

This is the peer-reviewed, accepted manuscript of an article published by Taylor & Francis in the *International Journal of Social Research Methodology* on 9 October 2012, available online (<https://doi.org/10.1080/13645579.2012.730704>):

Daigneault, Pierre-Marc, Steve Jacob et Mathieu Ouimet (2012). 'Using systematic review methods within a Ph.D. dissertation in political science: Challenges and lessons learned from practice'. *International Journal of Social Research Methodology*, 17 (3), 267-283.

Using systematic review methods within a PhD dissertation in political science: Challenges and lessons learned from practice¹

Abstract: Systematic review and synthesis methods have gained wide acceptance within the social sciences and, as a result, many postgraduate students now consider using them for their thesis or dissertation research. However, students are rarely aware of all the concrete implications that their decision entails. This reflective narrative reports the experience of a political science student who began to conduct a systematic review as part of his PhD dissertation but who did not complete it. The aim of this article is to identify challenges and lessons learned from this experience and to formulate recommendations for postgraduate students who wish to make an informed choice with respect to the use of these methods.

Keywords: Systematic literature review; scoping study; systematic map; research synthesis; social science practice; empirical research; reflective narrative.

Systematic reviews and the evidence movement

Under the leadership of organizations such as the Cochrane and Campbell Collaborations and the Evidence for Policy and Practice Information and Co-ordinating Centre (**EPPI-Centre**), and what is more generally labelled the 'evidence movement', systematic literature reviews, rapid reviews, scoping studies and research syntheses are becoming increasingly popular in the social sciences (Arksey & O'Malley, 2005; Davies, Nutley & Smith, 1999; Grant & Booth, 2009; Hansen & Rieper, 2009; Learmonth & Harding, 2006; Pawson, 2002; Petticrew & Roberts, 2006; Sandelowski & Barroso, 2006; Sanderson, 2002). The evidence movement has permeated various disciplines and policy sectors with differentiated force and timing. Although health-related disciplines such as medicine, health policy and nursing were the first to be swept up by this new paradigm, social science fields such as education (Oakley, 2003), management (Tranfield, Denyer & Smart, 2003) and public administration (Heinrich, 2007) are now 'full members' of the evidence movement. This popularity is quite understandable since systematic reviews have a comparative advantage over 'traditional' or 'conventional' literature reviews with respect to the validity of their conclusions (Littell, Corcoran & Pillai, 2008; Oakley, 2003):

Literature reviews, even those written by experts, can be made to tell any story one wants them to, and failure by literature reviewers to apply scientific principles to the process of reviewing the evidence, just as one would to primary research, can lead to biased conclusions, and to harm and wasted resources. Yet, traditional literature reviews frequently summarize highly unrepresentative samples of studies in an unsystematic and uncritical fashion ... (Petticrew & Roberts, 2006, p. 5)

In essence, what distinguishes systematic reviews from conventional literature reviews is the fact that they conform to the methodological standards used in primary research, namely transparency, rigor, comprehensiveness and reproducibility (Briner, Denyer & Rousseau, 2009; Denyer & Tranfield, 2009; Littell, Corcoran & Pillai, 2008). It 'is a specific methodology that locates existing studies, selects and evaluates contributions, analyses and synthesizes data, and reports the evidence in such a way that allows reasonably clear conclusions to be reached about what is and is not known' (Denyer & Tranfield, p. 672). Once research questions are specified, conducting a systematic review generally involves the following stages: identifying references, selecting relevant references, assessing study quality, data extraction, and data synthesis (Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007).

An approach which appeal to students

Even though the enthusiasm for systematic review is not universally shared (see Hammersley, 2001; 2008), the influence of this class of methods on scientific practices is beyond dispute. It is not surprising, therefore, that many master- and doctoral-level students consider conducting a systematic review as part of, or even as, their thesis or dissertation. The proportion of students using systematic review and synthesis methods for their postgraduate research project will certainly increase as these methods become more widely known and as their value is increasingly recognized in the social sciences. Besides the enhanced validity of their conclusions, substantial benefits have been associated with the use of systematic review and synthesis methods by undergraduate and postgraduate students. First of all, Armitage and Keeble-Allen (2008) have found that students who have used these methods 'gained a greater depth and insight into the subject they were researching' (p. 103). Systematic and integrative reviews can also help students develop critical reasoning, problem-solving skills, methodological expertise and information-technology skills (Minnie, van der Walt, Klopper & Cummings, 2010; Sambunjak & Puljak, 2010). Moreover, PhD students involved in a Cochrane review may 'develop functional networks of mentoring and research partnerships' (Sambunjak & Puljak, 2010, p. 321). Last but not least, conducting a systematic review is a rewarding experience that may allow students to make a significant contribution to their field of study, even though they are relatively inexperienced scholars (Jones, 2004; Owens, Baez & Tillman, 2006; Perry & Hammond, 2002).

