada (17), the UK (17), Australia (11), Taiwan (10), China (5). STEM disciplines contributed the greatest number of papers (46), followed by a group, which did not clearly list a discipline (41), then Health (35), Arts, Humanities and Social Sciences (22), and Business and Law (16). Education contributed 11 papers, and 15 additional papers were cross-disciplinary.

Importantly, the results need to be meaningful in terms of research questions, in providing some answers to the questions originally posed. Depending on the type of analysis undertaken, this may take many forms. It is also customary to include a "mother" table to accompany the review. This table records all included citations, and their relevant extracted information, such as when the study was carried out, a description of participants, number of participants, context of the study, aims and objectives, and findings or outcomes related to the research questions. This table is helpful for readers who wish to seek out particular individual studies.

## 9 Reflections on the Review Process

There are several key areas, which we wish to discuss in further depth, representing the authors' reflections on and learning from the process of undertaking a systematic review on the topic of student engagement. We feel a more lengthy

discussion of the problematic issues around the processes may be helpful to others, and we make recommendations to this effect.

## 9.1 Establishing Topic and Definitional Clarity

The research team spent a considerable amount of time discussing various definitions of engagement as we needed conceptual clarity in order to decide which articles would be included or excluded in the review, and to code the data extracted from those articles. Yet the primary reason for doing this systematic review was to better understand the range (or diversity) of views in the literature. Our sometimes circular conversations eventually became more iterative as we became more familiar with the common patterns and issues within the engagement literature. We used both popular conceptualisations and problematic exemplars as talking points to generate guiding principles about what we would rule in or rule out. Some decisions were simple, such as the context. For example, with our specific focus on higher education, it was obvious to rule out an article that was situated in vocational education. Definitions of engagement however were a little more problematic. As our key purpose was to describe and compare the breadth of engagement research, we needed to include many different perspectives as possible. This included articles that we as individual researchers may not have accepted as legitimate or relevant research of the engagement concept.

Having a comprehensive understanding of the breadth of the literature might seem obvious, but all members of our team were surprised at how many different approaches to engagement we found. Often, experienced researchers doing systematic reviews will be well versed in the literature that is part of, or closely related to, their own field of study, but systematic reviews are often the province of junior researchers with less experience and exposure to the field of inquiry as they undertake honours or masters research, or work as research assistants. For this reason, we feel that stating the obvious and recommending due diligence in pre-reading within the topic area is an essential starting point.

## 9.2 Review Aims: Identifying a Purpose

We found several papers that had attempted to provide some historical context or a frame of reference around the body of literature that were helpful in developing our own broad schema of the extant literature. For example, Vuori (2014), Azevedo (2015), and Kahu (2013) all noted the conceptual confusion around stu-

dent engagement which was borne out in our investigation. Such papers were useful in helping the research team to gain a broad perspective of the field of enquiry. At this point, we needed to make decisions about what we wanted to investigate. We limited our search to empirical papers, as we were interested in understanding what empirical research was being conducted and how it was being operationalised. It would have been a simpler exercise if we had picked a few of the more popular or well-defined conceptualisations of engagement to focus upon. This would have resulted in more well-defined recommendations for a composite conceptualisation or a selection of 'best practice' conceptualisations of student engagement, however this would have required the exclusion of many of the more 'fuzzy' ideas that exist in this particular field. We chose instead to cast our research net wide and provide a more realistic perspective of the field, knowing that we would be unlikely to generate a specific pattern that scholars could or should follow from this point forward. The result of this decision was, we hope, to provide a comprehensive understanding of the student engagement corpus and the complexities and difficulties that are embedded in the research to date. However, we note that our broad approach does not preclude a more narrow subsequent focus now the data set has been created.

Researchers should be clear from the beginning on what the research goals will be, and to continue to iterate the definitional process to ensure clarity of the concepts involved and that they are appropriately scoped (whether narrow or wide) to achieve the objectives of the review. In the case of this systematic review on student engagement in higher education, the complex process of iterating conceptual clarity served us well in exposing and summarising some of the complex problems in the engagement literature. However, if our goal had been to collapse the various definitions into a single over-arching conception of engagement, then we would have needed a narrower focus to generate any practical outcome.

## 9.3 Building and Expanding Understanding

As we worked our way through the multitude of articles in this review, we developed an iterative model where we would rule papers as clearly in, clearly out, and a third category of 'to be discussed'. Having a variety of views of engagement amongst our team was particularly useful as we were able to continually challenge our own assumptions about engagement as we discussed these more problematic articles. Our experience has led us to think that an iterative process can be useful when the scope of the topic of investigation is unclear. This allowed us to continually improve and challenge our understanding of the topic as we slowly

generated the final topic scope through undertaking the review process itself. When the topic of investigation is already clearly defined and not in debate, this process may not be required at initial stages of scoping. If this describes your project, dear reader, we envy you. Having a range of views within the investigative team was however helpful in assuring we did not simply follow one or more of the popular or prolific models of engagement, or develop confirmation bias, especially during the analytical stages: data interpretation may be assisted through the input of multiple analysts (Varpio et al. 2017). If agreement in inclusion or subsequent coding is of interest, inter-rater reliability may be calculated through a variety of methods. Cohen's kappa co-efficient is a common means of expressing agreement, however the simplest available method is usually sufficient (Multon and Coleman 2018). In our work, establishing shared understanding has been more important given the diversity of included papers and so we did not calculate an inter-rater reliability.

## 9.4 Choosing an Appropriate Type of Review Method

Given the heterogeneity of the research topic and the revised aim of documenting the field in all its diversity, the type of review conducted (in particular the extraction and analysis phase) shifted in nature. We had initially envisioned a qualitative synthesis where we would consolidate the where we could draw "conclusions regarding the collective meaning of the research" (Bearman and Dawson 2013, p. 252). However, as described already, coming to a consensus on a single conceptualisation of student engagement was deemed futile early on in the review. Instead we sought to document the range of conceptualisations and measures used. What was needed here then was more of a content rather than thematic analysis and synthesis of the data. Content analysis is a family of analytic methods for identifying and coding patterns in data in replicable and systematic ways (Hsieh and Shannon 2005). This approach is less about abstraction from the data but still involved interpretation. We used a directed content analysis method where we iteratively identified codes (using pre-existing theory and those derived from the empirical studies) and then using these to categorise the data systematically then counting occasions of the presence of each code. The strength of a directed approach is that existing theory (in our case conceptualisations of student engagement) can be supported and extended (Hsieh and Shannon 2005). Although seemingly straightforward, the research team needed to ensure consistency in our understanding of each conception of student engagement through a codebook and multiple team meetings where definitional issues are discussed and

ambiguities in the papers declared. Having multiple analysts who bring different lenses to bear on a research phenomenon, and who discuss emerging interpretations, is often considered to support a richer understanding of the phenomena being studied (Shenton 2004). However, in this case perhaps what mattered more was convergence rather than comprehensiveness.

## 9.5 Ensuring Ongoing Motivation to Undertake the Review

There are several difficult steps at any stage of a systematic review. The first is to finalise the yield of the articles. This was a large systematic review with—given our broad focus on conceptualisations—an extensive yield. We employed help from a research assistant to assist with the initial screening process at title and abstract level but we needed deep expertise as to what constituted engagement and (frequently) research when making final decisions about including full texts. This is a common mistake in systematic reviews: knowing the subject domain is essential to making nuanced decisions about yield inclusions. Strict inclusion and exclusion criteria do not mean that a novice can make informed judgements about how these criteria are met. This meant that we, with a more expert view of student engagement, all read an extremely large number of full texts—819 collectively—these needed to be read, those included had to have data extracted, and then the collective meaning of this data needed to be discussed against the aims of our review.

This was unquestionably, a dull and uninspiring task. The paper quality was poor in this particular systematic review relative to others we have conducted. As noted, engagement is by its nature difficult to conceptualise and this clearly caused problems for research design. In addition, while we are interested in engagement, we are less interested in the particular classroom interventions that were the focus of many papers. We found ourselves reading papers that often lacked either rigour or inherent interest to us. One way we surmounted this task was setting a series of deadlines and associated regular meetings where we met and discussed particular issues such as challenges in interpreting criteria and papers.

Motivation can be a real problem for systematic review methodology. Unlike critical reviews, the breadth of published research can mean wading through many papers that are not interesting to the researcher or of generally poor quality. It is important to be prepared. And it is also important to know that time will not be kind to the review. Most systematic reviews need to be relatively up-to-date

at the time of acceptance for publication, so the review needs to be completed within a year if at all possible.

The next motivational challenge in our experience of this systematic review is the data extraction. While the data sets were somewhat smaller (260), this was still a sizeable effort. Within each paper we were required to locate conceptualisations and measures of engagement—which were often scattered throughout the paper—and categorise these according to our agreed criteria. In the process of extraction we identified several papers again which did not adhere to our inclusion criteria, resulting in a final yield of 186 papers. Maintaining uniformity of interpretation and extraction was a matter of constant iterative discussion and again, this task, was impossible without a deep understanding of engagement as well as qualitative and quantitative methods. We found ourselves scheduling social arrangements at the end of some of our meetings, to keep on task.

Finally, we needed to draw some conclusions from the collated data from a sizeable number of papers. Throughout this process, we found that returning to the fundamental purpose of the review acted as a lodestar. We could see that the collected weight of the papers was suggesting that there were significant challenges with how engagement research was being enacted, and that there were important messages about how things could be improved. One thing we struggled with is the point that everyone else also finds difficult. That is, what is the nature of engagement? In what ways can we productively conceptualise it and then, possibly more controversially, measure it? Within this framing, it has been difficult to come to some conclusions based on the results we have produced. While in some ways, this appears the last part of the marathon, it presents a very steep challenge indeed.

## 10 Recommendations to Prospective Researchers

Systematic review methods add rigour to the literature review process, and so we would recommend, where possible, and warranted, that a systematic review be considered. Such reviews bring together existing bodies of knowledge to enhance understanding. We highlight the following points to those considering undertaking a systematic review:

- Clarity is important to remain consistent throughout the review: This may require the researchers developing significant familiarity with the topic of the review: an iterative process may be helpful to narrow the scope of the review through ongoing discussion.

- Processes may be emergent: Despite best efforts to set out a protocol at the commencement of the review process, the data itself may determine what occurs in the extraction and analysis stages. While the objectives of the review may remain constant, the way in which the objectives are achieved may be altered.
- Motivation to persevere is required: Systematic literature reviews generally take longer than expected, given the size of team required to tackle any topic of a reasonable size. The early stages in particular can be tedious, so setting concrete goals and providing rewards may improve the rate of progress.

## References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development*, *37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913.

Azevedo, R. (2015). Defining and measuring engagement and learning in science: conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069.

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, *31*(5), 625–640. https://doi.org/10.1080/07294360.2012.702735.

Bearman, M. (2016).Quality and literature reviews: beyond reporting standards. *Medical Education*, 50(4), 382–384. https://doi.org/10.1111/medu.12984.

Bearman, M. & Dawson, P. (2013). Qualitative synthesis and systematic review in health professions education. *Medical Education*, *47*(3), 252–60. https://doi.org/10.1111/medu.12092.

Coates, H. (2005).The value of student engagement for higher education quality assurance. *Quality in Higher Education*, *11*(1), 25–36. https://doi.org/10.1080/13538320500074915.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper & Row.

Dawson, P. (2014). Beyond a definition: toward a framework for designing and specifying mentoring models. *Educational Researcher*, *43*(3), 137–145. https://doi.org/10.3102/0013189x14528751.

Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, *42*(3), 1–14. https://doi.org/10.1080/02602938.2015.1111294.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevant of evidence. *Research Papers in Education*, *22*(2), 213–228.

Hsieh, H.-F. & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687.

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758–773. https://doi.org/10.1080/03075079.2011.598505.

Kahu, E. R. & Nelson, K. (2018). Student engagement in the educational interface: understanding the mechanisms of student success. *Higher Education Research & Development*, *37*(1), 58–71. https://doi.org/10.1080/07294360.2017.1344197.

Kuh, G. D. (2001). Assessing what really matters to student learning *Inside The National Survey of Student Engagement*, *Change: The Magazine of Higher Learning*, *33*(3), 10–17. https://doi.org/10.1080/00091380109601795.

Mandernach, B. J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *International Journal of Learning, Teaching and Educational Research*, *12*(2), 1–14. Retrieved February 12, 2018 from http://www.ijlter.org/index.php/ijlter/article/view/367.

Multon, K. D., & Coleman, J. S. M. (2018). Inter-Rater Reliability. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 863–865). Thousand Oaks, CA: SAGE Publications. https://doi.org/10.4135/9781506326139.n344.

Olsson, C., Ringnér, A., & Borglin, G. (2014). Including systematic reviews in PhD programmes and candidatures in nursing—"Hobson's choice"? *Nurse education in practice*, *14*(2), 102–5. https://doi.org/10.1016/j.nepr.2014.01.005.

Pickering, C. & Byrne, J. (2013). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development*, *33*(3), 534–548. https://doi.org/10.1080/07294360.2013.841651.

