Conducting a Systematic Review: Demystification for Trainees in Sport and Exercise Psychology

David Tod & Martin Eubank

School of Sport and Exercise Sciences

Liverpool John Moores University

Abstract

The purpose of this article is to define and detail the steps in conducting systematic reviews for trainees and supervisors. We also offer suggestions garnered from our experiences reading, conducting, publishing, and reviewing such manuscripts. Steps include: developing specific questions and inclusion/exclusion criteria; undertaking a multi-strategy literature search; implementing replicable data extraction methods; assessing study quality; and employing transparent procedures for synthesising and presenting results. Suggestions include: developing a proposal and having it reviewed, allowing sufficient time to conduct a review, keeping meticulous records, and adhering to established procedures.

Conducting a Systematic Review: Demystification for Trainees in Sport and Exercise Psychology

Systematic reviews involve the application of a transparent and systematic method to synthesise research and reduce reviewer bias (Lichtenstein, Yetley, & Lau, 2008). Transparency emerges from the documentation of decision making procedures and public availability of the data search, extraction, and analysis results. Reviewers attempt to reduce bias by describing and following a predetermined replicable protocol. Systematic review procedures have a long history in the medical and health communities to underpin practice guidelines, identify research directions, and evidence consensus statements (Lichtenstein et al., 2008). The employment of systematic reviews in sport and exercise psychology has a less developed tradition, but these types of reviews are becoming more popular in the literature. For example, in its first volume, none of the 11 articles in the *International Review for Sport and Exercise Psychology* were systematic reviews, as defined above. In 2016, 8 of the 11 articles had some degree of systematic review 'flavour'. One reason for their increasing popularity is that systematic reviews are considered of higher quality than narrative reviews, a view influenced by the "black box" characteristic of the latter compared with the transparency of the former. As a second reason, in addition to synthesising knowledge, systematic reviews allow for the rigorous assessment of the quality of the underpinning research. Nevertheless, based on our experiences garnered from our editorial and educational roles, we have found that there is misunderstanding among a number of sport and exercise psychology trainees, supervisors, and educators about what is involved in designing, implementing, writing, and publishing systematic reviews. For example, as part of the Stage 2 Qualification in Sport and Exercise Psychology, trainees may opt to conduct a systematic review to evidence the threshold level of doctoral level research competence required for the award of the qualification. Based on the reviews that have been done to date,

assessor feedback has frequently brought into question the systematic nature of the trainee's work, and resubmission requirements designed to strengthen the systematic elements of the review systematic are common-place. The purpose of the current article is first, to overview the process involved, and second, to offer advice on how to undertake systematic reviews. It is hoped that its content will serve to demystify the systematic review and help trainees and supervisors alike to conduct them more effectively.

According to the *Cochrane handbook for systematic reviews of interventions* (Higgins & Green, 2011):

> "A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made" (section 1.2.2).

Higgins and Green further suggest systematic reviews (a) have clearly stated objectives, (b) have pre-defined criteria for study inclusion and exclusion, (c) have an explicit and reproducible method, (d) involve a systematic search that attempts to identify all eligible studies, (e) include an assessment of study quality and the validity of their findings, and (f) a systematic synthesis and presentation of the included studies' characteristics, findings, and quality. Some people consider a systematic review to be the same beast as a meta-analysis, but the two are different. A meta-analysis involves the use of statistical procedures designed to synthesise the results from primary research (Borenstein, Hedges, Higgins, & Rothstein, 2009). It is possible to use these procedures to summarise any set of studies, including those that have not been located through systematic procedures (in which case the results are likely to be inaccurate and biased). A systematic review does not need to include statistical procedures to synthesis findings, whereas high quality meta-analyses will be conducted on a

dataset located through systematic means. It is also possible to conduct a systematic review on qualitative studies (Paterson, Thorne, Canam, & Jillings, 2001), either on their own or in combination with quantitative investigations.