Yet, whereas the literature reporting practical experiences, challenges and lessons learned in using systematic review in various contexts has grown substantially in the last decade (see e.g., Atkins, Lewin, Smith, Engel, Fretheimm & Volmink, 2008; Brereton et al., 2007; Oakley, 2003), the number of studies which specifically focus on students' experience in conducting systematic reviews as part of their degree is still limited, especially in the social sciences. This constitutes a serious problem on two grounds. For one thing, students have few models and guidelines adapted to their specific situation that they can rely on. Most reported students' experiences with systematic reviews are 'positive' in the sense that, despite some real difficulties and challenges, they were successfully completed. However, a lot can be learned from less successful cases, especially as students who consider the systematic review option are rarely aware of all the concrete and serious implications that their decision entails. For instance, the prescription 'be patient' (see Owens et al., 2006) when conducting a review is certainly wise in some contexts but totally inappropriate in others, as it will soon be clear. For another, systematically reviewing and synthesizing the literature in social science poses unique challenges due to the fragmented nature of many disciplinary fields (Denyer & Tranfield, 2009; Tranfield, Denyer & Smart, 2003). Many systematic reviews in the health sector are conducted under the leadership of the Cochrane Collaboration which subscribes to an 'evidence hierarchy' based on the internal validity of the

inferences derived from particular study designs. In practice, most Cochrane reviews only include studies that have used RCTs (Hansen & Rieper, 2009). While the Campbell Collaboration has a distinct social science orientation (i.e., it covers fields such as criminology, education and social services), its methodological standards and practices follow a 'mimetic process' based on those of the Cochrane Collaboration (Hansen & Rieper, 2009, p. 146).² By contrast, there is less consensus on what constitutes appropriate epistemology and methodology in most social science fields and the use of RCTs is less common for feasibility and ethical reasons (Tranfield, Denyer & Smart, 2003). When one adopts the more 'ecumenical' perspective on varieties of evidence and study designs typically found in most social science disciplines, identifying, screening, assessing and synthesizing evidence becomes a significantly more complex endeavour. To be sure, these challenges do not exclusively impact students, but they can nevertheless represent a disproportionate burden for them since students are more likely to be less experienced, less knowledgeable about methodology and their substantive field and have fewer resources than more established scholars.

Study Purpose and Approach

This article presents a reflective narrative based on the experience of a political science student conducting a systematic review as part of his PhD dissertation. The case presented is not an exemplar. Various reasons — including time constraints, an insufficiently focused review scope and the scientific competition from peers who work on the same research topic — prevented the successful completion of the systematic review as originally planned. Useful lessons were nevertheless learned from the challenges posed by this review process as there is now a growing realization that more attention should be devoted by scholars and practitioners to less-than-successful experiences (in the field of evaluation, see for instance Gervais, 2012). Whereas some challenges were a direct consequence of the fact that the systematic review was conducted by a PhD student for his dissertation, others are general challenges found in the social science literature which took a distinctive 'flavour' in this context. The primary aim of this article is to share useful lessons with other social science students who are considering doing a systematic review for their thesis or dissertation to help them make an informed choice. That being said, the issues discussed in this article should also be relevant to social scientists who plan to conduct a systematic review.

The reflective narrative is a very appropriate type of inquiry for that purpose. Reflective narratives are

...essays grounded in the observations and experience of the narrator(s). These are entirely rich and vivid ways of knowing about practical phenomena and [...] have considerable value in uncovering the subtle nuances and the complex interrelationship among variables. (Cousins, 2005, pp. 203-204)

Reflective narratives are a form of autobiographical self-study in which the private experience of the self can engage and illuminate the experience of others who share the same practice setting. In order for self-study to be of significance to others (i.e., to provide an answer to the 'so what?' question), a fine balance must be struck between the self and the others (Bullough & Pinnegar, 2001).

After a short — and, at times, candid — presentation of the case background, the challenges and lessons learned are discussed in light of the practice-based and prescriptive literature on the application of systematic review methods. This article concludes with a recapitulation of the main lessons learned and a few recommendations for social science students considering conducting a systematic review as their thesis or dissertation.

Background: The review project

Although conducted within a political science department, the PhD dissertation described here was geared towards the interdisciplinary — or more accurately ‘transdisciplinary’ — field of policy and program evaluation (Scriven, 2004; Jacob, 2008). It focused on the non-utilisation or under-utilisation of evaluation by decision-makers, one of the most researched issues in the field (Christie, 2007; Henry & Mark, 2003). The original aim of this project, and the one defended before the student’s dissertation proposal committee, was to systematically review and synthesize the empirical literature on the factors associated with evaluation use.

Because this project was too broad in scope and a similar review on this topic had just been published (i.e., Johnson et al., 2009), it was decided that the dissertation be refocused on the influence of an important hypothesised predictor of use, namely stakeholder participation in evaluation (e.g., see Cousins, 2003). This reorientation towards stakeholder participation had the additional benefit of allowing an article that the student had already published on this topic during his doctoral studies to be included in the dissertation (Dagneault & Jacob, 2009). To take into account the considerable heterogeneity of the scientific production of the field in terms of research design and constructs definition and measurement, the review approach used was more flexible and less rigid than the Cochrane review (for a similar rationale in the field of management, see Briner, Denyer & Rousseau, 2009; Denyer & Tranfield, 2009), but was nevertheless guided by the systematic standards and procedures as described by Petticrew and Roberts (2006). To keep the project manageable, it was decided that a ‘systematic map’ or ‘scoping study’ (Arksey & O’Malley, 2005; Gough, 2004) of the literature including a methodological critique of the field would be completed first. The relevance and feasibility of conducting a ‘full’ systematic review (i.e., including study quality assessment and results synthesis) for the dissertation would then be assessed.