Schaufeli, W. B. (2006). The measurement of work engagement with a short questionnaire: a cross-national study. *Educational and Psychological Measurement*, *66*(4), 701–716. https://doi.org/10.1177/0013164405282471.

Sharma, R., Gordon, M., Dharamsi, S., & Gibbs, T. (2015). Systematic reviews in medical education: a practical approach: AMEE Guide 94., *Medical Teacher*, 37(2), 108–124. https://doi.org/10.3109/0142159x.2014.970996.

Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*(2), 63–75. Retrieved on April 10, 2019 from https://www.researchgate.net/publication/228708239_Strategies_for_Ensuring_Trustworthiness_in_Qualitative_Research_Projects.

Tai, J., Ajjawi, R., Bearman, M., & Wiseman, P. J. (2018). Exploring the mismatch between conceptualisations and measures of student engagement. In *EARLI-SIG1*. Helsinki: European Association for Research on Learning and Instruction: Special Interest Group 1, Assessment and Measurement Conference, 29–31 August.

Tai, J., Molloy, E., Haines, T., & Canny, B. (2016). Same-level peer-assisted learning in medical clinical placements: A narrative systematic review. *Medical Education*, *50*(4), 469–484. https://doi.org/10.1111/medu.12898.

Varpio, L., Ajjawi, R., Monrouxe, L. V., O'Brien, B. C., & Rees, C. E. (2017). Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Medical Education*, *51*(1), 40–50. https://doi.org/10.1111/medu.13124.

Vuori, J. (2014). Student engagement: buzzword of fuzzword? *Journal of Higher Education Policy and Management*, *36*(5), 509–519. https://doi.org/10.1080/1360080x.2014.936094.

Westman, S. & Bergmark, U. (2018). Re-considering the ontoepistemology of student engagement in higher education. *Educational Philosophy and Theory*, https://doi.org/10.1080/00131857.2018.1454309.

Wiseman, P. J., Kennedy, G. E. & Lodge, J. M. (2016). Models for understanding student engagement in digital learning environments. In *Proceedings of Ascilite 2016, Show Me The Learning, Adelaide, November 27–30* (pp. 666–671). Retrieved April 10, 2019 from http://2016conference.ascilite.org/wp-content/uploads/ascilite2016_wiseman_concise.pdf.

Zepke, N. (2014). Student engagement research in higher education: questioning an academic orthodoxy. *Teaching in Higher Education*, *19*(6), 697–708. https://doi.org/10.1080/13562517.2014.901956.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Learning by Doing? Reflections on Conducting a Systematic Review in the Field of Educational Technology

Svenja Bedenlier, Melissa Bond, Katja Buntins, Olaf Zawacki-Richter and Michael Kerres

## 1 Introduction

In 1984, Cooper and Hedges stated that

> "scientific subliteratures are cluttered with repeated studies of the same phenomena. Repetitive studies arise because investigators are unaware of what others are doing, because they are skeptical about the results of past studies, and/or because they wish to extend…previous findings…[yet even when strict replication is attempted] results across studies are rarely identical at any high level of precision, even in the physical sciences…" (p. 4 as cited in Mullen and Ramírez 2006, p. 82–83).

S. Bedenlier (✉) · M. Bond · O. Zawacki-Richter
Centre for Open Education Research (COER), Institute of Education, Carl von Ossietzky University Oldenburg, Oldenburg, Germany
e-mail: svenja.bedenlier@uni-oldenburg.de

M. Bond
e-mail: melissa.bond@uni-oldenburg.de

O. Zawacki-Richter
e-mail: olaf.zawacki.richter@uni-oldenburg.de

K. Buntins · M. Kerres
Department of Educational Sciences, Learning Lab, University Duisburg Essen, Duisburg/Essen, Germany
e-mail: katja.buntins@uni-due.de

M. Kerres
e-mail: michael.kerres@uni-due.de

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_7

Presumably due to the reasons cited here, systematic reviews have recently garnered interest in the field of education, including the field of educational technology (for example Joksimovic et al. 2018).

Following the presentation and discussion of systematic reviews as a method in the first part of this book, in this chapter, we outline a number of challenges that we encountered during our review on the use of educational technology in higher education and student engagement. We share and discuss how we either met those challenges, or needed to accept them as an unalterable part of the work. "We" in this context refers to our review team, comprised of three Research Associates with backgrounds in psychology and education, and with combined knowledge in quantitative and qualitative research methods, under the guidance of two professors from the field of educational technology and online learning. In the following sections, we provide contextual information of our systematic review, and then proceed to describe and discuss the challenges that we encountered along the way.

## 2 Systematic Review Context

Our systematic review was conducted within the research project *Facilitating student engagement with digital media in higher education* (ActiveLeaRn), which is funded by the German Federal Ministry of Education and Research as part of the funding line 'Digital Higher Education', running from December 2016 to November 2019. The second-order meta-analysis by Tamim et al. (2011) found only a small effect size for the use of educational technology for successful learning, herewith showing that technology and media do not make learning better or more successful per se. Against this background, we posit that educational technologies and digital media do have, however, the potential to make learning *different* and more *intensive* (Kerres 2013), depending on the pedagogical integration of media and technologies for learning (Higgins et al. 2012; Popenici 2013).

The use of educational technology has been found to have the potential to increase student engagement (Chen et al. 2010; Rashid and Asghar 2016), improve self-efficacy and self-regulation (Alioon and Delialioglu 2017; Northey et al. 2015; Salaber 2014), and increase participation and involvement in courses and within the wider institutional community (Alioon and Delialioglu 2017; Junco 2012; Northey et al. 2015; Salaber 2014). Given that disengagement negatively impacts on students' learning outcomes and cognitive development (Ma et al. 2015), and is related to early dropout (Finn and Zimmer 2012), it is crucial to investigate how technology has been used to increase engagement.

Departing from the student engagement framework by Kahu (2013), this systematic review seeks to identify the conditions under which student engagement

is supported through educational technology in higher education. Given that calls have been made for further investigation into how educational technology affects student engagement (Castañeda and Selwyn 2018; Krause and Coates 2008; Nelson Laird and Kuh 2005), as well as further consideration of the student engagement concept itself (Azevedo 2015; Eccles 2016), a synthesis of this research can provide guidance for practitioners, researchers, instructional designers and policy makers. The results of this systematic review will then be discussed with experts and practitioners in the field of (German) higher education, to validate or controversially discuss the findings, providing both an impetus for evidence-based practice in the field of technology-enhanced learning and to gain insights relevant for further research projects.

**Theory is one thing, practice another: What happened along the way**
Whilst in theory, literature on conducting systematic reviews provides guidance in quite a straightforward manner (e.g. Gough et al. 2017; Boland et al. 2017), potential challenges (even though mentioned in the literature) take shape only in the actual execution of a review. Coverdale et al. (2017) describe some of the challenges that we encountered from a journal editor's point of view. They summarize them as follows:

> "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" (p. 250).

In the remainder of this chapter, we will centre our discussion around three main aspects of conducting our review, namely two broad areas of challenges that we faced, as well as a discussion of the chances that emerged from our specific review experience.

## 3 Challenge One: Defining the Review Scope, Question and Locating Literature for Inclusion

### 3.1 Broad vs. Narrow Questions

Research questions are critical parts of any research project, but arguably even more so for a systematic review. They need "to be clear, well defined, appropriate, manageable and relevant to the outcomes that you are seeking" (Cherry and

Dickson 2014, p. 20). The review question that was developed in a three-day workshop at the EPPI Centre[1] at the University College London was: 'Under which conditions does educational technology support student engagement in higher education?'. This is a broad question, without very clearly defined components and thus, logically, impacted on all ensuing steps within the review. 'Conditions' could be anything and therefore could not be explicitly searched for, so we chose to focus on students and learning. 'Educational technology' can mean different things to different people, therefore we chose to search as broadly as possible and included a large amount of different technologies explicitly within the search string (see Table 1) as we will also discuss in the further sections. This was a question of sensitivity versus precision (Brunton et al. 2012). However, this then resulted in an extraordinary amount of initial references, and required more time to undertake screening.

Had we not had as many resources to support this review, and therefore time to conduct it, we could have used the PICO framework (Santos et al. 2007) to define our question. This allows a review to target specific populations (in this case 'higher education'), interventions (in this case 'educational technology'), comparators (e.g. face to face as compared to blended or online learning), and outcomes (in this case 'student engagement'). The more closed those PICO parameters, the more tightly defined and therefore the more achievable a review potentially becomes.

Reflecting on the initial question from our current standpoint, it was the right decision in order to approach this specific topic with its often times implicit understandings and definitions of concepts. The challenge to grasp the student engagement concept is very illustratively captured by Eccles (2016), stating that it is like "3 blind men describing an elephant" (p. 71), or, more neutrally described as an "umbrella concept" (Järvelä et al. 2016, p. 48). As will be detailed below, the lack of a clear-cut concept in the review question that could directly be addressed in a database search demanded a broader search in order to identify relevant studies. Subsequently, to address this broad research question appropriately, we paid the price of tremendously increasing the scope of the review and not being able to narrow it down to have a simple and "elegant" answer to the question.

[1] http://eppi.ioe.ac.uk

**Table 1** ActiveLeaRn systematic review search string

| Topic and cluster | Search terms |
|---|---|
| Student | Learner* OR student* |
| Higher education | "higher education" OR universit* OR college OR undergrad* OR graduate OR postgrad* NOT ("K-12" OR kindergarten OR "corporate training" OR "professional training" OR "primary school" OR "middle school" OR school OR "vocational education" OR "adult education") |
| Educational technology | "educational technology" OR "learning technology" OR "digital technology" OR "digital media" |
| Tools | "social media" OR "social network*" OR "social web" OR vodcast OR podcast* OR "digital broadcasting" OR blog* OR weblog OR "electronic publishing" OR microblog* OR "interactive whiteboard*" OR simulation OR forum* "computer-mediated communication" OR "computer communication network*" OR ePortfolio OR e-Portfolio OR e-Assessment OR eAssessment OR "computer-based testing" OR "computer-assisted testing" OR OER OR "open educational resources" OR "open access" OR "open source technology" OR "information and communication technolog*" OR "information technology" OR "social tagging" OR "app" OR tablet* OR "handheld device*" OR "mobile device*" OR "electronic books" OR eBooks |
| Internet | "Web 2.0" OR "user generated content" OR "cyber space" |
| Learning environments | "virtual classroom*" OR "personal learning environment*" OR "virtual learning environment" OR "virtual reality" OR "augmented reality" OR "learning management system*" |
| Computer | "computer-based learning" OR "computer-based instruction" OR "computer-supported learning" OR "computer-supported collaborative learning" OR "computer-supported cooperative learning" OR "computer-supported cooperative work" OR "computer-mediated learning" OR "computer-assisted instruction" OR "computer-assisted language learning" |
| Web | "web-enhanced learning" OR "web-enhanced instruction" OR "web-based training" OR "web-based instruction" OR MOOC OR "massive open online course*" OR "online instruction" OR "online education" |
| Technology | "technology-enhanced learning" OR "technology-mediated learning" |

(continued)

**Table 1** (continued)

| Topic and cluster | Search terms |
|---|---|
| Mobile | "mobile learning" OR m-Learning OR "mobile communication system*" OR "mobile-assisted language learning" OR "mobile computing" |
| E-Learning | eLearning OR e-Learning OR "electronic learning" OR "online learning" |
| Mode of delivery | "distance education" OR "blended learning" OR "virtual universit*" OR "open education" OR "online course*" OR "distance learning" OR "collaborative learning" OR "cooperative learning" OR "game-based learning" |

## 3.2 Student Engagement: Focus on a Multifaceted Concept

To further explain why both our question and especially our search string can be considered rather sensitive than precise in the understanding of Brunton et al. (2012), discussing the concept of student engagement is vital. Student engagement is widely recognised as a complex and multi-faceted construct, and also arguably constitutes an example of "'hard-to-detect' evidence" (O'Mara-Eves et al. 2014, p. 51). Prior reviews of student engagement have chosen to include the phrase 'engagement' in their search string (e.g. Henrie et al. 2015), however this restricts search results to only those articles including the term 'engagement' in the title or abstract. To us—and including our information specialist who assisted us in the development of the search string—the concept of student engagement is a broad and somewhat fuzzy term, resulting in the following, albeit common, challenge:

> "[T]he main focus of a review often differs significantly from the questions asked in the primary research it contains; this means that issues of significance to the review may not be referred to in the titles and abstracts of the primary studies, even though the primary studies actually do enable reviewers to answer the question they are addressing" (O'Mara-Eves et al. 2014, p. 50).

Would this line need to be moved up? Given the contested nature of student engagement (e.g. Appleton et al. 2008; Christenson et al. 2012; Kahu 2013), and the vast array of student engagement facets, the review team therefore felt

that this would seriously limit the ability of the search to return adequate literature, and the decision was made to leave any phrase relating to engagement out of the initial string. Instead, the engagement and disengagement facets that had been uncovered—published elsewhere (Bond and Bedenlier, 2019)—were used to search within the initial corpus of results.