Steps Involved in a Systematic Review

Higgins and Green's (2011) description above highlights the steps involved in conducting a systematic review and includes: (a) developing a set of clear, specific questions; (b) the use of predetermined inclusion and exclusion criteria for study selection; (c) the undertaking of a systematic multi-strategy search; (d) implementing an explicit and replicable method for data extraction; (e) an assessment of study quality; and (f) employment of procedures for synthesising and presenting results. In the following paragraphs, we will briefly describe these steps, based on our experiences. Our intention is not to document definitively how to conduct a systematic review, but instead share reflections that may help trainees and supervisors as they fulfil the requirements of their doctoral level training.

Develop clear and specific purposes. We have observed that reviewers may make the scope of their reviews too vague. Focusing on specific purposes as much as possible helps individuals identify that in which they are most interested in examining and helps a great deal in directing feasible reviews. For example, rather than "review self-talk research in sport," a more focused specific purpose might be to "synthesise the experimental literature examining the influence of self-talk interventions on motor skill execution." Focused purposes help professionals make decisions in the development and implementation of their reviews. For example, based on the above self-talk example, the inclusion and exclusion criteria are likely to state that experimental rather than descriptive investigations will be included.

Well-conducted systematic reviews are time and labour intensive, and they involve an opportunity cost (e.g., an academic might have spent the time developing teaching materials

or conducting another type of investigation). The resource expense involved often needs to be justified, and it is worthwhile constructing the rationale for the review, especially if there is a desire to publish it. Two guiding considerations include: knowledge advancement and real world impact. Strong justifications emerge when reviewers detail the specific ways in which their work will advance knowledge and help others. For example, authors might argue that synthesising the self-talk intervention research may estimate numerically the effect and variability of such strategies, along with identifying moderating variables (representing knowledge advancement). Further, such information may inform practice guidelines (signifying potential impact).

There are existing tools to help authors frame (and then answer) their review questions. One widespread example is the PICO acronym, standing for Problem/population, Intervention, Control, and Outcome (Higgins & Green, 2011). To illustrate, authors might decide to review research focused on athletes (population) using specified types of self-talk (intervention) in research where its efficacy was compared against acceptable control conditions (control) for its influence on objectively measured motor movements (outcome). The PICO acronym is suitable for intervention research and other tools are available for difference types of literature and purposes. In addition to helping authors focus their review questions, these tools can also help people in other ways, such as developing inclusion/exclusion criteria and identifying search terms.

Identify explicit inclusion and exclusion criteria. The development of clear explicit inclusion and exclusion criteria is possibly one of the more neglected components in the review process. Similar to the review's guiding question, however, inclusion/exclusion criteria may influence the scope and quality of the results (Card, 2011). If not considered, then authors may apply implicit criteria without realising or discussing notable consequences. For example, a large proportion of systematic reviews in sport and exercise psychology are

limited to articles published in English. Inclusion of articles written in other languages is not normally feasible, because reviewers may not have the ability to read them and translation costs can be prohibitive, yet studies published in other dialects may be relevant (e.g., Tod & Edwards, 2015). Failure to include these studies may be a form of publication bias. Rather than ignoring limitations, such as language, openly discussing the boundaries of a systematic review may assist others in identifying solutions that stimulate future knowledge advancements.

The PICO tool described above, along with its compatriots, can help authors develop suitable inclusion/exclusion criteria. Additional issues to consider include the articles' publication year, their participants' country or cultural origins, and their methodological rigour (Meline, 2006). Addressing these issues allows reviews to be replicable, a commonly viewed desirable feature of the genre. Even if replication is not desired (or possible), addressing these issues helps readers interpret the review's contribution to knowledge. As discussed below, for example, tools exist to assist reviewers in evaluating the methodological quality of included/excluded studies. Some reviewers might decide to exclude studies based on the methodological assessment due to their desire to summarise only the highest quality evidence available. Others writers argue that including all studies that have been located, but then assessing the influence methodological features have had on results may inform interpretations or future research design (Card, 2011). The previous example illustrates that developing inclusion and exclusion criteria is not straightforward or uncontentious. Nevertheless, when combined with a focused review question, time spent developing explicit criteria can assist the next step: the literature search. Given the transparency expected in these types of articles, journal reviewers and Stage 2 assessors would expect to see justification regarding the stated inclusion/exclusion criteria.