It is important to describe the research process in sufficient detail to allow for a contextual understanding of this reflective narrative. The tasks of reference search and screening as well as data extraction were conducted individually by the student over a period of approximately two years because other activities (research, consulting and teaching) slowed down the process. The literature search was conducted by combining terms related to evaluation use. The interdisciplinary nature of evaluation warranted searching electronic databases from different academic disciplines. All results were imported into the Endnote, a bibliographic management software. To ensure a comprehensive search that compensate for deficiencies in database indexing, the search process was completed by a hand search of *all* references of relevant journals and of literature reviews on evaluation use (Littell et al., 2008; Sandelowski and Barroso, 2006). Overall, a total of 24,795 references were identified once duplicates were excluded. Since the search process was carried on *before* narrowing the review focus to stakeholder participation, the search terms combinations were clearly less specific than what would be warranted by best practices in systematic reviews. In addition, it must be stressed that the search strategy deliberately emphasized *sensitivity* (i.e., the ability to identify *all* relevant studies) over *specificity* (i.e., the ability to identify *only* relevant references or, in other words, its accuracy) (Petticrew and Roberts, 2006).

The following criteria were used to screen identified references: an *empirical study* (no restriction on design or type of publication) in which the independent variable is *stakeholder participation*, in *program evaluation* and the dependent variable is *evaluation use*, published in English or French between 1970 and 2010. Screening and data extraction were performed individually by the student, but a third party (research assistant or supervisor) checked for a subsample of his work to ensure for reliable and valid data. Because of the low specificity of the

search strategy, a fair amount of 'noise' (i.e., irrelevant references) was observed in the results. Therefore, most references could be relatively easily excluded by one person on the basis of their title and abstract, emphasizing over-inclusiveness. A total of 838 references could not be excluded on the basis of title and abstract. A copy of each available reference was obtained through electronic databases, the local university library, interlibrary loans or by directly contacting the authors of the study. The full-text screening process based on the selection criteria led to the inclusion of 140 references.

Near the end of the data extraction process (only a few references left), it became clear that this review could not be completed on time to activate a postdoctoral fellowship that the student was awarded. Furthermore, the student became aware, through informal contacts, of a review and synthesis of a larger scope on a similar topic that was about to be published at the time (Cousins & Chouinard, 2012). The review process was therefore a race against the clock, but also a race against other scholars. The student thus decided to forgo a systematic review for his dissertation and turn to a validation of a research instrument that purports to measure stakeholder participation in evaluation (Daigneault, Jacob & Tremblay, in press; Daigneault & Jacob, forthcoming).

Issues and challenges

While it is too late for regrets, it is important to look back on this review experience to reflect on the challenges encountered and the lessons that were learned during the process. The following discussion is geared towards the issues that surfaced during the systematic review process and the learning that they spurred on. More specifically, the presentation follows the political (P), conceptual/intellectual (C) and technical (T) challenge categories used by Oakley (2003).

Systematic reviews: Status and acceptance by scholars (P)

Although systematic review methods have been used for quite some time in fields such as medicine, psychology and, more recently, education, most political scientists are only slowly beginning to discover them and appreciate their potential. Regardless of their discipline, students must thus be aware of the fact that everyone on their dissertation committee might not recognize the value of systematic review and synthesis methods. Speaking from the field of education, Oakley (2003) has summarized the main criticisms put forward by what she labelled the 'anti-evidence movement' such as the narrow focus on quantitative effectiveness evidence to the detriment of theory and other kinds of evidence such as professional experience and the limitations to academic freedom (see also Hammersley, 2001; 2008). She argued:

...those engaged in doing systematic reviews in education must often actively contest with their colleagues and others the position that this approach is a suspicious and inappropriately mechanistic procedure for reaching conclusions about what educational research can reliably tell us. Such antagonism and misunderstanding multiplies the difficulties of work in an area that is challenging for other structural and intellectual reasons. (Oakley, 2003, p. 26)

In the present case, the dissertation proposal committee had a relatively open-minded attitude toward systematic reviews as nobody voiced criticisms in line with the anti-evidence movement. That being said, the fact that systematic review methods are relatively novel and/or less familiar to scholars in the field of political science was raised during the proposal as a hurdle to a future academic career. For instance, a member of the committee argued: 'While I am sure that your work will make a fine contribution to the field and that it will be widely cited, I do not think that any political science department will hire you *only* on the basis of a "literature review".' This professor

strongly urged the student to enlarge his dissertation focus to include primary data collection and/or sophisticated theoretical development. The political challenge of systematic reviewing not only relates to the epistemological struggle between the 'pro-evidence' and 'anti-evidence' scholars, but also to the 'so what?' reaction of scholars who are less familiar with this class of methods. These scholars do not see the value added or the amount of work that a systematic review entails. For them, a systematic review is a conventional literature review which is perhaps a bit more comprehensive.

As a result, it is important for students to carefully select, in collaboration with their supervisor, the members of their dissertation committee so that they will already have a basic understanding of systematic reviews. In addition, a section of the dissertation proposal must clearly and convincingly explain the difference between systematic and traditional literature reviews as well as the comparative advantage of the former over the latter.

Review scope and time requirements to process identified references (C & T)

Defining the research question(s) and selection criteria of a systematic review is no trivial matter: 'One reason why specifying review questions and criteria can be such a lengthy, iterative process is because such discussions are precisely not just technical; they provide a forum for debate about important theoretical and ideological issues' (Oakley, 2003, p. 28). Yet, the decisions taken at this stage can have tremendous logistical consequences on the technical tasks of the review process itself, notably on reference screening and data extraction. It is thus important that the scope of the review be sufficiently narrow for the review to be manageable by a student working alone.