## 3.3 Developing the Search String: Iterations and Complexity

Developing a search string, which is appropriate for the purpose of the review and ensures that relevant research can be identified, is an advanced endeavor in itself, as the detailed account by Campbell et al. (2018) shows. Resulting from our initially broad review question, we were subsequently faced with the task to create a search string that would reflect the possible breadth of both student engagement (facets) (see Bond and Bedenlier, 2019) as well as be inclusive of a diverse range of educational technology tools. The educational technology tools were, in the end, identified in a brainstorming session of the three researchers and the guiding professors; trying to be comprehensive whilst simultaneously realizing the limitations of this attempt. As displayed in the search string below, categories within educational technology were developed, which were then applied in different combinations with the student and higher education search terms, and were run in four different databases, that is ERIC, Web of Science, PsycINFO and SCOPUS.

Not only due to slight differences in the make up of the databases, e.g. different usage of truncations or quotation marks, but also grounded in misleading educational technology terms, the search string underwent several test runs and modifications before final application. Initially included terms such as "website" or "media" proved to be dead ends, as they yielded a large number of studies including these terms but that were off topic. Again, reflecting from today's point of view, the term "simulation" was also ambivalent, sometimes used in the understanding of our review as an educational technology tool, but often times also used for in-class role plays in medical education, without the use of further educational technology.

However, the broadness of the search string made it possible to identify research that, with a more precise search focusing on "engagement" would have been lost to our review—demonstrated in the simple fact that within our final corpus of 243 articles, only 63 studies (26%) actually employ the term "student engagement" in their title or abstract.

# 4 Challenge Two: Retrieving, Analyzing and Describing the Research

## 4.1 Accuracy of Title and Abstract

As we began to screen the titles and abstracts of the studies that met our predefined criteria (English language, empirical research, peer-reviewed journal articles, published between 2007–2016; focused on students in higher education, educational technology and student engagement), we quickly realized that the abstracts did not necessarily provide information on the study that we needed, e.g. whether it was an empirical study, or if the research population was students in higher education. This problem, dating back to the 1980s, was also mentioned by Mullen and Ramírez (2006, p. 84–85), and was addressed in the field of medical science by proposing guidelines for making abstracts more informative. Whilst we were cognizant of the problem of abstracts—and also keywords—being misleading (Curran 2016), there proved no way around this issue, and we subsequently included abstracts for further consideration that we thought unlikely to be on topic, but which could not be excluded due to the slight possibility that they might be relevant.

## 4.2 The Sheer Size of It…. Using a Sampling Strategy

As described in Borah et al. (2017), "the scope of some reviews can be unpredictably large, and it may be difficult to plan the person-hours required to complete the research" (p. 2). This applied to our review as well. Having screened 18,068 abstracts, we were faced with the prospect of screening 4152 studies on full text. This corresponds roughly to the maximum number of full texts to be screened (4385) in the study by Borah et al. (2017, p. 5) who analyzed 195 systematic reviews in the medical field to uncover the average amount of time required to complete a review. However, retrieval and screening of 4152 articles was not feasible for a part-time research team of three, within the allotted time and the other research tasks within the project. As a result of this challenge, it was decided that a sample would be drawn from the corpus, using the sample size estimation method (Kupper and Kafner 1989) and the R Package MBESS (Kelley et al. 2018). The sampling led to two groups of 349 articles each that would need to be retrieved, screened and coded. Whilst the sampling strategy was indeed a time saver, and the sample was representative of the literature in terms of geographical

representation, methodology and study population, the question remains as to the results we might have uncovered, had we had the resources to review the entire corpus.

## 4.3 Study Retrieval

Although authors such as Gough et al. (2017) mention that the retrieval of studies requires time and effort, this step in the review certainly assumed both time and human resources—and also a modest financial investment. We attempted to acquire the studies via our respective institutional libraries, or ordered them in hard copy via document delivery services, contacted authors via ResearchGate (with mixed results), and finally also took to purchasing articles when no other way seemed to work. However, we also had to realise that some articles would not be available, e.g. in one case, the PDF file in question could not be opened by any of the computers used, as it comprised a 1000 page document, which inevitably failed to load.

Trying to locate the studies required time. Some of the retrieval work was allocated to a student assistant whose searching skills were helpful for easy to retrieve studies, but this required us to follow up on harder to find studies. Thus, whilst the step of study retrieval might sound rather trivial on first sight, this phase actually evolved into a much larger consideration. As a consequence, we would strongly recommend to have this factored in attentively into the time line of the review execution, and particularly when applying for funding.

## 4.4 Using Software Within the Review

In order to manage a large corpus of literature, it is highly recommended to use software, in order to make the screening and coding steps easier in particular. Popular low cost options include using Excel spreadsheets, Google Sheets, or reference management software, such as Endnote, Citavi or Zotero. Spreadsheets are straightforward to use and are familiar applications, however they can result in an unwieldy amount of information on one screen at a time, and reference management software has limited filtering and coding functionality. Software that has been specifically designed for undertaking systematic reviews can therefore be a more attractive option, as their design can produce quick and easy reports, speeding up the synthesis and trend identification process. Rayyan (Ouzzani et al. 2016)

is a free web-based systematic review platform, which also has a mobile app for coding. However, we decided to use EPPI-Reviewer software, developed by the EPPI-Centre at the University College London.

Whilst not free, the software does have an easy-to-use interface, it can produce a number of helpful reports, and the support team is fantastic. However, more training in how to use the software was needed at the beginning to set the review up, and the lack thereof meant that we were not only learning on the job, but occasionally having to learn from mistakes. The way that we designed our coding structure for data extraction, for example, has now meant that we need to combine results in some cases, whereas they should have been combined from the beginning. This is all part of the iterative review experience, however, and we would now recommend spending more time on the coding scheme and thinking through how results would be exported and analyzed, prior to beginning data extraction.

Another area, where using software can be extremely helpful, is in the removal of duplicates across databases. We highly recommend importing the initial search results from the various databases (e.g. Web of Science, ERIC) into a reference management software application (such as Endnote or Zotero), and then using the 'Remove Duplicates' function. You can then import the reduced list into EPPI-Reviewer (or similar software) and run the duplicate search again, in case the original search missed something. This can happen due to the presence of capitals in one record but not in another, or through author or journal names being indexed differently in databases. We found this was the case with a vast number of records and that, despite having run the duplicate search multiple times, there were still some duplicates that needed to be removed manually.

## 4.5 Describing Studies

Against the backdrop of our review being very large, as well as employing an extensive coding scheme, we engaged in discussion of how to present a descriptive account of this body of research that would both meaningfully display the study characteristics, as well as take into account that even this description constitutes a valuable insight into the research on student engagement and educational technology. Finding guidance in the article by Miake-Lye et al. (2016) on "evidence maps" (p. 2), we decided to dedicate one article publication to a thorough description of our literature corpus, thereby providing a broad overview of the theoretical guidance, methods used and characteristics of the studies (see Bond et al., Manuscript in preparation), and then to write field of study-specific articles with the actual synthesis of results (e.g. Bedenlier, Bond et al., Forthcoming).

To handle the coded articles, all data and information were exported from EPPI-Reviewer into Excel to allow for necessary cross tabulations and calculations—and also to ensure being able to work with the data after the expiry of the user accounts in EPPI-Reviewer. Most interestingly, the evidence map—structured along four leading questions[2]—emerged to be a very insightful and helpful document, whose main asset was to point us towards a potentially well-suited framework for our actual synthesis work. Thus, following the expression 'less is more', the wealth of information, concepts and insights to be gained from the mere description of the identified studies is worth an individual account and presentation—especially if this helps to avoid an overladen article that can neither provide a full picture of the included research nor an extensive synthesis due to space or character constraints.

## 5 Chances

Whilst we encountered the challenges described here—and there are more, which we cannot include in this chapter—we were also lucky enough to have a few assets in conducting our review, which emerged from our specific project context and which we would also like to alert others to.

### 5.1 Involvement of the Information Specialist

As suggested in Beverley et al. (2003), information specialists can assume ten roles in a systematic review, comprising "traditional librarian responsibilities, such as literature searching, reference management and document supply, as well as a whole range of progressive activities, such as project leadership and management, critical appraisal, data extraction, data synthesis, report writing and dissemination" (p. 71). Whilst the same authors point out that information specialists

[2]What are the geographical, methodological and study population characteristics of the study sample? What are the learning scenarios, modes of delivery and educational technology tools employed in the studies? How do the studies in the sample ground student engagement and align with theory, and how does this compare across disciplines? Which indicators of cognitive, affective, behavioural and agentic engagement were identified due to employing educational technology? Which indicators of student disengagement?

are often consulted and involved in the more traditional tasks, this is also how we consulted the librarian in charge of our research field.

In our case, we were lucky enough to have an information specialist who not only attended the systematic review workshop jointly with us, but also played an integral part in setting up the search string—including making us cognizant of pitfalls such as potential database biases (e.g. ERIC being predominantly US-American focused), and the need to adapt search strings to different databases (e.g. changing truncations). On a general note, we can add that students and faculty who are seeking assistance in conducting systematic reviews increasingly frequent the research librarian for education at our institution. This not only shows the current interest in systematic reviews in education but also emphasizes the role that information specialists and research librarians can play in the course of appropriate information retrieval. It also relates back to Beverley's et al. (2003) discussion of information specialists engaging in various parts of the review—and strengthening their capacity beyond merely being a resource at the beginning of the review.

Thus, although researchers are familiar with searching databases and information retrieval, an external perspective grounded in the technical and informational aspect of database searches is helpful in order to carry out searches and understanding databases as such.

## 5.2 Multilingualism

Our team was comprised of five researchers; two project leaders, who joined the team in the crucial initiation and decision-making phase, and who provided in-depth content expertise based on the extensive knowledge of the field, as well as three Research Associates, who carried out the actual review. The three Research Associates are located at the two participating universities; University of Oldenburg and the University of Duisburg-Essen. Whilst Katja and Svenja are native speakers of German, Melissa is a native speaker of (Australian) English, which proved to be of enormous help in phrasing the nuances of the search string and defining the exact tone of individual words. However, Australian English differs from American, British and other English variations, which therefore has implications of context on certain phrases used.

Additionally, we now know that authors from Germany do not always use terms and phrases that are internationally compatible (e.g. "digital media in education" = digitale Medien in der Bildung), rather, terms have been developed that

are specific to the discourse in Germany (Buntins et al. 2018). A colleague also observed the same for the Spanish context. Both of these examples suggest a need for further discussion of how this influences the literature in the field and also how this potentially (mis)leads authors from these countries (and other countries as well) in their indexing of articles via author-given keywords. Thus, our different linguistic backgrounds alerted us to these nuances in meaning whilst this also raises the question about potential linguistic "blind spots" in monolingual teams. This could be a topic of further investigation.

### 5.3 Teamwork

Beyond the challenges that occurred at specific points in time, we would like to stress one asset that emerged clearly in the course of the (sometimes rather long) months we spent on our review: Working in a research and review *team*. We started out as a team who had not worked together before, and therefore only knew about each other's potentially relevant and useful abilities beforehand: quantitative and qualitative method knowledge, English native speaker and plans to conduct a PhD in the field of educational technology in K-12 education. In the course of the work, adding the function of a (rough) time keeper and also the negotiation of methodological perfection, rigor and practicability, emerged to be important issues that we solved within the team and that would have been hardly, if at all, solvable if the review had been conducted by a single person.

As every person in the team—as in all teams—brings certain abilities, it is the sum of individual competencies and the joint effort that enabled us to carry out a review of this size and scope. Thus, in the end, it was the constant negotiation, weighing the pros and cons of which way to go, and the ongoing discussions, that were the strongest contributor to us meeting the challenges encountered during the work and also successfully completing the work.

## 6 Hands-on Advice and Implications

The review has been—and continues to be—a large and dominating part within ActiveLeaRn. Working together as a team was the greatest motivational force and help in the conduct of this review, not only in regards to achieving the final write up of the review but also in regards to having learned from one another.

Going back to the title of this chapter "Learning by doing", we can confirm that this holds true for our experience. Although method books do provide help and guidance, they cannot fully account for the challenges and pitfalls that are individual to a certain review—hence all reviews, and all other research for that matter—are to some part learning by doing. And transferring what we learnt from this review might not even be fully applicable to other future reviews we might conduct.

Unfortunately we do not have the space here to discuss all of the lessons learned from our review, such as tackling the question of quality appraisal, issues of synthesizing findings, and which parts of the review to include in publications, a discussion of which would complement this chapter. Likewise, the experiences throughout our review and our solutions to them certainly also constitute limitations of our work—as will also be discussed in the publications ensuing the review. However, it is our hope that by discussing them so openly and thoroughly within this chapter, other researchers who are conducting a systematic review for the first time, or who experience similar issues, may benefit from our experience.

**Acknowledgements** We gratefully acknowledge the support offered by Dr. Oliver Schoenbeck (Information Specialist at the University of Oldenburg Library) throughout our systematic review project.

## References

Alioon, Y., & Delialioglu, O. (2017). The effect of authentic m-learning activities on student engagement and motivation. *British Journal of Educational Studies*. https://doi.org/10.1111/bjet.12559.

