Journal of the Association for Information Systems (2020) **21**(5), 1103-1114
doi: 10.17705/1jais.00630

EDITORIAL

ISSN 1536-9323

(Re)considering the Concept of Literature Review Reproducibility

W. Alec Cram[1], Mathieu Templier[2], Guy Paré[3]
[1]University of Waterloo, Canada, wacram@uwaterloo.ca
[2]Université Laval, Canada, mathieu.templier@fsa.ulaval.ca
[3]HEC Montréal, Canada, guy.pare@hec.ca

Abstract

Literature reviews play a key role in academic research by describing, understanding, explaining, and testing the constructs and theories within a particular topic area. In recent years, various commentaries, debates, and editorials in the information systems (IS) field's top journals have highlighted the importance of a trustworthy literature review process, including detailed discussions on systematicity and transparency. Although the reproducibility of a literature review has also been noted as important, it remains less recognized because of several terminology-related issues. This ambiguity could result in misunderstandings regarding the degree of trust that should be placed in a literature review's process. In this research essay, we seek to clarify what makes a literature review reproducible, how it is distinct from related concepts, and when achieving it is desirable and feasible. We propose a series of clarifications and remedies to assist scholars within and outside the IS field in the preparation of stand-alone reviews.

Keywords: Literature review, Reproducibility, Repeatability, Replicability, Systematicity, Transparency

Dorothy E. Leidner was the accepting senior editor. This editorial was submitted on April 7, 2020.

1 Introduction

Literature reviews act as the backbone of an academic field. By periodically examining the research that has been done in the past, scholars can cultivate valuable insights by describing a phenomenon (e.g., a narrative review), understanding a phenomenon (e.g., a scoping review), explaining a phenomenon (e.g., a realist review), or testing a theory (e.g., a meta-analysis) (Paré et al., 2015; Rowe, 2014; Templier & Paré, 2018). In contrast to the conceptual foundations or background sections within traditional empirical papers, literature reviews in this context represent full-length, stand-alone academic studies. Taken as a whole, IS scholars have been provided with a range of opinions and suggestions on how to conduct literature reviews. Although the advice has been contradictory at times (for example, refer to the discussion below on differing opinions related to systematicity), most would acknowledge the overall value that has emerged from this dialogue. Indeed, much of the recent guidance has been oriented around the characteristics that make a literature review process trustworthy. In particular, a good deal of recent attention has been paid to the role of *systematicity*, which refers to literature reviews that are conducted in an organized and orderly manner (Paré et al., 2016), as well as *transparency*, which is achieved when the elements of the review process are explicitly detailed (Templier & Paré, 2018). A variety of views on either systematicity, transparency, or both have figured prominently in the recent IS literature review guidance of Boell & Cecez-Kecmanovic (2015a, 2015b), Paré et

al. (2016), Rowe (2014), Templier and Paré (2018), and vom Brocke et al. (2015).

A third characteristic, *reproducibility*, broadly refers to elements of the review process that can be duplicated by an independent party. Although reproducibility has been noted as important (Leidner, 2018; Paré et al., 2016; Templier & Paré, 2018), we believe that it remains less recognized in the IS field because of three main issues. First, there are competing perspectives on the precise definition of reproducibility, including inconsistencies relative to other scientific disciplines. Specifically, most IS research does not clearly distinguish between reproducibility that refers to an independent party who is able to reperform the *methods* used in a literature review (e.g., the literature search) and reproducibility that refers to an independent party who is able to duplicate the *results* of a literature review using an existing dataset (e.g., recalculate effect sizes in a meta-analysis). This ambiguity in the terminology could cause a misunderstanding that leads to a false sense of trust in the paper (i.e., the authors claim that their work is reproducible in a methodological sense, but a reviewer/editor might interpret this as referring to the reproducibility of the study's results). Likewise, reproducibility misunderstandings could lead to underestimating the trustworthiness of a paper's approach (i.e., the authors claim their work is reproducible in a results sense, but a reviewer/editor interprets this as referring to methodological reproducibility).

The second issue refers to the uncertainty regarding how reproducibility is distinct from the seemingly similar concept of replicability, which broadly refers to a completely new study that seeks to corroborate or refute a previous study's results based on independent data collection and analysis. Since both reproducibility and replicability have elements of duplication by an independent party, the difference between the two concepts is unclear. If authors, reviewers, and editors are uncertain about how reproducibility is distinct from replicability, it becomes increasingly challenging for authors to satisfy their desired objective. That is, if there is an ambiguous relationship between the concepts, authors may falsely claim that their research is replicable, when in fact it is reproducible.

Third, there are unanswered questions related to the desirability and feasibility of achieving reproducibility in literature reviews. The differing perspectives voiced by recent commentators on the value of reproducibility introduce uncertainty into the literature review writing process. Specifically, it remains unclear what elements of a literature review could (or should) be reproducible, as well as the types of literature reviews that reproducibility applies to. On the one hand, this uncertainty can lead to wasted time for authors who attempt to pursue reproducibility if it isn't actually valued by reviewers and editors; on the other hand, authors may erroneously determine that reproducibility is unimportant and unintentionally contribute to grounds for a rejected manuscript.