Undertake a systematic and multi-strategy search. Given the volume of research available it is necessary to use electronic databases to identify relevant studies. Electronic searching is a skill that develops with practice and some review teams consult or include librarians, because they can make a valuable contribution to projects. Review teams will typically search multiple databases, because such archives do not always overlap with each other. Examples of common databases used in sport and exercise psychology include SPORTDiscus, Web of Science, PubMed, PsychINFO, and PsychARTICLES. It is also advisable to include databases that reference 'grey' literature, such as Open Grey and PsycEXTRA. It is beyond the scope or the space of the current article to provide specific guidance on electronic database searching, because there are several components and each may involve decision making and piloting. For example, identifying suitable lists of keywords may require discussions with stakeholders and piloting. In addition, search strategies often need to be tailored to specific search engines (Egan, MacLean, Sweeting, & Hunt, 2012).

Electronic databases are not exhaustive, however, with estimates of the percentage of studies included in systematic reviews coming from these sources ranging from 40-80% (Hopewell, Clarke, Lefebvre, & Scherer, 2007). Best practice involves a combination of electronic and various hand search methods. One hand search strategy is a review of journals' tables of contents and indexes. Another practice is a backward search where reviewers trawl through located articles' reference lists. Some electronic databases allow investigators to do both a backward and forward search (where articles citing a located study can be viewed). It is also possible to contact recognised authors requesting any unpublished research that might be relevant (as noted below, it is often useful to contact authors to request information that may be missing from publication that if not provided will lead to exclusion).

Systematic reviews attempt to locate and synthesise all the relevant literature on a topic. Although a desirable goal, based on our experience it is unrealistic for numerous reasons: databases are incomplete, researchers may not disseminate their work publically, published work (especially older studies) may lack necessary information for them to be included, authors may not respond to email, and reviewers may not have the funds or ability to include reports written in languages other than their own. The extent to which reviewers have surveyed the total population of relevant studies is unknowable in most cases, but they will enhance their work's credibility by acknowledging limitations and demonstrating they have taken the necessary steps to include as much as reasonably possible.

The literature search is not a onetime discrete step in the process. Instead searching is an ongoing process that occurs throughout the project's lifecycle. Also, authors may find that they refine their search strategies as they proceed. They might uncover new keywords, stumble across previously unknown archives, or need to modify their inclusion/exclusion criteria. With the way many systematic reviews are presented, there is the impression that authors created the optimal search strategy 'off the bat'. Perhaps teams of experienced researchers working in fields they know well are able to develop good protocols that work perfectly first time. However, trainees and other inexperienced reviewers may need practice and it is likely their protocols would benefit from pilot testing and modification. It can also build readers' confidence in the project's rigour if authors implement checks to assess the efficacy of their search, such as comparing results from team members working independently or with previous reviews.

Implement an explicit and replicable data extraction method. Systematic review data can include located publications' citation information, methodological features, conceptual characteristics, results, and any other attributes necessary to answer the review questions in a clear informed fashion. The number of included studies can range from a few

to hundreds depending on the scope of the questions being addressed. Employing data extraction tools help to ensure consistency across time and review teams. Data extraction accuracy does not increase with systematic review experience (Horton et al., 2010). There are numerous tools available ranging from paper-based crib sheets to web-based software packages, but they are only as helpful as the person using them. 'Off the shelf' tools may be best considered a starting point because they may not be entirely suited to the specific project being conducted: pilot work and adjustments may be needed. Issues warranting consideration include: initial set up and running costs, versatility, training requirements, portability, and ability to store, analysis, and present data (Elamin et al., 2009).