Appleton, J., Christenson, S.L., & Furlong, M. (2008). Student Engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369–386. https://doi.org/10.1002/pits.20303.

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069.

Beverley, C. A., Booth, A., & Bath, P. A. (2003). The role of the information specialist in the systematic review process: a health information case study. *Health Information and Libraries Journal*, *20*(2), 65–74. https://doi.org/10.1046/j.1471-1842.2003.00411.x.

Bedenlier, S., Bond, M., Buntins, K., Zawacki-Richter, O., & Kerres, M. (Forthcoming). Facilitating student engagement through educational technology in higher education: A systematic review in the field of Arts & Humanities. *Australasian Journal of Educational Technology.*

Bond, M., & Bedenlier, S. (2019). Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education*, 2019(1), 12. https://doi.org/10.5334/jime.528

Boland, A., Cherry, M. G., & Dickson, R. (2017). *Doing a systematic review: a student's guide* (2nd edition). Thousand Oaks, CA: SAGE Publications.

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*(2), e012545. https://doi.org/10.1136/bmjopen-2016-012545.

Brunton, G., Stansfield, C., & Thomas, J. (2012). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 107–134). London; Thousand Oaks, CA: SAGE.

Buntins, K., Bedenlier, S., Bond, M., Kerres, M., & Zawacki-Richter, O. (2018). Mediendidaktische Forschung aus Deutschland im Kontext der internationalen Diskussion. Eine Auswertung englischsprachiger Publikationsorgane von 2008 bis 2017. In B. Getto, P. Hintze, & M. Kerres (Eds.), *Digitalisierung und Hochschulentwicklung Proceedings zur 26. Tagung der Gesellschaft für Medien in der Wissenschaft e. V.* (Vol. 74, pp. 246–263). Münster, New York: Waxmann.

Campbell, A., Taylor, B., Bates, J., & O'Connor-Bones, U. (2018). Developing and applying a protocol for a systematic review in the social sciences. *New Review of Academic Librarianship*, *24*(1), 1–22. https://doi.org/10.1080/13614533.2017.1281827.

Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education*, *15*(1), 211. https://doi.org/10.1186/s41239-018-0109-y.

Chen, P-S.D., Lambert, A.D., & Guidry, K.R. (2010). Engaging online learners: The impact of web-based learning technology on college student engagement. *Computers & Education, 54*(4), 1222–1232. https://doi.org/10.1016/j.compedu.2009.11.008.

Cherry, M.G., & Dickson, R. (2014). Defining my review question and identifying inclusion criteria. In A. Boland, M.G. Cherry, & R. Dickson (Eds.), *Doing a Systematic Review: a student's guide*. London: SAGE Publications.

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of Research on Student Engagement*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-2018-7.

Coverdale, J., Roberts, L. W., Beresin, E. V., Louie, A. K., Brenner, A. M., & Balon, R. (2017). Some potential "pitfalls" in the construction of educational systematic reviews. *Academic Psychiatry*, *41*(2), 246–250. https://doi.org/10.1007/s40596-017-0675-7.

Curran, F. C. (2016). The State of Abstracts in Educational Research. *AERA Open*, *2*(3), 233285841665016. https://doi.org/10.1177/2332858416650168.

Eccles, J. (2016). Engagement: Where to next? *Learning and Instruction*, *43*, 71–75. https://doi.org/10.1016/j.learninstruc.2016.02.003.

Finn, J., & Zimmer, K. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 97–131). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-2018-7_5.

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd edition). Los Angeles: SAGE.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005.

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The impact of digital technology on learning: a summary for the education endowment foundation*. Durham University.

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction*, 43, 39–51. https://doi.org/10.1016/j.learninstruc.2016.01.005

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., … Brooks, C. (2018). How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs. *Review of Educational Research*, *88*(1), 43–86. https://doi.org/10.3102/0034654317740335.

Junco, R. (2012). The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers & Education*, *58*(1), 162–171. https://doi.org/10.1016/j.compedu.2011.08.004.

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758–773. https://doi.org/10.1080/03075079.2011.598505.

Kelley, K., Lai, K., Lai, M.K., & Suggests, M. (2018). *Package 'MBESS'*. Retrieved from https://cran.r-project.org/web/packages/MBESS/MBESS.pdf.

Kerres, M. (2013). *Mediendidaktik. Konzeption und Entwicklung mediengestützter Lernangebote*. München: Oldenbourg.

Krause, K.-L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, *33*(5), 493–505. https://doi.org/10.1080/02602930701698892.

Kupper, L.L., & Hafner, K.B. (1989). How appropriate are popular sample size formulas? *The American Statistician, 43*(2), 101–105.

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education, 24*, 26–34. https://doi.org/10.1016/j.iheduc.2014.09.005.

Miake-Lye, I. M., Hempel, S., Shanman, R., & Shekelle, P. G. (2016). What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Reviews*, *5*(1). https://doi.org/10.1186/s13643-016-0204-x.

Mullen, P. D., & Ramírez, G. (2006). The promise and pitfalls of systematic reviews. *Annual Review of Public Health*, *27*(1), 81–102. https://doi.org/10.1146/annurev.publhealth.27.021405.102239.

Nelson Laird, T. F., & Kuh, G. D. (2005). Student experiences with information technology and their relationship to other aspects of student engagement. *Research in Higher Education*, *46*(2), 211–233. https://doi.org/10.1007/s11162-004-1600-y.

Northey, G., Bucic, T., Chylinski, M., & Govind, R. (2015). Increasing student engagement using asynchronous learning. *Journal of Marketing Education*, *37*(3), 171–180. https://doi.org/10.1177/0273475315589814.

O'Mara-Eves, A., Brunton, G., McDaid, D., Kavanagh, J., Oliver, S., & Thomas, J. (2014). Techniques for identifying cross-disciplinary and 'hard-to-detect' evidence for systematic review: Identifying Hard-to-Detect Evidence. *Research Synthesis Methods*, *5*(1), 50–59. https://doi.org/10.1002/jrsm.1094.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, (5), 1–10. https://doi.org/10.1186/s13643-016-0384-4.

Popenici, S. (2013). Towards a new vision for university governance, pedagogies and student engagement. In E. Dunne & D. Owen (Eds.), *The Student Engagement Handbook: Practice in Higher Education*. United Kingdom: Emerald.

Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, *63*, 604–612. https://doi.org/10.1016/j.chb.2016.05.084.

Salaber, J. (2014). Facilitating student engagement and collaboration in a large postgraduate course using wiki-based activities. *The International Journal of Management Education*, *12*(2), 115–126. https://doi.org/10.1016/j.ijme.2014.03.006.

Santos, C. M. C., Pimenta, C. A. M., & Nobre, M. R. C. (2007). The PICO strategy for the research question construction and evidence search. *Rev Latino-am Enfermagem, 15*(3), 508–5011.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Review of Educational Research*, *81*(1), 4–28. https://doi.org/10.3102/0034654310393361.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# Systematic Reviews on Flipped Learning in Various Education Contexts

Chung Kwan Lo

## 1 Introduction

In recent years, numerous studies about the flipped (or inverted) classroom approach have been published (Chen et al. 2017; Karabulut-Ilgu et al. 2018). In a typical flipped classroom, students learn course materials before class by watching instructional videos (Bishop and Verleger 2013; Lo and Hew 2017). Class time is then freed up for more interactive learning activities, such as group discussions (Lo et al. 2017; O'Flaherty and Phillips 2015). In contrast to a traditional lecture-based learning environment, students in flipped classrooms can pause or replay the instructor's presentation in video lectures without feeling embarrassed. These functions enable them to gain a better understanding of course materials before moving on to new topics (Abeysekera and Dawson 2015). Moreover, instructors are no longer occupied by direct lecturing and can thus better reach every student inside the classroom. For example, Bergmann and Sams (2008) provide one-to-one assistance and small group tutoring during their class meetings.

The growth in research on flipped classrooms is reflected in the increasing number of literature review studies. Many of these are systematic reviews (e.g., Betihavas et al. 2016; Chen et al. 2017; Karabulut-Ilgu et al. 2018; Lundin et al. 2018; O'Flaherty and Phillips 2015; Ramnanan and Pound 2017). One would expect that if the scope of review has remained unchanged, contemporary reviews would include and analyze more research articles than the earlier

C. K. Lo (✉)
University of Hong Kong, Pok Fu Lam, Hong Kong
e-mail: cklohku@gmail.com

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_8

**Table 1** Summary of the systematic reviews of flipped classroom research written by the author (in chronological order)

| Study | Focus | Studies reviewed (*n*) | Search period |
|---|---|---|---|
| Lo (2017) | History education | 5 | Up to Jun 2016 |
| Lo and Hew (2017) | K-12 education | 15 | 1994 to Sep 2016 |
| Lo et al. (2017) | Mathematics education | 61 | 2012 to 2016 |
| Hew and Lo (2018) | Health professions | 28 | 2012 to Mar 2017 |
| Lo and Hew (2019) | Engineering education | 31 | 2008 to 2017 |

reviews. Moreover, because flipped classroom practice is becoming more innovative (e.g., gamified flipped classroom), recent reviews should provide new insights into future research and practice. However, this is not always the case.

With this in mind, this chapter highlights possible strategies to improve the quality of systematic reviews. The chapter is based on my experiences of and reflections on systematic reviews of flipped classroom research in various contexts (Table 1). It begins by presenting the rationale for conducting systematic reviews. The chapter then discusses how systematic reviews contribute to the flipped learning field. In contrast to several existing reviews, it then shares my reflections on practical aspects of systematic reviews, including literature search, article selection, and research synthesis. The chapter concludes with a summary.

## 2 Rationale for Conducting Systematic Reviews

To avoid repeating previous research efforts, researchers should first understand the current state of the literature by either examining existing reviews or conducting their own systematic review. Phrases such as "little research has been done" and "there is a lack of research" are extensively used to justify a newly written article. However, I sometimes doubt the grounds for these claims. There is no longer a lack of research in the field of flipped learning. In mathematics education alone, for example, 61 peer-reviewed empirical studies were published between 2012 to 2016 (Lo et al. 2017). Karabulut-Ilgu et al. (2018) found 62 empirical research articles on flipped engineering education as of May 2015. Through a systematic review of the literature, a more comprehensive picture of current research can be revealed.

In fact, before conducting my studies of flipped learning in secondary schools, I carried out a systematic review in the context of K-12 education (Lo and Hew 2017). At the time of writing (October 2016), only 15 empirical studies existed.

We therefore knew little (at that time) about the effect of flipped learning on K-12 students' achievement under this instructional approach. With such a small number of research published, the systematic review thus provided a justification for our planned studies (see Lo et al. 2018 for a review) and those of other researchers (e.g., Tseng et al. 2018) to examine the use of the flipped classroom approach in K-12 contexts.

In addition to understanding the current state of the literature, systematic reviews help identify research gaps. In flipped mathematics education, for example, Naccarato and Karakok (2015) hypothesized that instructors "used videos for the delivery of procedural knowledge and left conceptual ideas for face-to-face interactions" (p. 973). However, researchers have not reached a consensus on course planning using the lens of procedural and conceptual knowledge. While Talbert (2014) found that students were able to acquire both procedural and conceptual knowledge by watching instructional videos, Kennedy et al. (2015) discovered that flipping conceptual content might impair student achievement. More importantly, we found in our systematic review that very few studies evaluated the effect of flipping specific types of materials, such as procedural and conceptual problems (Lo et al. 2017). To flip or not to flip the conceptual knowledge? That is a key question for future studies of flipped mathematics learning.

## 3 Contribution of Systematic Reviews

A systematic review should not be merely a summary of existing studies. Instead, the review should contribute to the body of knowledge. Researchers must figure out the purpose of their systematic review and ensure the significance of their work. This section illustrates several possible goals of research synthesis. Table 2 shows that in our systematic review, we aimed to achieve two main goals: (1) To inform future flipped classroom practice, and (2) to compare the overall effect of flipped learning to traditional lecture-based learning.

First, the overarching goal of some of our systematic reviews was to inform future flipped classroom practice. Using the findings of the reviewed studies, we have developed a 5E flipped classroom model for history education (Lo 2017), made 10 recommendations for flipping K-12 education (Lo and Hew 2017), and established a set of design principles for flipped mathematics classrooms (Lo et al. 2017). Taking the design principles for flipped mathematics classrooms as an example, our Principle 4 suggested that short videos could be used to enable effective multimedia learning. This principle was based on the problem (reported in the literature) that students tend to disengage when watching long videos.