Being realistic with respect to their thesis or dissertation scope is a common (and wise) recommendation routinely made to students (e.g., Mauch & Park, 2003). This prescription applies even more forcefully to students conducting a systematic review because, contrary to their colleagues who use more conventional research methods, it can be more difficult for them to narrow their dissertation focus while conducting a review (i.e., they cannot drop a few respondents or cases just because they don't have time to analyse them). Students can thus quickly become overwhelmed by the number of studies to screen, assess and extract. Reviewers should plan about half-a-day/person for extracting data from a study, depending on the complexity of the study and data extraction form (for similar figures, see Oakley, 2003). Whereas selection criteria can be modified during the review process so that they are made more restrictive, a systematic review *must* be systematic, as its name indicates, thus posing distinctive challenges for the student working on an individual basis.

In the present case, the number of studies on the topic was grossly underestimated, as well as the time required for screening them as the relevance of identified references was not always clear. Systematic reviews in the social sciences are indeed often characterized by a low yield of usable studies (Oakley, 2003). A first problem was that the transdisciplinary nature of policy evaluation made the review scope unduly large. In order for the review to take into account all the evidence published on the topic, the literature search had to be conducted in electronic databases and journals from various disciplines (education, psychology, social services, public administration, health, etc.), which generated an impressive number of references. A second problem was that the review focus and inclusion criteria were not specific enough given the fact that there is less consensus on what constitute legitimate research questions and methods in social science compared, for example, to medicine (see Denyer & Tranfield, 2003). For one thing, empirical studies of all types and designs were to be included (even though there is an 'evidence hierarchy'

for the assessment of causal question, the idea was first to 'map' and assess the scientific production of the field). For another, the main variables of interest (stakeholder participation and evaluation use) were defined inclusively in order to avoid missing any potentially relevant reference. This is especially important as concepts are not always neatly defined and measured in many social science disciplines (for management and organizational studies, see Denyer & Tranfield, 2009; Tranfield, Denyer & Smart, 2003). For instance, use was defined to include process and findings use, instrumental, conceptual and symbolic use, and evaluation influence (see Alkin & Taut, 2003).

The last problem was the fact that it was difficult to identify and establish the relevance of many references on the basis of title and abstract alone, a problem encountered by other reviewers in various disciplines (Atkins et al, 2008; Brereton et al., 2007; Jones, 2004; Oakley, 2003), but which was aggravated by the fact that only one person (i.e., the student) performed the screening. Many references have uninformative titles and their abstracts — when they possess one — are rarely as explicit as the detailed abstracts found in medicine and other health sciences. Time-consuming hand searches were therefore required. In addition, the time required for screening references is significantly increased since a copy of the 'grey zone' references (neither clearly relevant nor clearly irrelevant) must be obtained and read in full.

In retrospect, it was clear that the review scope was too wide for being manageable by a single person in a reasonable time. Pre-review or scoping study have been recommended as a 'good practice' to circumscribe the review so that it is both relevant and manageable (Brereton et al., 2007; Denyer & Tranfield, 2009; Jones, 2004). This appears to be good advice, even when, like in the present case, the review is itself a scoping study. In the end, however, the decision to conduct a pre-review must be based on the nature of the reviewed body of literature: 'Where fields comprise of semi-independent and autonomous sub-fields, then this process may prove difficult and the researcher is likely to struggle with the volume of information and the creation of transdisciplinary understanding' (Tranfield, Denyer & Smart, 2003, p. 215). The PICO framework (*Population, Patient or Problem; Intervention, Comparison; Outcome*) and its social science equivalent, the CIMO framework (*Context; Intervention; Mechanism; Outcome*) can be used to restrict the scope of a review before or while it is conducted (Denyer & Tranfield, 2009; Littell, Corcoran & Pillai, 2008; Tranfield, Denyer & Smart, 2003). In the present case, the review could have been restricted to one sector (e.g., health), one population (e.g., public managers) or one type of evaluation use (e.g., instrumental use). Although his dissertation supervisors insisted on the importance of narrowing down the review topic, the student's scientific ambition was a problem; he wanted to review everything that was relevant to the topic and was thus more inclusive than less. In addition, the high level of uncertainty with respect to the number of relevant studies that would be identified contributed to the student's decision to maintain a broad review scope, despite recommendations by his supervisors to be more restrictive.

A second way to restrict the scope of a review is to use less demanding but nevertheless systematic methods of reviewing such as the 'rapid review' or the Rapid Evidence Assessment (REA) (Ganann, Ciliska & Thomas, 2010; Petticrew & Roberts, 2006). Rapid reviews 'aim to be rigorous and explicit in method and thus systematic but make concessions to the breadth or depth of the process by limiting particular aspects of the systematic review process' (Butler et al., 2005, as cited in Grant & Booth, 2009, p. 100). This class of methods is increasingly popular and the scholarly literature now contains many examples in which they have been used when decision-makers need answers in a timely manner, usually a few months (Ganann, Ciliska & Thomas, 2010; Daigneault,

Jacob & Tereraho, 2012). While the production contexts of rapid reviews and PhD dissertations differ substantially, the scope limitations used in rapid reviewing can help a postgraduate student conducting a review without the support of a full review team to keep the project manageable within a reasonable time frame.

What is empirical (and what is not)? (C)

The last challenge to using systematic review methods in a social science dissertation is related to the meaning of 'empirical study'. At the general level, there seems to be a quasi-consensus on what constitutes empirical research. For instance, 'by *empirical knowledge*, I mean experientially based knowledge acquired through formal study (Smith, 1983)' (Christie, 2003, p. 7). Others have defined the adjective 'empirical' in a similar way:

The word empirical means information gained by experience, observation, or experiments. The central theme in scientific method is that all evidence must be empirical which means it is based on evidence. In scientific method the word "empirical" refers to the use of working hypothesis that can be tested using observation and experiment. (Hani, 2009, n.p.)