We have found that students and trainees sometimes underestimate the challenges involved in data extraction. Even with crib sheets and explicit decision making rules, it may be difficult to obtain data. For example, empirical research reports, especially older ones, frequently do not contain information that might be expected to be included, such as descriptive statistics (e.g., authors may sometimes only report whether *P* was greater or less than .05 or those results supporting their hypotheses). Details may be reported inconsistently, such as the sample size being reported differently in the abstract, methods, and results. When reviewing a large number of studies there is the challenge of maintaining attention to detail. We believe many students would benefit from greater attention to detail and patience. Having some method of data extraction evaluation can help reviewers check their consistency and accuracy. Stage 2 trainees could consider, for example, collaborating with other trainees, supervisors, or colleagues in conducting their review as a way of checking their progress.

Assess study quality and use the resulting information. Most readers will recognise that variation in empirical study quality can influence the results. The degree of confidence in an investigation's method will enhance trust in the results (Ryan, Hill, Prictor, & McKenzie, 2013). Similarly, the faith readers have in a systematic review will be

influenced by the quality of both the included investigations and review methodology. Fortunately, there exist numerous critical appraisal tools and checklists to help guide reviewers. Often these checklists facilitate a numeric score, a metric that provides an imperfect insight into the investigations' quality, relative to similar studies.

To illustrate, Timmer, Sutherland, and Hilsden (2003) produced a checklist of 19 items against which a study can be evaluated numerically. Items focused on topics such as question clarity, use of randomization, control of confounds, and presentation of results. The checklist is flexible and scoring can be tailored to suit reviewers' needs. Although initially designed for conference abstracts, the checklist has been adopted by systematic reviewers. Initial evidence suggested the tool was associated with high inter-rater reliability. Although we have found Timmer et al.'s tool useful, it is one of many that exist. A Google search will yield various examples from which trainees can select one that suits their purposes.

One common omission from many systematic reviews is the *analysis* of study quality results. Having assessed study quality, which often involves numerical scoring, authors vary considerably in the degree to which they make use of this information. Some might list the information when presenting each investigation's characteristics and leave the interpretation to readers, and others may not even present or refer to the information in the results or discussion sections. We believe that in such cases authors are doing themselves a disservice. They have probably spent considerable time obtaining this information, yet we would wonder why they did so if they then failed to use the data obtained. Their reviews could certainly have made a greater contribution and impact if they had drawn on the information. In a meta-analysis, for example, information from a quality assessment (e.g., experimental design, questionnaire reliability, participant type, control condition) could be used to identify methodological moderators. In the synthesis of qualitative studies, reviewing researchers' adopted methodological theoretical orientations may help inform the interpretation of results.

Explicitly detail the data synthesis and presentation method employed. Another way that we believe authors have done themselves a disservice in systematic review presentation is by just describing the results of their searches. They might, for example, produce a table that presents the located studies' characteristics, which are then briefly summarised in the text. These articles would have greater impact if they did more than simply assemble the raw results from their searches. A strong systematic review also includes the more creative aspects of synthesizing, analysing, and interpreting the raw data. At a minimum a review might be expected to include (a) a description of the search results, (b) a presentation of the located studies' characteristics, (c) an attempt to analyse the data, and (d) the authors' interpretation of the analysis. The reader is better able to evaluate the outcomes of the results and discussion if the underpinning decision making process has been stated explicitly and transparently rather than being left implicit.

Sometimes reviewers are able to refer to published guidelines on data analysis and presentation. For example, there are several textbooks presenting details on how to conduct a meta-analysis. Further, given the transparency principle, ideally authors could also replicate methods used in other reviews. One feature of systematic approaches is the scope for creative and new ways of analysing and presenting data. It is permissible, for example, to present results as text, as numbers, and in graphical ways as long as authors provide a clear, concise, and informative answer to the question and they conform to requirements of any stakeholder, such as a funding body or journal editor. Common examples of data analysis in sport and exercise psychology include meta-analysis, vote counting, and qualitative thematic analysis. Each has advantages and disadvantages and is suited to specific questions and types of literature. For example, although meta-analysis allows for precise estimates of effect size in quantitative research, the procedures do not help reviewers provide rich description about a phenomenon examined in qualitative investigations.