**Table 2** Some possible goals and contributions of systematic reviews of flipped classroom research

| Goal | Major contributions |
|---|---|
| To inform future flipped classroom practice | • Develop a 5E flipped classroom model for history education (Lo 2017).<br>• Make 10 recommendations for flipping K-12 education (Lo and Hew 2017).<br>• Establish a set of design principles for flipped mathematics classrooms (Lo et al. 2017). |
| To examine the overall effect of flipped learning | • Provide quantitative evidence that flipped learning improves student achievement more than traditional learning in the following subject areas.<br>– Mathematics education (Lo et al. 2017)<br>– Health professions (Hew and Lo 2018)<br>– Engineering education (Lo and Hew 2019)<br>• Reveal that the use of the following instructional activities can further promote the effect of flipped learning.<br>– A formative assessment of pre-class materials (Hew and Lo 2018; Lo et al. 2017)<br>– A brief review of pre-class materials (Lo and Hew 2019) |

To avoid making similar mistakes, we recommended that each video be limited to six minutes and all combined video segments be no more than 20–25 min. With this principle applied, Chen and Chen (2018) confirmed that the assigned workload was bearable for the students in their flipped research methodology course.

Second, the goal of our systematic reviews was to examine the effect of flipped learning versus traditional learning on student achievement. These reviews focus on flipped mathematics education (Lo et al. 2017), health professions (Hew and Lo 2018), and engineering education (Lo and Hew 2019). Researchers have conducted several systematic reviews of flipped learning in the health professions (Chen et al. 2017; Ramnanan and Pound 2017) and engineering education (Karabulut-Ilgu et al. 2018). Ramnanan and Pound (2017) reported that medical students were generally satisfied with flipped learning and preferred this instruction approach to traditional lecture-based learning. However, strong satisfaction with learning does not necessarily mean improved achievement. Examining student learning outcomes, Karabulut-Ilgu et al. (2018) classified their flipped-traditional

comparison studies into five categories: (1) More effective, (2) more effective and/or no difference, (3) no difference, (4) less effective, and (5) less effective and/or no difference. As in Chen et al. (2017), they presented the effect size of each flipped-traditional comparison study. However, as Karabulut-Ilgu et al. (2018) acknowledged, no definitive conclusion can be made without a meta-analysis of student achievement in flipped classrooms.

We therefore attempted to examine the overall effect of flipped learning on student achievement through systematic reviews of the empirical research. The findings enhance our understanding of this instructional approach. Using a meta-analytic approach, a small but significant difference in effect in favor of flipped learning over traditional learning was found in all three contexts (i.e., mathematics education, health professions, and engineering education). Most importantly, our moderator analyses provided quantitative support for a brief review and/or formative assessment of pre-class materials at the start of face-to-face lessons. The effect of flipped learning was further promoted when instructors provided such an assessment (for mathematics education and health professions) and/or review (for engineering education) in their flipped classrooms. These findings not only extend our understanding of flipped learning, but also inform future practice of flipped classrooms (e.g., offering a quiz on pre-class materials at the start of face-to-face lessons).

## 4 Reflections on Some Practical Issues of Conducting Systematic Reviews

The following sections cover some practical aspects of systematic reviews of flipped classroom research, including literature search, article selection, and research synthesis.

### 4.1 Literature Search

Abeysekera and Dawson (2015) shared their experiences of searching for articles on flipped classrooms. They performed their search using the term "flipped classroom" in the ERIC database. In June 2013, they found only two peer-reviewed articles on flipped learning. Although not much research had been published at that time, this scarcity of search outcome has prompted us to reflect on (1) the design of the search string and (2) the choice of databases when conducting a systematic review.

## 4.2 The Design of Search String

The search term "flipped classroom" is very specific in that it cannot include other terms used to describe this instructional approach, such as flipped learning, flipping classrooms, and inverted classrooms. From my observation, some authors use even more flexible wording. For example, Talbert (2014) entitled his article "Inverting the Linear Algebra Classroom" (p. 361). If certain keywords are not included in their title, abstract, and keywords, their articles might not be retrieved through a narrow database search.

Although it is the authors' responsibility to use well-recognized keywords, researchers producing systematic reviews should make every effort to retrieve as many relevant studies as possible. To this end, we used the asterisk as a wild card to capture different verb forms of "flip" (i.e., flip, flipping, and flipped) and "invert" (i.e., invert, inverting, and inverted). The asterisk also allowed the inclusion of both singular and plural forms of nouns (e.g., class and classes, classroom and classrooms). Furthermore, Boolean operators (i.e., AND and OR) were applied to separate each search term to increase the flexibility of our search strings. In this way, we were able to include some complicated expressions used in flipped classroom research, such as "Flipping the Statistics Classroom" (Kuiper et al. 2015, p. 655). Table 3 shows the search strings that we used in the systematic reviews of flipped history education (Lo 2017), K-12 education (Lo and Hew 2017), and mathematics education (Lo et al. 2017).

Our search strings comprised two parts: (1) The instructional approach, and (2) the context. In the first part, "(flip* OR invert*) AND (class* OR learn*)" allowed us to capture different combinations of terms about flipped learning. In the second part, we used various search terms to specify the research contexts (e.g., K12 OR K-12 OR primary OR elementary OR secondary OR "high school" OR "middle school") or subject areas (e.g., math* OR algebra OR trigonometry OR geometry OR calculus OR statistics) that we wanted. As a result, we were able to reach research items that had seldom been downloaded and cited.

However, upon completion of the systematic reviews in Table 3, we realized that researchers might use other terms to describe the flipped classroom approach, such as "flipped instruction" (He et al. 2016, p. 61). Therefore, we further included "instruction*" and "course*" in our search strings. Table 4 shows the improved search strings that we used in the systematics reviews of flipped health professions (Hew and Lo 2018) and engineering education (Lo and Hew 2019).

As a side note about the design of search strings, one researcher emailed me about our systematic review of flipped mathematics education (Lo et al. 2017). He told me that our review has missed his article, an experimental study of flipped

**Table 3** Search strings used in systematic reviews of flipped classroom research

| Study | Search string | | | | |
|---|---|---|---|---|---|
| | Instructional approach | | | | Context |
| History education (Lo, 2017) | (flip* OR invert*) | AND | (class* OR learn*) | AND | History |
| K-12 education (Lo and Hew 2017) | (flip* OR invert*) | AND | (class* OR learn*) | AND | (K12 OR K-12 OR primary OR elementary OR secondary OR "high school" OR "middle school") |
| Mathematics education (Lo et al. 2017) | (flip* OR invert*) | AND | (class* OR learn*) | AND | (math* OR algebra OR trigonometry OR geometry OR calculus OR statistics) |

**Table 4** Improved search strings used in systematic reviews of flipped classroom research

| Study | Search string | | | | |
|---|---|---|---|---|---|
| | Instructional approach | | | | Context |
| Health professions (Hew and Lo 2018) | (flip* OR invert*) | AND | (class* OR learn* OR instruction* OR course*) | AND | (medic* OR nurs* OR pharmac* OR physiotherap* OR dental OR dentist* OR chiropract*) |
| Engineering education (Lo and Hew 2019) | (flip* OR invert*) | AND | (class* OR learn* OR course*) | AND | engineering |

mathematics learning. After careful checking, his study perfectly fulfilled all inclusive criteria for our systematic review. However, I could not find any variations of "mathematics" or other possible identifiers of subject areas (e.g., algebra, calculus, and statistics) in his title, abstract, and keywords. That is why we were unable to retrieve his article through database searching using our search string.

At this point, I still believe that the context part of our search string of flipped mathematics education (i.e., math* OR algebra OR trigonometry OR geometry OR calculus OR statistics) is broad enough to capture the flipped classroom research conducted in mathematics education. However, this search string cannot capture studies that do not describe their subject domain at all. Without this information, other readers would have no idea about where the work is situated within the broader field of flipped learning if they only scan the title, abstract, and keywords. Most importantly, this valuable piece of work cannot be retrieved in a database search. Other snowballing strategies, such as tracking the reference lists of reviewed studies (see Lo 2017; Wohlin 2014 for a review), should be applied to find these articles in future systematic reviews.

## 4.3 The Choice of Databases

In our systematic reviews, we performed our literature search across databases, such as Academic Search Complete, TOC Premier, and ERIC. For the systematic review of flipped health professions (Hew and Lo 2018), we further used databases of medicine education, including PubMed, PsycINFO, CINAHL Plus, and British Nursing Index. In my experience, there are relatively few documents about flipped learning in the ERIC database. For example, Fig. 1 shows that we obtained 1611 peer-reviewed journal articles (though not all articles were related to flipped learning) in Academic Search Complete using our search string of health professions, but only 14 in ERIC. This situation was similar to the systematic review of flipped engineering education by Karabulut-Ilgu et al. (2018), in which we only found two documents in ERIC. Therefore, flipped classroom research reviewers should not restrict their searches to this database.

Records identified through database searching (n = 2125)

- Academic Search Complete (n = 1611)
- PubMed (n = 759)
- PsycINFO (n = 126)
- CINAHL Plus (n = 94)
- TOC Premier (n = 49)
- British Nursing Index (n = 31)
- ERIC (n = 14)

**Fig. 1** The search outcome of flipped classroom research across databases in health professions (Hew and Lo 2018, p. 4)

Apart from the aforementioned databases, other researchers (e.g., Lundin et al. 2018; O'Flaherty and Phillips 2015; Ramnanan and Pound 2017) have used the following databases in their systematic reviews of flipped learning: Cochrane library, EMBASE, Joanna Briggs Institute, Scopus, and Web of Science. In future systematic reviews, relevant databases need to be consulted. Researchers can follow existing reviews in their research field or consult librarians for advice on which databases to use.

## 4.4 Article Selection

After obtaining the search outcomes, we selected articles based on our inclusion and exclusion criteria. Other existing systematic reviews also develop criteria for article selection. However, they have a few constraints (Table 5) that reviewers may disagree and could significantly limit the number of studies included. As a result, the representativeness and generalizability of the reviews could be impaired. Researchers should thus provide strong rationales for their inclusion and exclusion criteria for article selection.

Taking a recent systematic review by Lundin et al. (2018) as an example, they reviewed the most-cited publications on flipped learning. They only included publications that were cited at least 15 times in the Scopus database. With such a constraint, 493 out of 530 documents were excluded in the early stage of their

**Table 5** A few controversial criteria for article selection

| Criteria | Inclusion | Exclusion | Possible concerns |
|---|---|---|---|
| Citation (Lundin et al. 2018) | Studies are cited at least 15 times in Scopus | Studies are cited less than 15 times in Scopus | Why "15 times" is set, instead of 10 or other possibilities? Will this criterion become a threshold for recently published studies? |
| Ethics clearance (O'Flaherty and Phillips 2015) | Studies with approved ethics notification | Studies without approved ethics notification | Is it common practice in the field of flipped learning to explicitly acknowledge ethical approval in journal writing? |

review. Only 31 articles were ultimately included in their synthesis. This particular criterion could block the inclusion of recently published articles because it takes time to accumulate a number of citations. The majority of the articles that they included were published in 2012 ($n=6$), 2013 ($n=16$), and 2014 ($n=5$), with only a scattering of articles from 2000 ($n=1$), 2008 ($n=1$), and 2015 ($n=2$). No documents after 2016 were included in their systematic review. The authors argued that citation frequency is "an indicator of which texts are widely used in this emerging field of research" (p. 4). However, further justification may help highlight the value of examining this particular set of documents instead of a more comprehensive one. They also have to provide a strong rationale for their 15+ citation threshold (as opposed to 10+ or other possibilities).

In our systematic reviews, we also added a controversial criterion for article selection, the definition of the flipped classroom approach. In my own conceptualization, "Inverting the classroom means that events that have traditionally taken place inside the classroom now take place outside the classroom and vice versa" (Lage et al. 2000, p. 32). What traditionally takes place inside the classroom is instructor lecturing. Therefore, I agree with the definition of Bishop and Verleger (2013) that instructional videos (or other forms of multimedia materials) must be provided for students' class preparation. For me, the use of pre-class videos is a necessary element for flipped learning, although it is not the whole story. Merely asking students to read text-based materials on their own before class is not a method of flipping. As one student of Wu et al. (2017) said, "Sometimes I couldn't get the meanings by reading alone. But the instructional videos helped me understand the overall meaning" (p. 150). Using instructional videos, instructors of flipped classrooms still deliver lectures and explain concepts for their students (Bishop and Verleger 2013). Most importantly, this instructional medium can "closely mimic what students in a traditional setting would experience" (Love et al. 2015, p. 749).

However, a number of researchers have challenged the definition provided by Bishop and Verleger (2013). For example, He et al. (2016) asserted that "qualifying instructional medium is unnecessary and unjustified" (p. 61). During the peer-review process, reviewers have also questioned our systematic reviews and disagreed with the use of this definition. In response to the reviewers' concern, we added a section discussing our rationale for using the definition by Bishop and Verleger (2013). We also acknowledged that our systematic review "focused specifically on a set of flipped classroom studies in which pre-class instructional videos were provided prior to face-to-face class meetings" (Lo et al. 2017, p. 50). Without a doubt, if instructors insist on "flipping" their courses using pre-class

text-based materials only, they will not find our review very useful. Therefore, in addition to explaining the criteria for article selection, future systematic reviews should detail their review scope and acknowledge the limitations of reviewing only a particular set of articles.