Yet, the operationalisation of the concept in a clear inclusion criterion is far from straightforward. Should a practice-based reflective case narrative, for example, be considered empirical? It is surely empirical in the sense that is not speculative and is based on direct experience but, at the same time, it does not always make use of or report formal data collection and analysis methods. Perhaps more confusing, even when data collection and analysis methods are explicitly reported in a study, authors may nevertheless qualify their work as a non-empirical study or a non-formal empirical study. For instance, Mertens, Berkeley & Lopez (1995) argue that their '...study provided the basis for an *indirect* examination of the effects of participatory evaluation, not a formal study...' (p. 153, italics added). Forss, Kruse, Taut and Tendén (2006) similarly discussed the reflexive character of their study:

As with many other articles in the evaluation literature, this essay reflects on our practice as evaluators in one specific case [...] A note of warning: this is not a research paper. We call it an essay, and that is what it is – an experiment to develop the understanding of capacity building through our shared reflections on a common evaluation experience. We did consult those who were involved in the process (members of the internal evaluation Task Force – a sort of reference group set up to follow the evaluation) through a set of interviews at UNESCO headquarters, and by an email questionnaire to people in field offices who were focal points for country visits performed as part of the evaluation. But a few interviews and email responses are not sufficient to claim that we have a definitive idea about the actual impact of the evaluation. (p. 130)

These definitional problems have been discussed by others (for an excellent discussion on this topic, see Cousins and Chouinard, 2012) and various legitimate solutions are available to tackle them. In the present case, although the student conducting the review considered reflective case narratives as an empirical and legitimate way of knowing³, he finally chose to restrict his review focus to studies that explicitly reported their methods. This was first and foremost a pragmatic decision intended to keep the number of references manageable. This challenge is especially salient in social science reviews that incorporate qualitative evidence 'since a surprising proportion of studies are very inexplicit, not to say vague, about the designs they have used' (Oakley, 2003, p. 29; see also Atkins et al., 2008; Jones, 2004).

Although the preceding discussion has focused on designs and methods, it is important to stress that this challenge was related to other important constructs of the review such as 'evaluation use' and 'policy evaluation'. Two 'near miss' cases (i.e., cases that seemed relevant *prima facie* but were subsequently excluded) will suffice to illustrate this point. In the first case (Vanlandingham, 2010), the title, abstract and keywords of the reference suggests that the study reported is about evaluation utilization in a legislative evaluation context whereas, in reality, the dependent variable is 'legislative value', an index that contain a few items related to evaluation use (e.g., 'The office's work helps set the legislative issue agenda'), but also to totally different constructs than use (e.g., 'The office is easy to work with') (p. 6). The second case (Oh & Myeong, 2002) is also framed as a study on the use of policy evaluation as its abstract makes clear — 'this study examines the current state of e-government evaluation in Korea and to what uses the evaluation results are put in building an effective e-government' (p. 33). However, the data used in this study is about the use of information in general, including — and thus not limited to — policy evaluation (see also n. 2, p. 40). Whereas evaluation is undoubtedly a type of information, there is nevertheless a fundamental difference between the former and the latter (Daigneault, 2012; Rossi, Lipsey & Freeman, 2004). These two cases illustrate the issues of relevance and construct validity faced by reviewers in social science. Of course, it has been known for a while that scholarly publications are subject to what could be called 'framing effects' by their authors⁴:

Researchers/writers, in turn, employ various writing conventions and literary devices in order to appeal to readers, and to shape and control their readings. [...] Indeed, the research report is itself better viewed, not as an end-stage write-up, but rather as a — dynamic "literary technology" — (Shapin, 1984) whereby writers use literary devices [...] to engage readers rhetorically to accept their study procedures and findings as valid. (Sandelowski & Barroso, 2006, pp. xvii-xviii)

Students can be more vulnerable to these definitional challenges since their methodological and substantive knowledge — as well as their level of confidence in their own judgment — is often not as developed as those possessed by established scholars.

Expertise (T)

When conducting a systematic review, students must be able to count on sufficient substantive, methodological and information technology expertise. Unfortunately, the graduate training currently received by many social science students rarely prepares them for undertaking systematic review work (Denyer & Tranfield, 2009; Oakley, 2003). The student was fortunate enough to have access to policy evaluation expertise through his dissertation supervisor, knowledge use and systematic review expertise through his co-supervisor and information technology expertise through the personnel of his university library. He could also rely on a systematic review consultant (free of charge) through his co-supervisor. This person was of great help with the definition of search terms and subjects as well as reference screening on bibliographic management software such as Endnote. The student could therefore count on a 'team' of researchers that harboured different types of expertise and had different perspectives on the review process. Teamwork can contribute to provide insights and better interpretation of the evidence in cases where students use systematic review methods (see Jones, 2004). In addition to reading books and articles on the subject of systematic reviewing, the student received a financial grant from his supervisor to attend the course 'Introduction to systematic reviewing' at the Essex Summer School in Social Science and Data Analysis. He was thus well-supported and able to conduct the review, which is not the case for every student. Despite the invaluable support described above, the aims and scope of the systematic review were not specific enough to be manageable by only one person. This can be explained by the student's scientific ambition — that

is, he wanted to review *all* empirical studies from *all* disciplines which examined the relationship between stakeholder participation on the one hand and *all* types of evaluation use on the other — but also by other factors. Indeed, the literature search conducted before the reorientation of the review was not only on stakeholder participation, but on all factors affecting evaluation use. Moreover, deficient indexing and the transdisciplinary nature of the field would have made difficult to conduct a more specific search without missing relevant references. As a result, the number of references for which screening and data extraction was needed was grossly underestimated.

