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 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 End-Note. 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/10.1080/13600 80x.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_con cise.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/10.108 0/13562517.2014.901956.