## 4.5 Research Synthesis

The difficulty of the research synthesis is somewhat correlated to the number of studies to be analyzed. My research synthesis of flipped history education (Lo 2017) was not difficult. In this systematic review, I found only five empirical studies at the time of writing (June 2016). I first extracted the data on learning activities, learning outcomes, benefits, and challenges reported in the reviewed studies. These data were then organized and presented in a logical sequence (e.g., from pre-class to in-class). Similarly, Betihavas et al. (2016) also reviewed and identified themes from only five empirical studies of flipped nursing education. They focused on study characteristics, academic performance outcomes, student satisfaction, and challenges in implementing flipped classrooms. With a limited number of studies, Betihavas et al. (2016) were able to discuss the findings of each reviewed study in detail.

In contrast, synthesizing the findings of a large number of studies is challenging and time-consuming. In our systematic review of flipped mathematics education (Lo et al. 2017), we included and analyzed 61 empirical studies. We read through all of the texts, focusing particularly on the results/findings and discussion sections. One of our research objectives was to understand how the flipped classroom approach benefits student learning, and the challenges of flipping mathematics courses. Codes were assigned to pieces of data (i.e., the benefits and challenges reported in the reviewed studies). Thanks to previous efforts in flipped classroom research, we were able to adopt the frameworks by Kuiper et al. (2015) and Betihavas et al. (2016) as our initial analytic frameworks for benefits and challenges, respectively. Despite the large amount of data to be analyzed, these established frameworks made our research synthesis easier.

Taking the challenges of implementing flipped classrooms as an example, Betihavas et al. (2016) defined three kinds of challenges in their systematic review of flipped nursing education, namely (1) student-related challenges, (2) faculty challenges, and (3) operational challenges. This framework basically covered every aspect involved in implementing a flipped classroom. We therefore adopted this framework as our initial analytic framework for flipped mathematics education

**Table 6** Thematic analysis of the challenges of flipped mathematics education. (Lo et al. 2017, p. 61)

| Theme | Sub-themes (Count) |
|---|---|
| Student-related challenges | • Unfamiliarity with flipped learning ($n=26$)<br>• Unpreparedness for pre-class learning tasks ($n=14$)<br>• Unable to ask questions during out-of-class learning ($n=13$)<br>• Unable to understand video content ($n=11$)<br>• Increased workload ($n=9$)<br>• Disengaged from watching videos ($n=3$) |
| Faculty challenges | • Significant start-up effort ($n=21$)<br>• Not accustomed to flipping ($n=10$)<br>• Ineffectiveness of using others' videos ($n=4$) |
| Operational challenges | • Instructors' lacking IT skills ($n=3$)<br>• Students' lacking IT resources ($n=3$) |

(Lo et al. 2017). With these three kinds of challenges defined as the major themes, all of the identified challenges were then organized into sub-themes (Table 6).

Furthermore, we quantified our thematic analysis by counting the number of studies that contributed to a theme. In this way, our findings could be more specific. Most importantly, such an analysis provided a foundation to develop our design principles to address these challenges. For example, the most-reported student-related challenge was students' unfamiliarity with flipped learning. Therefore, our Principle 1 was to manage their transition to the flipped classroom. We recommended that instructors introduce students to (1) the rationale for flipped learning, (2) the potential benefits and challenges of this instructional approach, (3) the logistics of their flipped course, and (4) the tasks that students need to do (Lo et al. 2017).

## 5 Summary

This chapter shared some experiences of conducting systematic reviews of flipped classroom research. Table 7 recaps the recommendations for future systematic reviews. First, researchers can understand the current state of the literature and identify research gaps by conducting systematic reviews. Systematic reviews can inform future practice or examine the overall effect of instructional strategies.

This chapter discussed several practical aspects of systematic reviews such as literature search, article selection, and research synthesis. To identify relevant documents, researchers should design more flexible search strings using

**Table 7** Recommendations for future systematic reviews

| Aspect | Implications |
|---|---|
| Rationale for conducting systematic reviews | • Understand the current state of the literature<br>• Identify research gaps |
| Contribution of systematic reviews | • Inform future practice, for instance by establishing design principles or instructional models<br>• Examine the overall effect of instructional strategies |
| Literature search | • Use the asterisk and Boolean operators to increase the flexibility of search strings<br>• Perform literature search across relevant databases, instead of relying on only a few databases |
| Article selection | • Provide strong rationale for inclusion and exclusion criteria<br>• Acknowledge the limitations resulted from confining the scope of review |
| Research synthesis | • Adopt established frameworks as initial analytic frameworks, especially when a large number of studies are being reviewed<br>• Quantify the thematic analysis by counting the number of studies that contribute to a theme |

the asterisk and Boolean operators. Moreover, relevant databases should be consulted in the literature search. Researchers should also provide strong rationales for inclusion and exclusion criteria for article selection. Meanwhile, they should acknowledge any possible limitations of their review scope. For the research synthesis, researchers can adopt established frameworks as initial analytic frameworks. Finally, the thematic analysis can be quantified by counting the number of studies that contribute to a theme. Taking these recommendations into account, the quality of future systematic reviews can be improved.

## References

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development, 34*(1), 1–14.

Bergmann, J., & Sams, A. (2008). Remixing chemistry class. *Learning and Leading with Technology, 36*(4), 24–27.

Betihavas, V., Bridgman, H., Kornhaber, R., & Cross, M. (2016). The evidence for 'flipping out': A systematic review of the flipped classroom in nursing education. *Nurse Education Today, 38*, 15–21.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. In *120th ASEE national conference and exposition, Atlanta, GA (paper ID 6219)*. Washington, DC: American Society for Engineering Education.

Chen, F., Lui, A. M., & Martinelli, S. M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education, 51*(6), 585–597.

Chen, T. L., & Chen, L. (2018). Utilizing wikis and a LINE messaging app in flipped classrooms. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(3), 1063–1074.

He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction, 45*, 61–71.

Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Medical Education, 18*(1), article 38.

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411.

Kennedy, E., Beaudrie, B., Ernst, D. C., & St. Laurent, R. (2015). Inverted pedagogy in second semester calculus. *PRIMUS, 25*(9–10), 892–906.

Kuiper, S. R., Carver, R. H., Posner, M. A., & Everson, M. G. (2015). Four perspectives on flipping the statistics classroom: Changing pedagogy to enhance student-centered learning. *PRIMUS, 25*(8), 655–682.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education, 31*(1), 30–43.

Lo, C. K. (2017). Toward a flipped classroom instructional model for history education: A call for research. *International Journal of Culture and History, 3*(1), 36–43.

Lo, C. K., & Hew, K. F. (2017). A critical review of flipped classroom challenges in K-12 education: possible solutions and recommendations for future research. *Research and Practice in Technology Enhanced Learning, 12*(1), article 4.

Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education.* Advance online publication. https://doi.org/10.1002/jee.20293.

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review, 22*, 50–73.

Lo, C. K., Lie, C. W., & Hew, K. F. (2018). Applying "First Principles of Instruction" as a design theory of the flipped classroom: Findings from a collective study of four secondary school subjects. *Computers & Education, 118*, 150–165.

Love, B., Hodge, A., Corritore, C., & Ernst, D. C. (2015). Inquiry-based learning and the flipped classroom model. *PRIMUS, 25*(8), 745–762.

Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., & Peterson, L. (2018). Higher education dominance and siloed knowledge: A systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education, 15*, article 20.

Naccarato, E., & Karakok, G. (2015). Expectations and implementations of the flipped classroom model in undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology, 46*(7), 968–978.

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education, 25*, 85–95.

Ramnanan, C. J., & Pound, L. D. (2017). Advances in medical education and practice: Student perceptions of the flipped classroom. *Advances in Medical Education and Practice, 8*, 63–73.

Talbert, R. (2014). Inverting the linear algebra classroom. *PRIMUS, 24*(5), 361–374.

Tseng, M. F., Lin, C. H., & Chen, H. (2018). An immersive flipped classroom for learning Mandarin Chinese: Design, implementation, and outcomes. *Computer Assisted Language Learning, 31*(7), 714–733.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *18th international conference on evaluation and assessment in software engineering, London, England, United Kingdom (article no.: 38).* New York, NY: ACM.

Wu, W. C. V., Chen Hsieh, J. S., & Yang J. C. (2017). Creating an online learning community in a flipped classroom to enhance EFL learners' oral proficiency. *Educational Technology & Society, 20*(2), 142–157.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

# The Role of Social Goals in Academic Success: Recounting the Process of Conducting a Systematic Review

Naska Goagoses and Ute Koglin

Motivational theorists have long subscribed to the idea that human behavior is fundamentally driven by needs and goals. A goal perspective provides us with insights on the organization of affect, cognition, and behavior in specific contexts, and how these may change depending on different goals (Dweck 1992). The interest in goals is also prominent in the educational realm, which led to a boom of research with a focus on achievement goals (Kiefer and Ryan 2008; Mansfield 2012). Although this research provided significant insights into the role of motivation, it does not provide a holistic view of the goals pursued in academic contexts. Students pursue multiple goals in the classroom (Lemos 1996; Mansfield 2009, 2010, 2012; Solmon 2006), all of which need to be considered to understand students' motivations and behaviors. Many prominent researchers argue that social goals should be regarded with the same importance as achievement goals (e.g., Covington 2000; Dowson and McInerney 2001; Urdan and Maehr 1995), as they too have implications for academic adjustment and success. For instance,

N. Goagoses (✉) · U. Koglin
Department of Special Needs Education & Rehabilitation,
Carl Von Ossietzky University of Oldenburg, Oldenburg, Germany
e-mail: naska.goagoses@uni-oldenburg.de
U. Koglin
e-mail: ute.koglin@uni-oldenburg.de

© The Author(s) 2020

O. Zawacki-Richter et al. (eds.), *Systematic Reviews in Educational Research*,
https://doi.org/10.1007/978-3-658-27602-7_9

studies have shown that social goals are related to academic achievement (Anderman and Anderman 1999), school engagement (Kiefer and Ryan 2008; Shim et al. 2013), academic help-seeking (Roussel et al. 2011; Ryan and Shin 2011), and learning strategies (King and Ganotice 2014). At this point it should be noted that the term *social goal* is a rather broad term, under which many types of social goals fall (e.g., prosocial, popularity, status, social development goals). Urdan and Maehr (1995) stated that there is a critical need for research to untangle and investigate the various social goals, as these could have different consequences for students' motivation and behavior.

Intrigued by social goals and their role in socio-academic contexts, we opted to pursue this line of research for a larger project. At the beginning of every research endeavor, familiarization with the relevant theories and current research is essential. It is an important step before conducting primary research as unnecessary research is avoided, current knowledge gaps are exposed, and it can help with the interpretation of later findings. Furthermore, funding bodies that provide research grants often require a literature review to assess the significance of the proposed project (Siddaway et al. 2018). Customary within our research group, this process is completed by conducting a thorough systematic review. In addition to being beneficial merely to the authors, systematic reviews also provide other researchers and practitioners with a clear summary of findings and critical reflections thereof. Considering that the research on social goals and academic success dates back nearly 30 years, we deemed this to be an ideal time to provide a systematic overview of the entire research.

## 1 Purpose of Review

Our main aim was to produce a comprehensive review, which adequately displays the significance of social goals for academic success. Gough et al. (2012) describe the different types of systematic reviews by exploring their aims and approaches, which include the role of theory, aggregative versus configurative reviews, and further ideological and theoretical assumptions. The purpose of the current review was to further theoretical understanding of the current phenomena by developing concepts and arranging information in a configurative way. As such we were interested in exploratively investigating the role of social goals on academic success, by identifying patterns in a heterogeneous range of empirical findings. With such reviews, the review question is rather open, concepts are emergent, procedures less formal, theoretical inferences are drawn, and insightful information is synthesized (Brunton et al. 2017). Levinsson and Prøitz (2017)

found that configurative reviews are rarely used in education, although they can be very beneficial for academic researchers, especially at the start of new research projects. Commonly researchers gather information from introductory sections of empirical journal articles, without considering that this information is cherry-picked to support the rationale and hypotheses of a research study. In order to thoroughly inform research and practice, a configurative and systematic summary of empirical findings are needed.

## 2 Methods

Although the aim of our systematic review was to learn something new about the relation between social goals and academic success, we did not tread into the process blindly. A paramount yet often overlooked step in the systematic review process is the exploration of relevant theoretical frameworks. Even though the theoretical framework does not need to be explicitly stated in the systematic review, it is of essential importance as it lays the foundation for every step of the process (Grant and Osanloo 2014). We thus first spent time understanding the theoretical backgrounds and approaches with which prominent research articles explored social goals and investigated their relation to academic success. This initial step helped us throughout review process, as it allowed us to better understanding the research questions and results presented in the articles, revealed interconnections with bordering topics, and gave us a more structured thought process. Naturally during the course of the systematic review, we underwent a learning process in which we gathered new theoretical knowledge and also updated previously held notions.