Financial resources to complete review tasks and access to documents (T)

The real costs of producing a full systematic review can be quite significant (Oakley, 2003; Petticrew & Roberts, 2006), especially if the number of relevant references is large. It was mentioned earlier that, though systematic, the review process was not as structured as a Cochrane review. Indeed, the student did not have sufficient funds to hire research assistants to perform each and every review task (reference screening, data extraction, etc.) in double. Yet, his supervisor and co-supervisor acted as a third party check at ‘critical junctures’ of the review. They also provided a small budget to hire a research assistant to perform pilot double-coding and double-data extraction on a subsample of the references (70 references and 3 studies, respectively). Taken together, these procedures served to enhance the validity of the review process despite significant resource constraints.

Access to a well-endowed university library is essential to any serious review endeavour. When faced with shortcomings in journals, books and dissertations coverage, students have a few options at their disposal such as requesting that their library make purchase on their behalf and interlibrary loans. However, depending on the number of documents ordered, their geographic origin and the library’s policy on interlibrary loans, costs can quickly become prohibitive for a researcher with a student’s budget. It is always a good idea for the student to plan a small budget with his or her research supervisor to cover such expenses. That being said, a very efficient and convenient way to obtain unavailable references — for students in particular — is to contact authors directly, by email, as was done in the present case. As most study authors are also academics, a quick Google search can often help to locate up-to-date contact information. When contacted through a valid email address, authors do not always reply for various reasons (spam filter, too busy, etc.) and when they do, they are not always able or willing to help you. According to this student’s experience, about 40-45% of the persons contacted replied to emails and, for those who did, about a quarter have sent a copy of their study. So overall, 10% of the needed documents were obtained through direct author contact, which represents a significant economy of resources. Moreover, it is often a good way to make the student and his or her research known to other scholars. In this case, the student ‘thanked’ contacted authors who replied to his query by sending them a copy of one his own publications related to the review topic. This led to a few stimulating (academic) discussions and a few citations of his work. Such fruitful exchanges and peer recognition are particularly important for new scholars.

Systematic review structured format and the scientific race towards progress (T)

As mentioned earlier, the student had to deal with the fact that other scholars had conducted and published the results of systematic reviews on the same topic as him (although there were a few differences). In such situations, the student was at a disadvantage on at least two grounds. First, he was alone for most of the tasks he had to perform whereas his competitors were well-endowed review teams made of established scholars. Second, the consequences of 'losing the race' are usually less severe for established scholars; students, by contrast, depend on the review process to obtain their degree. Contrary to more conventional studies, systematic reviews have an 'all-or-nothing' flavor derived from their methodological standards and practices. Indeed, the structured and even rigid character of systematic reviews, combined with the fact that the data on which they are based is more or less a 'fixed stock' (apart from ongoing and unpublished new studies), entails that the value added of an extra review on the same topic covering the same material is extremely low. By contrast, two different scholars doing independent simultaneous *primary* research on the same research topic (e.g., the 9/11 tragedy) will likely collect different data and analyze them differently. Their studies are thus likely to be significantly different but, even if they are not, the second study could still make a contribution as a 'replication' of the first study.

Lessons learned and recommendations

In the end, many lessons for research practice were learned from the political, conceptual and technical challenges encountered during this systematic review experience. These lessons are mainly directed to other postgraduate students interested in conducting a systematic review for their thesis or dissertation, but can also be relevant to a more general audience (reviewers operating under considerable time and resource constraints, reviewers interested by interdisciplinary topics or different types of empirical evidence etc.). Here are a few (prescriptive) highlights drawn from direct experience:

- *Carefully explain and justify your choice to conduct a systematic review to your supervisor and dissertation committee since you can never take for granted that your supervisor, dissertation committee and future academic employer will be familiar with and/or appreciative of these methods.*
- *Don't be overambitious with respect to the scope of your review: keep it manageable. The PICO or CIMO frameworks, rapid review methods, and conducting a pre/pilot review are useful tools to circumscribe your review aims and scope. It is preferable to be slightly more restrictive than to never complete your review, risk being 'scooped' by colleagues working on the same topic and never obtain your degree.*
- *Don't underestimate the time required for sifting through the identified references and extracting their information, especially if your review topic is interdisciplinary and/or includes various types of evidence, and if, as most students, you perform these tasks individually.*
- *Before you start your review, learn as much as possible about systematic reviews through relevant readings and courses. During your review, be sure to have access to sufficient substantive, methodological and library expertise.*
- *Each selection criterion in your review protocol should be clear and operational. In particular, special care must be taken in defining 'empirical study' when references come from various disciplinary and methodological traditions. To circumvent framing effects, your selection criteria should be independent of the terms used by study authors (to the extent possible).*

- *Don't hesitate to contact authors directly through email to obtain a copy of unavailable or difficult-to-obtain publications. In addition to reducing the costs of your review, direct contact can make yourself and your research known, which is especially important for doctoral students who will soon be on the job market.*
- *In case you don't have sufficient resources to hire personnel to perform reference screening and data extraction in double, use third party checks on a subsample of the references to significantly enhance the validity of your findings.*
- *Since the stakes of not completing a systematic review are high for students (i.e., not obtaining their degree), always have a 'Plan B' (conceptual development, methodological critique, instrument validation, etc.) to which you can turn to relatively easily if a problem occurs with your review.*