There are several guidelines available to help reviewers write their reports and include the information needed to inform readers and policymakers. One such example is the PRISMA checklist (PRISMA stands for Preferred Reporting Items for Systematic reviews and Meta-Analyses). The checklist contains items that detail the types of information expected to be included in a report, and there are extensions designed for various types of documents, such as protocols and abstracts. Some journals are now asking authors to include the checklist as part of a manuscript submission.

Reflections and Suggestions for Trainees

Based on our combined experiences of publishing, supervising, and assessing / moderating numerous systematic reviews (of varying quality), we have listed below some observations and suggestions that we think will assist trainees in completing 'passable and publishable' systematic reviews.

Write a protocol and have it critiqued. Good protocols detail specifically and concretely the review questions, rationale, methods, analyses, and presentation plans that will result in the final report. Many trainees (and some supervisors!) will probably not have previously undertaken a systematic review and will likely be unaware of potential challenges and obstacles. They might also identity topics that have been previously reviewed, might not be worth doing, or may not be feasible. Writing a protocol may help avoid these pitfalls. We have also found that would-be reviewers are typically not specific enough in identifying their questions and methods. Failure to focus these details sufficiently leads to difficulties (sometimes fatal) further down the line. A Google search will quickly locate proposal examples and templates to assist trainees.

Once a protocol has been written, trainees can seek formal feedback from knowledgeable peers and supervisors. Even experienced folks may overlook details, especially if reviewing unfamiliar bodies of literature. Constructive feedback will help

trainees expand their own knowledge and enhance the likelihood that they will conduct a high quality review.

We have come across a perception that systematic review protocols must be registered with some organisation prior to being undertaken, such as the International Prospective Register of Systematic Reviews (PROSPERO). Although there are benefits to registering systematic reviews prior to being conducted, such as allowing for peer review, we are unaware of any legislation that requires such mandatory action. A trainee could produce a review that satisfied requirements without registering their protocol. There are situations, however, when funding bodies, journal editors, or other stakeholders might stipulate some form of registration. Trainees hoping to publish their work could avoid difficulties by considering dissemination plans when planning their projects.

Consider the time needed to produce a document of high quality. We have come across colleagues and students who may view systematic reviews as a "soft publication" that are easy to do and a quick way to secure a highly valued and cited article. These attitudes quickly change once the individual embarks on the process, realising that is it hard, or at least far from soft. For example, it can be demoralising when an individual suddenly uncovers an obscure database that contains a large number of relevant documents that have not been identified in more conventional catalogues. High quality systematic reviews are increasingly held in high regard by the research community, so they are worth doing, and worth doing well.

It may be possible to do a systematic review or meta-analysis quickly if the questions and inclusion/exclusion criteria are so narrow that a search yields a small number of studies (we have seen reviews published with an $N = 4$). It is also possible, however, that such reviews do not add a great deal to the literature and the outcome does not justify the effort expended (which would include the authors', editors', and reviewers' time). There may not

be enough information to provide a meaningful interpretation. For example, there is a misperception that meta-analyses have high power to detect significant findings, however, there are times when they can have quite lower power, such as when there are few primary studies and great between-study variance (Borenstein et al., 2009). There is a trade-off between question specificity and worthiness that individuals need to balance. In situations where there are not many studies, reviewers need to ensure that they have a strong rationale: that the information gained will make a useful increment to knowledge or will have an impact on practice.