Before starting with the systematic review, we checked whether there are already existing reviews on the topic. We initially checked the Cochrane Database of Systematic Reviews and PROSPERO, which revealed no registered systematic reviews on the current topic. Being skeptical that these databases include non-medical reviews or that social scientists would register their reviews on these databases, we opted to search for existing systematic reviews on social goals through the Web of Science Core Collection. We identified two narrative reviews, which specifically related to social goals in the academic context (Dawes 2017; Urdan and Maehr 1995), and one narrative review on social goals (Erdley and Asher 1999). We decided that a current systematic review was warranted to provide an updated and more holistic view of the literature on social goals in relation to academic adjustment and success. We drafted a protocol, which included background and aims of the review, as well as selection criteria, search

strategy, screening and data extraction methods, and plan for data synthesis. As our research agenda follows a configurative approach, we adapted the protocol iteratively when certain methods and procedures were found to be incompatible (see Gough et al. 2012); these changes are reflected transparently throughout the review. We did not register the systematic review on any database; upon request to the corresponding author the protocol can be acquired.

## 3 Literature Search

Systematic review searches should be objective, rigorous, and inclusive, yet also achieve a balance between comprehensiveness and relevance (Booth 2011; Owen et al. 2016). Selecting the "right" keywords that find this balance is not always easy and may require more thought than simply using the terms of the research question. A particular problem within psychology and educational research may lie within the used constructs, as the (in)consistency in terminology, definition, and content of constructs is a plight known to many researchers. The *déjà-variable* (Hagger 2014) and the *jangle fallacy* (Block 1995) are phenomena in which similar constructs are referred to by different names; this presents a particular challenge for systematic reviews, as entire literatures may be neglected if only a surface approach is taken to identify construct terminologies (Hagger 2014). Researchers may also be lured into relying on hierarchical or umbrella terms, in which a range of common concepts are covered with a single word. This is problematic, as literature which uses specific and detailed terminology instead of umbrella terms will be overlooked.

We were faced with such dilemmas when we decided to embark on a systematic review which investigates *social goals*; relying on the one term (and its synonyms) was deemed insufficient to comprehensively extract all appropriate articles. We thus referred back to the three identified reviews on the topic and systematically extracted all types of social goals that were mention in these articles. As the dates of the reviews range from 1995 to 2017, we assumed that they would encompass a range of approaches and specific social goals. We acknowledge that this is by no means extensive and other conceptualizations of goals exist (e.g., Chulef et al. 2001; Ford 1992; McCollum 2005). Nonetheless our search can be deemed both systematic and comprehensive, and resulted in 42 keywords for the term *social goals*. This large number of keywords might seem unusual, but specifically address our quest to investigate the role of various social goals for student's academic success (as requested by Urdan and Maehr 1995).

For the second part of the search string, we used general keywords contextual to the field of academia to reflect the differential definitions and operationalizations that exist of academic success (e.g., achievement, effort, engagement). Keeping the keywords for academic success broad, meant our systematic review would take on a rather open nature. This delineates from most other reviews in which the outcome is more narrowly set. In retrospect, we found this to be quite effortful as we had to keep updating our own conceptualization of academic success and apply these to further decision-making processes. Nonetheless, we maintain that this allowed us to develop a well-rounded systematic review, in which our pre-existing knowledge did not bias our exploration of the topic. In Appendix A are our final keywords, embedded in a Boolean search string as they were used. In addition to combining or terms with the OR and AND operators, we added an asterisk (*) to the term goal to include single and plural forms.

To locate relevant articles, in March 2018 we entered our search string in the following electronic bibliographic databases: Web of Science Core Collection, Scopus, and PsycINFO. These were entered as free-text terms and thus applied to title, abstract, and keywords (depending on database). It is advisable to use multiple databases, as variations in content, journals, and period covered exists even in renowned scientific electronic databases (Falagas et al. 2008). In January 2019 we conducted an update as our initial search was more than six months ago; we entered the same keywords into Web of Science Core Collection and Scopus.

## 4 Selection Criteria

To be included in this review, articles were required to

- *Be relevant for the topic under investigation.* Commonly excluded for being topic irrelevant were articles, which used the social goal keywords in a different way (e.g., global education goals), or only focused on social goals without explicating any academic relevance (e.g., social goals of bullies), or focused on non-social achievement goals (e.g., mastery goals).
- Be *published in a peer-reviewed journal*. Dissertations, conference papers, editorials, books, and book chapters were excluded. If after extensive research it was unclear whether or not a journal was peer-reviewed it was excluded.
- *Report empirical studies*. Review articles and theoretical papers were excluded.

- To constitute of a *population of students*, articles which focused on teachers or parents were excluded. No age restrictions were set.
- *Not examine special populations* (e.g., children with disabilities, ADHD).
- *Provide full-texts in English*. We acknowledge that this introduces a language bias, but due to limited resources we were unable to translate non-English articles. Although both authors are bilingual, we felt the inclusion of specifically German articles would arguably introduce more bias.

We opted not to impose a publication date restriction; thus, the search covered articles from the first available date until March 2018.

## 5 Study Selection

Appendix B provides a flow diagram of the study selection process, which has been adapted from Moher et al. (2009). All potential articles obtained via the electronic database searches were imported into EPPI-Reviewer 4, and duplicate articles were removed. A title and abstract screening ensued, which resulted in the exclusion of all articles that did not meet the selection criteria. If these did not provide sufficient information, the article was shifted into the next phase. For articles that were excluded on the bases that they were not empirical, a backward reference list checking was conducted. Specifically, the titles of articles in the listed references were screened and resulted in the addition of a few new articles. Reference list checking is acknowledged as a worthwhile component of a balanced search strategy in numerous systematic review guidelines (Atkinson et al. 2015).

We were able to locate all but three articles via university libraries and online searches (e.g., ResearchGate). Full-text versions of the preliminarily included articles were obtained and screened for eligibility based on the same selection criteria. Wishing to explore research gaps in the area, as well as having an interest in the developmental changes of social goals, we originally intended to keep the level of education very broad (primary, secondary, and tertiary). During the full-text screening we were however reminded of the dissimilarities between these academic contexts, as well as school and university students; we also realized that the addressed research questions in tertiary education varied from the rest (e.g., cheating behavior, cross-cultural adjustment). Thus, articles dealing with tertiary education students were also eliminated at this point.

Although we initially planned to include both quantitative and qualitative articles, we came to realize during the full-text screening that this may be more problematic than first anticipated. Qualitative articles are often excluded from

systematic reviews, although their use can increase the worth and understanding of synthesized results (CRD 2009; Dixon-Woods et al. 2006; Sheldon 2005). While strides have been made in guiding the systematic review process of qualitative research, epistemological and methodological challenges remain prominent (CRD 2009; Dixon-Woods et al. 2006). Reviewing quantitative research in conjunction with qualitative data is even more challenging, as qualitative and quantitative research varies in epistemological, theoretical, and methodological underpinnings (Yilmaz 2013). With an increased interest in mixed-methods research (Johnson and Onwuegbuzie 2004; Morgan 2007), the development of appropriate systematic review methodologies needs to be boosted. Due to the differential methodologies described for systematic reviews of qualitative and quantitative articles and a lack of clear guidance concerning their convergence, we opted to exclude all qualitative and mixed-method articles at this stage. To not lose vital information provided by these qualitative articles, we incorporated some of their findings into other sections of the review (e.g., introduction).

During the full-text screening we came to realize that we had a rather idealistic plan of conducting a comprehensive yet broad systematic review. To not compromise on the depth of the review, we opted to narrow the breadth of the review. Nonetheless, having a broader initial review question subsequently followed by a narrower one, allows us to create a synthesis in which studies can be understood within a wider context of research topics and methods (see Gough et al. 2012). Furthermore, we maintain that the inclusion of multiple social goals as well as different academic success and adjustment variables still provides a relatively broad information bank, from which theories can be explored and developed. Our review thus followed an iterative yet systematic process.

## 6 Data Extraction

We created an initial codebook, which included numerous categories of information to be extracted from each article. Piloting the codebook ensures that all relevant data is captured and that resources are not wasted on extracting unrequired information (CRD 2009). After piloting the codebook on some of the included articles we realized adaptations needed to be made. We carefully deliberated on which information needs to be extracted to accurately map the articles, indicate research gaps, and provide relevant information for a well-rounded synthesis on the current topic. Information regarding the identification of articles was already incorporated in EPPI-Reviewer when articles were first identified. We included both open and categorical coding schemes to extract theoretical information

(i.e., overarching aim, social goal type and approach, research questions, hypotheses), participant details (i.e., number, age range, education level, continent of study), methodological aspects (i.e., design, time periods, variables, social goal measurement tools), and findings (i.e., main results, short conclusions).

Extracting theoretical information from the articles, such as the overarching aim and research questions was fairly simple. Finding the suggested hypotheses was a bit more complicated, as many articles did not explicitly report these in one section. Surprisingly, almost a third of the articles did not mention a priori hypotheses. Uniformly extracting the description of the participants also required some maneuvering. We found that the seemingly simple step of extracting the number of participants required some tact, as articles differentially reported these numbers (e.g., before-after exclusion, attrition, multiple studies). Studies differentially described the age of participants, with some reporting only the mean, others the age range, and some not mentioning the age at all (i.e., reporting only the grade level). We opted not to extract additional descriptive participant data, such as socio-economic status and sex-ratio, as these were not central to the posed research question and results of the included articles. Attributing study design was easily completed with a closed categorical coding scheme, whilst listing all the included variables required an open coding scheme.

Extracting which measurements (i.e., scales and questionnaires) were used to assess social goals with their respective references was constructive for our review. Engaging with the operationalizations provided us with a deeper understanding of the concepts, lead to new insights about the various forms of conceptualizations, and also revealed stark inconsistencies albeit sharing the same term. To extract the main results, we combed through the results section of the article, whilst at the same time having the research question(s) at hand. With this strategy we did not extract information about the descriptive or preliminary analyses, but specifically focused on the important analyses pertaining only to social goals. Although we did not extract any information from the discussion, we did find it useful to read this section as it provided us with confirmation that we extracted the main results correctly and allowed us to place them in a bigger theoretical context. As a demonstrative example, Table 1 shows a summary of some of the extracted data from five of the included articles.

**Table 1** Summary of extracted data from five reviewed articles

| Author | Sample | Grade | Social Goals | Other constructs | Method |
|---|---|---|---|---|---|
| Anderman and Anderman (1999) | 254 North American middle school children | Grades 3–6 | Prosocial | Gender; academic success disclosures; perceptions of friend's responses and disclosures as normative; achievement (grades in school subjects) | Cross-sectional |
| Kiefer and Ryan (2008) | 718 (time 1) and 656 (time 2) North American middle school children | Grades 6–7 | Intimacy | Self-reported disruptive behavior and peer-reported disobedience; self and peer-reported effort; achievement (GPA); gender; ethnicity | Longitudinal |
| King et al. (2012) | 1147 Asian secondary school children | Not specified | Affiliation, approval, concern, status, responsibility | Achievement goals, engagement (behavioral, emotional, & cognitive) | Cross-sectional |
| Shim et al. (2013) | 478 North American middle school children (Time 1) | Grades 6–8 | Social development, demonstration-approach, demonstration-avoid | Classroom goal structures, academic engagement, social adjustment | Longitudinal |
| Wentzel (1993) | 423 North American middle school children | Grades 6–7 | Responsibility | Mastery and evaluation goals; classroom goals; achievement (GPA and test scores); causal belief systems | Longitudinal |

## 7 Synthesis

In our protocol we stated that we would conduct a narrative synthesis, as this would be most appropriate to the array of (quantitative) studies we hoped to include in the systematic review. A narrative synthesis is a textual approach to the systematic review, which involves summarizing and explaining the findings of multiple studies with primarily words and text (Popay et al. 2006). We have since come to realize that the term 'narrative synthesis' is rather generic, describing a collection of methods for synthesizing data narratively (Snilstveit et al. 2012). Upon inspection of the range of available methods (see Barnett-Page and Thomas 2009; Dixon-Woods et al. 2005), we decided that we would use a thematic analysis (synthesis), as it is a good method when dealing with a broad range of findings. A thematic analysis involves creating summaries of prominent and recurrent themes in the articles in a systematic way. We aimed to create an intertwined web of results from all the studies.

The synthesis is probably the most cumbersome step in the systematic review, as the content, results, and surrounding theories become central and generic guidelines can only be adopted to a certain extent. A challenge in the thematic synthesis was that educational and psychological studies often boasted a high number of variables and investigated complicated relations. We found that only few studies included the same social goals and academic outcomes, and the ones that did often reported contradictory results. Although "vote counting" has received some criticism, Popay et al. (2006) describe it as a useful descriptive tool in which studies are categorized as showing significant or non-significant results. However, due to the reported influences of various individual and contextual factors drawing such simple conclusions was not easy. To do justice to the articles, we additionally had to find a balance between elaborating and highlighting the key results. A common fallacy during the synthesis is simply summarizing the findings from each study, without reaching a meta-perspective. Siddaway et al. (2018) maintain that the findings need to be interpreted, integrated, and critiqued in order to advance theoretical understanding.