Looking ahead

Although the systematic review described in this article was not completed as intended as part of a PhD dissertation, the end of doctoral studies does not necessarily signal the end of the review. Indeed, since a lot of valuable work has been accomplished during the review process, to terminate it would represent a considerable waste of time and resources. Therefore, a version of the original review that builds on the findings of the most recent systematic review on the topic (i.e., Cousins & Chouinard, 2012) and on the lessons learned during the process presented here is currently under preparation by the first author. The topic of this revised review is the same, that is, the relationship between stakeholder participation in evaluation and evaluation use, with the only notable difference that it now focuses exclusively on quantitative evidence.

Perseverance is certainly a virtue for those involved in conducting systematic reviews, but students need to remain open-minded and flexible about the aims and scope of their review in the face of developments. While a certain level of ambition is essential to completing a thesis or dissertation, being overambitious is certainly the best way to never obtain one's degree. Indeed, PhD students should keep in mind that 'the best dissertation is a done dissertation' and, by extension, that 'the best systematic review is a done review'.

Notes

1. Systematic reviews are generally conceived as involving many steps, including study quality appraisal and data integration or synthesis. As used in this article, however, systematic review simply refers to an "overview of primary studies that use explicit and reproducible methods" (Greenhalgh, 1997, as cited in Petticrew & Roberts, 1997, p. 283). This broad definition of systematic review thus includes scoping studies, systematic maps and rapid reviews in the extension of the concept. We indeed argue that reviews which use explicit and reproducible methods to search, screen and analyze the literature are fundamentally closer to 'full' systematic reviews (i.e., involving study quality appraisal and synthesis) than they are to conventional literature reviews. Furthermore, systematic review methods share common roots (i.e., the evidence movement) and their various applications such as scoping studies or rapid reviews are usually included within a more general discussion on systematic reviews (see e.g., Briner, Denyer & Rousseau, 2009; Gough, 2004).
2. Other organizations involved in the evidence movement have adopted a more open-minded perspective on what constitute sound empirical evidence. For instance, reviews

- conducted under the hospices of the EPPI-Centre include a variety of methodological designs (Hansen & Rieper, 2009).
3. It is not the value of reflective narrative for professional learning that is disputed — to do so would naturally be self-defeating as this article precisely reports a reflective narrative — but rather its ontological status as empirical study.
 4. To be sure, *all* scholarly publications — including this article — make use of framing devices. The point that is made here is that practice poses a particular challenge for systematic reviewers who have to assess reference relevance.

References

- Alkin, M.C., & Taut, S.M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1-12. Retrieved from: [http://dx.doi.org/10.1016/S0191-491X\(03\)90001-0](http://dx.doi.org/10.1016/S0191-491X(03)90001-0)
- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8, 19-32. Doi: 10.1080/1364557032000119616
- Armitage, A., & Keeble-Allen, D. (2008). Undertaking a structured literature review or structuring a literature review: tales from the field. *The Electronic Journal of Business Research Methods*, 6, 103-114. Retrieved from: www.ejbrm.com
- Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: Lessons learnt. *BMC Medical Research Methodology*, 8(1), 21.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4), 571-583. doi: 10.1016/j.jss.2006.07.009
- Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-Based Management: Concept Cleanup Time? *Academy of Management Perspectives*, 23(4), 19-32. doi: 10.5465/amp.2009.45590138
- Bullough, R. V., & Pinnegar, S. (2001). Guidelines for Quality in Autobiographical Forms of Self-Study Research. *Educational Researcher*, 30(3), 13-21. doi: 10.3102/0013189x030003013
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, (97), 7-35. doi: 10.1002/ev.72
- Christie, C.A. (2007). Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation*, 28, 8-25. 10.1111/1467-9302.00144
- Cousins, J.B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan & D.L. Stufflebeam (Eds.) *International handbook of educational evaluation* (Vol. 9, pp. 245-265). Dordrecht: Kluwer Academic.
- Cousins, J. B. (2005). Will the real empowerment evaluation please stand up? A critical friend perspective. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment Evaluation in Practice* (pp. 183-208). New York, NY: Guilford Press.
- Cousins, J.B., & Chouinard, J.A. (2012). *Participatory evaluation up close: A review and integration of research-based knowledge*. Charlotte, NC: Information Age Press.
- Daigneault, P.-M. (2012). La participation à l'évaluation: du concept à la mesure [Stakeholder participation in evaluation: from the concept to measurement]. Unpublished PhD Dissertation, Université Laval, Quebec City.