Engage in meticulous record keeping. The aim of systematic reviews is to assess as much of the relevant research as possible and provide detailed informative answers to specific questions. Readers expect to see levels of precision not normally associated with narrative reviews and to have opportunities to examine the raw results from data extraction for themselves, if they so wish. Meticulous record keeping is needed to fulfil such expectations. Record keeping might include, for example, the number of hits from a database on a particular date using an identified search strategy, the percentage break down of the reasons certain studies were excluded, or the number of participants failing to complete intervention regimes. In our experience the attention to detail needed sometimes dampens trainees' enthusiasm for a project, however, the use of crib sheets during data extraction and spread sheets can scaffold efficiently their time and effort expended. Several websites contain crib sheets that can be downloaded for various types of studies, and software is also available to help trainees keep track of their projects.

Follow a recognised systematic review framework and read existing reviews. Trainees do not need to reinvent the systematic review wheel. Guidelines exist that can provide assistance, and there are many examples of excellent reviews that can provide inspiration. When first learning a skill, individuals benefit from observing models to help

them develop a cognitive map of what they wish to produce, and this observation applies to trainees conducting a review. Once individuals have completed several reviews they may have the confidence to be creative in the way they conduct their reviews in the future. This takes time, and is one of those areas of professional development that can truly be regarded as continuous.

Conclusion

The purpose of the current article has been to provide an overview of the systematic review process to help trainees, including Stage 2 candidates, complete their qualification. Although there is considerable variety in publically available systematic reviews that reflects the adaptability of the process to suit different purposes, stakeholder needs, and types of literature, there are also numerous textbooks, websites, and other resources that can help individuals plan, conduct, and write up their projects. We hope that the steps discussed above are informative to trainees, supervisors, and assessors alike in demystifying the systematic review, and alleviate some of the misunderstandings that exist around the systematic review product and process so that authors can be satisfied and proud of their end product.

References

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

Card, N. A. (2011). *Applied meta-analysis for social science research.* New York, NY: Guilford.

Egan, M., MacLean, A., Sweeting, H., & Hunt, K. (2012). Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: a prospective comparative study of literature search methods. *BMJ Open, 2,* e001043.

Elamin, M. B., Flynn, D. N., Bassler, D., Briel, M., Alonso-Coello, P., Karanicolas, P. J., . . . Kunz, R. (2009). Choice of data extraction tools for systematic reviews depends on resources and review complexity. *Journal of clinical epidemiology, 62*, 506-510.

Higgins, J. P. T., & Green, S. (2011). Cochrane handbook for systematic reviews of interventions: version 5.1.0 [updated March 2011] Retrieved from www.cochrane-handbook.org

Hopewell, S., Clarke, M., Lefebvre, C., & Scherer, R. (2007). Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database Syst Rev, 2,* Art. No.: MR000001. doi: 10.1002/14651858.MR000001.pub2.

Horton, J., Vandermeer, B., Hartling, L., Tjosvold, L., Klassen, T. P., & Buscemi, N. (2010). Systematic review data extraction: Ccross-sectional study showed that experience did not increase accuracy. *Journal of Clinical Epidemiology, 63*, 289-298.

Lichtenstein, A. H., Yetley, E. A., & Lau, J. (2008). Application of systematic review methodology to the field of nutrition. *The Journal of Nutrition, 138*, 2297-2306.

Meline, T. (2006). Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders, 33*, 21-27.

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: Sage.

Ryan, R., Hill, S., Prictor, M., & McKenzie, J. (2013). Cochrane consumers and communication review group: Study quality guide. Retrieved from http://cccrg.cochrane.org/authorresources

Timmer, A., Sutherland, L. R., & Hilsden, R. J. (2003). Development and evaluation of a quality score for abstracts. *BMC medical research methodology*, *3*, 2.

Tod, D., & Edwards, C. (2015). A meta-analysis of the drive for muscularity's relationships with exercise behaviour, disordered eating, supplement consumption, and exercise dependence. *International Review of Sport and Exercise Psychology, 8*, 185-203. doi: 10.1080/1750984X.2015.1052089