## 8 Risk of Bias and Quality Assessment

Upon inspection of popular risk of bias (e.g., Cochrane Collaboration Risk of Bias Tool) and quality assessment tools (e.g., NHLBI and STROBE checklists), we found these to be unsuitable for the majority of articles included in the current systematic review. We were unable to apply these tools,

originally developed for randomized controlled trials in the health sciences, without tweaks to non-experimental social science studies. Revising these tools was deemed beyond the scope of the current systematic review. Interestingly, a moderate portion of systematic reviews do not conduct risk of bias analyses and many syntheses remain uninformed by the results of such analyses (Katikireddi et al. 2015). Some authors and methodologists reject the idea that a quality assessment needs to be conducted for articles that are included in configurative reviews, instead highlighting the need to prioritize relevance and contribution towards the synthesis (see Gough et al. 2012). As our review attempts to explore and generate theories on social goals in the academic context, we place a higher value on emergent concepts through a range of study contributions than precision by avoiding bias.

Furthermore, for a study to be included in our review it needed to be published in a peer-reviewed journal. Peer-reviews help validate research and raise the quality of articles by increasing robustness, legibility, and usefulness (Springer International Publishing AG 2018). Peer-reviews usually address aspects reflected in traditional quality assessment tools, such as reporting, validity, statistical tools, and interpretations (see Ramos-Álvarez et al. 2008). Although there is no guarantee that individual peer-reviewers adequately scrutinize each article, it has become a well-established method that the scientific community relies on. Quality assessment and risk of bias tools can also not account for frequently committed questionable research practices, such as selective reporting of variables, rounding down p-values, adjusting hypotheses after analyzing results, or falsifying data (see John et al. 2012).

## 9 Quality Assurance of the Systematic Review

The PRISMA statement is not a quality assurance instrument but does provide authors with a guide on how to transparently and excellently report their systematic review (Moher et al. 2009). The checklist provides a simple list with points corresponding to each section of the review (e.g., title—convey the type of review, information sources—name all databases and date searched). The majority of the items can be easily implemented, even for reviews within the field of psychology and education research. We followed this checklist and only deviated on certain points, such as those that referred to PICOS as it does not align with our review question. PICO(S) is limited in its applicability to reviews whose aim is not to assess the impact of an intervention (Brunton et al. 2017).

## 10 Experience and Communication

Some guidelines on systematic reviews propose that authors not only provide detailed descriptions of the review process, but also information about their experience with systematic reviews (see Atkinson et al. 2015). Our review team consisted of the two authors who worked closely together throughout the process. The second author has published and supervised numerous systematic reviews, whilst this is the first systematic review conducted by the first author. While an expert brings knowledge and skills, a novice viewpoint can ensure that the continuously advancing methods and tools for conducting a systematic review are incorporated into the process. As with any research endeavors, critical discussions, experience sharing, and help-seeking form part of the systematic review process. Working on a systematic review can at times feel tedious and endless, yet simply discussing the steps and challenges with others provides a new boost of enthusiasm. Conducting a systematic review is a time-consuming endeavor, which is not comparable to the process of writing an empirical article. Although numerous books and articles exist for self-study, having contact with an experienced author is invaluable.

## 11 Conclusion

This chapter details a current systematic review conducted in the realm of educational research concerning the role of social goals in academic adjustment and success. Unfortunately, reporting the findings of our systematic review and our synthesis is beyond the scope of the current chapter. Yet through methodological reflections and explicit descriptions, we hope to provide guidance and inspiration to researchers who wish to conduct a systematic review. In our example, we illustrate a possible strategy for keyword selection, setting selection criteria, conducting the study selection and data extraction. Once a precise question or aim has been set, selecting the keywords becomes a critical point with important consequences for the progression of the systematic review; unsuitable and/or limited keywords result in the loss of a comprehensive perspective that will pervade throughout the review. We recommend that selecting the keywords should be an iterative process, accompanied by careful consideration and reflection. Throughout the review process, each new stage should be

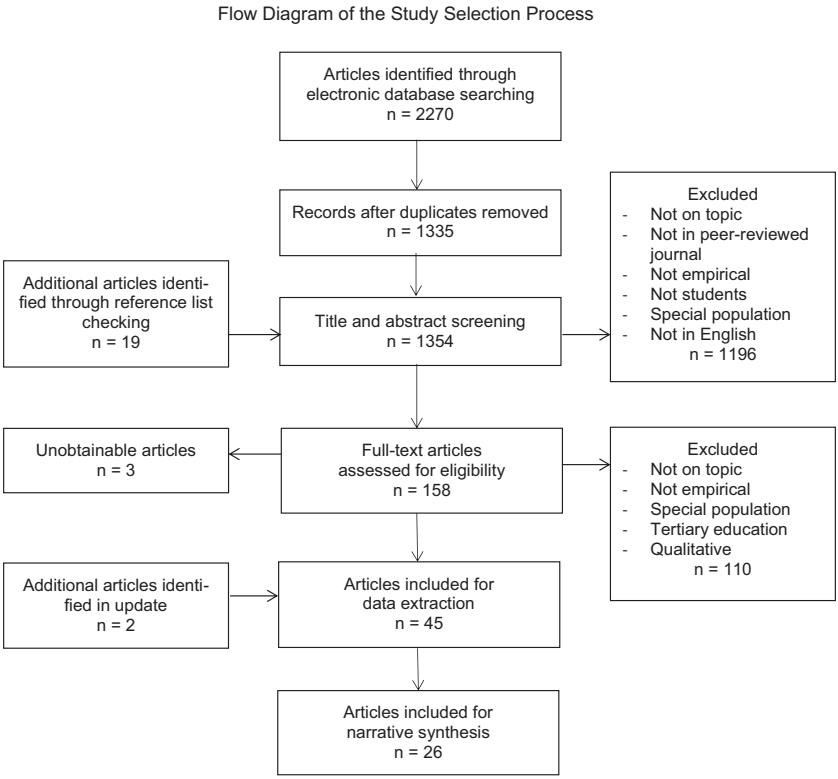
accompanied by a pilot phase to ensure appropriateness as new insights emerge (e.g., selection criteria, data extraction, thematic synthesis). We also wish to highlight the importance of moving beyond mere summarizing of studies in systematic reviews, and instead striving for a meta-perspective that allows for the results to contribute to a larger theoretical and practical context. Whilst conducting a configurative review, the initial protocol should not be viewed as a restraint; we were able to adjust the review process to the emerging needs and information obtained during the individual steps. Intensive reflection and meticulous documentation allow for necessary flexibility during the review process, whilst remaining systematic. The configurative approach is well suited for synthesizing the comprehensive and diverse studies often encountered in educational research, and could prove to be useful for future systematic reviews in the field.

## Appendix A

Search String

(“social goal*” or “interpersonal goal*” or “social status goal*” or “popularity goal*” or “peer preference goal*” or “agentic goal*” or “communal goal*” or “dominance goal*” or “instrumental goal*” or “intimacy goal*” or “prosocial goal*” or “social responsibility goal*” or “relationship goal*” or “affiliation goal*” or “social achievement goal*” or “social development goal*” or “social demonstration goal*” or “social demonstration-approach goal*” or “social demonstration-avoidance goal*” or “social learning goal*” or “social interaction goal*” or “social academic goal*” or “social solidarity goal*” or “social compliance goal*” or “social welfare goal*” or “belongingness goal*” or “individuality goal*” or “self-determination goal*” or “superiority goal*” or “equity goal*” or “resource acquisition goal*” or “resource provision goal*” or “in-group cohesion goal*” or “approval goal*” or “acceptance goal*” or “retaliation goal*” or “hostile social goal*” or “revenge goal*” or “avoidance goal*” or “relationship oriented goal*” or “relationship maintenance goal*” or “control goal*”) and (academic or school or classroom)

## Appendix B




## References

Anderman, L., & Anderman, E. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology, 24*(1), 21–37.

Atkinson, K., Koenka, A., Sanchez, C., Moshontz, H., & Cooper, H. (2015). Reporting standards for literature searches and report inclusion criteria: Making research syntheses more transparent and easy to replicate. *Research Synthesis Methods, 6*(1), 87–95.

Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology, 9*, 1:59.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215.

Booth, A. (2011). Chapter 3: Searching for studies. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions*. London, UK: Cochrane. Retrieved from http://cqrmg.cochrane.org/supplemental-handbook-guidance.

Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2017). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 93–122). London: Sage.

Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York: University of York. Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.

Chulef, A., Read, S., & Walsh, D. (2001). A hierarchical taxonomy of human goals. *Motivation and Emotion, 25*(2), 191–232.

Covington, M. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*(1), 171–200.

Dawes, M. (2017). Early adolescents' social goals and school adjustment. *Social Psychology of Education, 20*(2), 299–328.

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research & Policy, 10*(1), 45–53.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D., Miller, T., Sutton, A., … Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research, 6*(1), 27–44.

Dowson, M., & McInerney, D. M. (2001). Psychological parameters of students' social and work avoidance goals: A qualitative investigation. *Journal of Educational Psychology, 93*(1), 35–42.

Dweck, C. (1992). Article commentary: The study of goals in psychology. *Psychological Science*, *3*(3), 165–167.

Erdley, C., & Asher, S. (1999). A social goals perspective on children's social competence. *Journal of Emotional and Behavioral Disorders*, *7*(3), 156–167.

Falagas, M., Pitsouni, E., Malietzis, G., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google scholar: Strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342.

Ford, M. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage.

Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews, 1*:28.

Grant, C., & Osanloo, A. (2014). Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house". *Administrative Issues Journal, 4*(2), 12–26.

Hagger, M. (2014). Avoiding the "déjà-variable" phenomenon: Social psychology needs more guides to constructs. *Frontiers in Psychology*, *5*, 1–2.

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Katikireddi, S., Egan, M., & Petticrew, M. (2015). How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *Journal of Epidemiology and Community Health, 69*(2), 189–195.

Kiefer, S., & Ryan, A. (2008). Striving for social dominance over peers: The implications for academic adjustment during early adolescence. *Journal of Educational Psychology, 100*(2), 417–428.

King, R., & Ganotice, F. (2014). The social underpinnings of motivation and achievement: Investigating the role of parents, teachers, and peers on academic outcomes. *Asia-Pacific Education Researcher, 23*(3), 745–756.

King, R., McInerney, D., & Watkins, D. (2012). Studying for the sake of others: The role of social goals on academic engagement. *Educational Psychology, 32*(6), 749–776.

Lemos, M. (1996). Students' and teachers' goals in the classroom. *Learning and Instruction, 6*(2), 151–171.

Levinsson, M., & Prøitz, T. (2017). The (non-)use of configurative reviews in education. *Education Inquiry, 8*(3), 209–231.

Mansfield, C. (2009). Managing multiple goals in real learning contexts. *International Journal of Educational Research, 48*(4), 286–298.

Mansfield, C. (2010). Motivating adolescents: Goals for Australian students in secondary schools. *Australian Journal of Educational and Developmental Psychology, 10*, 44–55.

Mansfield, C. (2012). Rethinking motivation goals for adolescents: Beyond achievement goals. *Applied Psychology, 61*(1), 564–584.

McCollum, D. (2005). What are the social values of college students? A social goals approach. *Journal of College and Character, 6*, Article 2.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 6*, e1000097.

Morgan, D. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76.

Owen, K., Parker, P., Van Zanden, B., MacMillan, F., Astell-Burt, T., & Lonsdale, C. (2016). Physical activity and school engagement in youth: A systematic review and meta-analysis. *Educational Psychologist, 51*(2), 129–145.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., … Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews.* ESRC Methods Programme: University of Lancaster, UK.

Ramos-Álvarez, M., Moreno-Fernández, M., Valdés-Conroy, B., & Catena, A. (2008). Criteria of the peer review process for publication of experimental and quasi-experimental research in Psychology: A guide for creating research papers. *International Journal of Clinical and Health Psychology, 8*(3), 751–764.

Roussel, P., Elliot, A., & Feltman, R. (2011). The influence of achievement goals and social goals on help-seeking from peers in an academic context. *Learning and Instruction, 21*(3), 394–402.

Ryan, A., & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction, 21*(2), 247–256.

Sheldon, T. (2005). Making evidence synthesis more useful for management and policy-making. *Journal of Health Services Research & Policy, 10*, S1:1–5.
Shim, S., Cho, Y., & Wang, C. (2013). Classroom goal structures, social achievement goals, and adjustment in middle school. *Learning and Instruction, 23*, 69–77.
Siddaway, A., Wood, A., & Hedges, L. (2018). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*. Advance online publication. doi.org/https://doi.org/10.1146/annurev-psych-010418-102803.
Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness, 4*(3), 409–429.
Solmon, M. (2006). Goal theory in physical education classes: Examining goal profiles to understand achievement motivation. *International Journal of Sport and Exercise Psychology, 4*(3), 325–346.
Springer International Publishing AG. (2018). *Peer-review process.* Retrieved from https://www.springer.com/gp/authors-editors/authorandreviewertutorials/submitting-to-a-journal-and-peer-review/peer-review-process/10534962.
Urdan, T., & Maehr, M. (1995). Beyond a two-goal theory of motivation and achievement: A case for social goals. *Review of Educational Research*, *65*(3), 213–243.
Wentzel, K. (1993). Motivation and achievement in early adolescence: The role of multiple classroom goals. *Journal of Early Adolescence, 13*, 4–20.
Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, *48*(2), 311–325.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

View publication stats