- Daigneault, P.-M., & Jacob, S. (forthcoming). Unexpected but most welcome: Mixed methods for the validation and revision of the Participatory Evaluation Measurement Instrument. *Journal of Mixed Methods Research*.
- Daigneault, P.-M., & Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation, 30*(3), 330-348.
- Daigneault, P.-M., Jacob, S., & Tereraho, M. (2012). Understanding and improving the take-up of public programs: Lessons learned from the Canadian and international experience in human services. *International Journal of Business and Social Science, 30*(3), 39-50.
- Daigneault, P.-M., Jacob, S., & Tremblay, J. (in press). Measuring stakeholder participation in evaluation: An empirical validation of the Participatory Evaluation Measurement Instrument (PEMI). *Evaluation Review*.
- Davies, H.T.O., Nutley, S.M., & Smith, P.C. (1999). Editorial: What works? The role of evidence in public sector policy and practice. *Public Money and Management, 19*, 3-5. doi: 10.1111/1467-9302.00144
- Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. A. Buchanan & A. Bryman (Eds.), *The Sage handbook of organizational research methods* (pp. 671-689). London: Sage.
- Forss, K., Kruse, S.-E., Taut, S., & Tenden, E. (2006). Chasing a ghost? An essay on participatory evaluation and capacity development. *Evaluation, 12*, 128-144. doi: 10.1177/1356389006064190
- Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science, 5*(1), 56.
- Gervais, M. (2012). Guest editor's remarks: As I recall—Or how to take advantage of less-than-successful evaluation experiences. *Canadian Journal of Program Evaluation, 25*(3 (special issue 2010)), xi-xvii.
- Gough, D. (2004). Systematic research synthesis. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 44-62). Buckingham: Open University Press.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108. doi: 10.1111/j.1471-1842.2009.00848.x
- Hammersley, M. (2001). On 'Systematic' Reviews of Research Literatures: A 'narrative' response to Evans & Benefield. *British Educational Research Journal, 27*(5), 543-554. doi: 10.1080/01411920120095726
- Hammersley, M. (2008). Paradigm war revived? On the diagnosis of resistance to randomized controlled trials and systematic review in education. *International Journal of Research & Method in Education, 31*(1), 3-10. doi: 10.1080/17437270801919826
- Hani, D. (2009). Empirical research. *Experiment resources*, n.p., n.e.. Retrieved from: www.experiment-resources.com/empirical-research.html
- Hansen, H.F., & Rieper, O. (2009). The evidence movement: The development and consequences of methodologies in review practices. *Evaluation, 15*, 141-163. doi: 10.1177/1356389008101968
- Heinrich, C.J. (2007). Evidence-based policy and performance management. *The American Review of Public Administration, 37*, 255-277. doi: 10.1177/0275074007301957
- Henry, G.T., & Mark, M.M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*, 293-314. doi: 10.1177/109821400302400409

- Jacob, S. (2008). Cross-Disciplinarization: A New Talisman for Evaluation? *American Journal of Evaluation*, 29(2), 175-194. doi: 10.1177/1098214008316655
- Johnson, K., Greenesid, L.O., Toal, S.A., King, J. A., Lawrenz, F., & Volkov, B (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30, 377-410. doi: 10.1177/1098214009341660
- Jones, M. L. (2004). Application of systematic review methods to qualitative research: practical issues. *Journal of Advanced Nursing*, 48(3), 271-278. doi: 10.1111/j.1365-2648.2004.03196.x
- Learmonth, M., & Harding, N. (2006). Evidence-based management: The very idea. *Public Administration*, 84, 245-266. doi: 10.1111/j.1467-9299.2006.00001.x
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford: Oxford University Press.
- Mauch, J.E., & Park, N. (2003). *Guide to the successful thesis and dissertation: A handbook for students and faculty* (5th ed.), New York: Marcel Dekker.
- Mertens, D.M., Berkeley, T.R. & Lopez, S.D. (1995). Using participatory evaluation in an international context. In J. B. Cousins & L.M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 140-156). London/Washington D.C.: The Falmer Press.
- Minnie, K., van der Walt, C., Klopper, H. & Cummings, G. (2010, July). *Systematic or integrative review as research project for graduate study*. Paper presented at the International Nursing Research Congress, Orlando, FL. Retrieved from http://www.stti.iupui.edu/pp07/congress10/minnie_k.pdf
- Oakley, A. (2003). Research Evidence, Knowledge Management and Educational Practice: Early lessons from a systematic approach. *London Review of Education*, 1(1), 21-33. doi: 10.1080/14748460306693
- Oh, C. H., & Myeong, S.-H. (2002). Evaluating e-government: a lesson from Korea. *International Review of Public Administration*, 7(2), 33-44.
- Owens, E, Baez, M., & Tillman, S. (2006). Lessons learned: The student experience. *Contemporary issues in communication science and disorders*, 33, 74-78. doi: 1092-5171/06/3301-0074
- Pawson, R. (2002). Evidence-based policy: In search of a method. *Evaluation*, 8, 157-181. doi: 10.1177/1358902002008002512
- Perry, A., & Hammond, N. (2002). Systematic reviews: The experiences of a PhD student. *Psychology Learning and Teaching*, 2, 32-35. Retrieved from: <http://dx.doi.org/10.2304/plat.2002.2.1.32>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Sambunjak, D., & Puljak, L. (2010). Cochrane systematic review as a PhD thesis: An alternative with numerous advantages. *Biochemia Medica*, 20, 319-326. Retrieved from: <http://www.biochemia-medica.com/content/cochrane-systematic-review-phd-thesis-alternative-numerous-advantages>
- Sandelowski, M., & Barroso, J. (2006). *Handbook for synthesizing qualitative research*. New York, NY: Springer.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80, 1-22. doi: 10.1111/1467-9299.00292
- Scriven, M. (2004). Reflections. In M.C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 183-195). Thousand Oaks, CA: Sage.

- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207-222. doi: 10.1111/1467-8551.00375
- Vanlandingham, G. R. (2010). Escaping the Dusty Shelf: Legislative Evaluation Offices' Efforts to Promote Utilization. *American Journal of Evaluation*. doi: 10.1177/1098214010382768