The objective of this essay is to bring clarity to the terminology associated with reproducibility in the context of literature reviews by identifying a series of potential solutions to the three issues outlined above. Ultimately, we suggest that IS scholars employ the terminology in a manner that is consistent with other scientific fields, by using the term *repeatability* to refer to the methods that can be reperformed by an independent party in order to generate trust in the methodological process used by the authors. Further, we suggest that the term *reproducibility* be used to denote an independent party that is able to duplicate a review's results using an existing dataset in order to generate trust in the data analysis process used by the authors. Further, we emphasize the important difference between *reproducibility* and *replicability*. Here, we suggest that replicability is distinct in that it comprises a completely new, second study (compared to reproducibility, which focuses on only a single study), in which the objective is to confirm or refute the results of an earlier study (compared to reproducibility, which aims to generate trust in the process used by the authors). In clarifying these three key terms, we also distinguish between the literature review steps and types that could benefit from being repeatable, reproducible, and/or replicable. In particular, we believe that theory-testing reviews (e.g., meta-analysis, qualitative systematic reviews, and umbrella reviews) enjoy an improved level of trustworthiness by fulfilling one or more of these characteristics, while other review types, such as theory development reviews, do not.

The remainder of the paper is presented as follows. First, we present an overview of the main characteristics of IS literature reviews. This is followed by a discussion of the current issues related to reproducibility, their potential impact, and proposals for addressing concerns. We conclude with considerations for future directions that can further aid the field in facilitating literature reviews that are increasingly reproducible.

2 Characteristics of Literature Reviews

A variety of influential commentaries have been published over the years to provide assistance to IS scholars seeking to publish literature reviews. Early guidance in the field is often traced to the creation of the *MIS Quarterly* review department in 2001 and the advice of the first two senior editors of the section (Watson, 2001; Webster & Watson, 2002). During the past five years, the editors of other top IS journals have also weighed in with their views on literature reviews within IS, including at the *European Journal of Information Systems* (Rowe, 2014) and the *Journal of*

the Association for Information Systems (Leidner, 2018). In addition, there have been a number of debates and commentaries published in top IS journals that examined emerging issues related to the practice of writing literature reviews. For example, in 2015, the *Communications of the Association for Information Systems* published a special issue on IS literature reviews, oriented towards advancing review methodologies, improving rigor, and providing practical guidance to authors (Tate et al., 2015). Similarly, a paper published in the *Journal of Information Technology* by Boell and Cecez-Kecmanovic (2015b) considered the concept of being systematic in literature reviews, which prompted a range of responses (Chiasson, 2015; Oates, 2015; Schultze, 2015; Watson, 2015), as well as a rejoinder from the original authors (Boell & Cecez-Kecmanovic, 2015a). Finally, in addition to these important collections of opinions, several additional publications have provided valuable and insightful views on IS literature reviews, including Paré et al. (2016; 2015), Schwarz et al. (2007), Sylvester et al. (2013), Templier & Paré (2018), and Wolfswinkel et al. (2013).

As this guidance has expressed, undertaking a stand-alone literature review can follow several distinct paths, each of which has the potential to result in a trustworthy process. Past commentators have established a variety of frameworks and typologies that categorize these different approaches. For example, Rowe (2014) classifies literature reviews into four genres with respect to theory: describing, understanding, theory testing, and explaining. In comparison, Paré et al. (2015) propose a typology of nine ideal literature review types, consisting of narrative review, descriptive review, scoping/mapping review, meta-analysis, qualitative systematic review, umbrella review, theoretical review, realist review, and critical review. These types are distinguished by seven first-order dimensions: overarching goal, scope of questions, search strategy, nature of primary sources, explicit study selection, quality appraisal, and methods for synthesizing/analyzing findings. In addition, Leidner (2018), who focuses more specifically on theoretical reviews, suggests that such review papers can be framed as either organizing reviews, assessing reviews, specific theorizing reviews, or broad theorizing reviews, depending on the review's focus and objectives.

2.1 What Makes a Stand-Alone Literature Review Trustworthy?

Among the range of characteristics associated with a trustworthy IS literature review process, systematicity and transparency are the most commonly recognized within the recent commentaries, while reproducibility is only occasionally referred to.[1] Systematic and transparent literature reviews are important because they enhance credibility and provide helpful guidance to researchers for future studies (Paré et al., 2016). In order to achieve systematicity and transparency, literature reviews not only require an organized design and sound execution, but also a clear explanation of the methods used. In that sense, systematicity and transparency represent key characteristics of all review types because they impart confidence in the review outcomes and provide inspiration for further research in terms of the methods and techniques used. Refer to Table 1 for examples of how the characteristics have been referred to within IS, as well as outside of the discipline.

2.1.1 Systematicity

The first characteristic is systematicity refers to "a disposition towards organized, methodic, and orderly inquiry that uses various methods and processes to search, screen, assess, analyze and interpret relevant information with a view to achieving a set of specific research goals" (Paré et al., 2016, p. 596; cf. Valanides & Charoula, 2008; Borko et al., 2007)[2]. A common technique to demonstrate systematicity is to follow a series of predetermined, agreed-upon steps. One example in the recent guidance is Okoli's (2015) eight steps of identifying the purpose, drafting a protocol and training a team, applying a practical screen, searching for literature, extracting data, appraising quality, synthesizing the studies, and writing the review. Another example is from Fink's (2010) seven steps of selecting a research question, selecting sources, choosing search terms, applying practical screening criteria, applying methodological screening criteria, doing the review, and synthesizing the results. Although systematicity in conducting IS literature reviews is generally seen as a valuable objective, some commentators speak to the difficulties in actually being entirely systematic.

[1] We recognize that other characteristics of trustworthy literature review processes are occasionally noted in the literature, such as objectivity and comprehensiveness (refer to the Future Considerations section below for additional research opportunities); however, we focus on those that have been most prominently discussed in the recent IS commentaries.

[2] As pointed out by Paré et al. (2016), Schultze (2015), and vom Brocke et al. (2015), the characteristic of being systematic differs from the genre of a systematic literature review, which refers to a specific method common in other fields, such as medicine. Therefore, all literature reviews are at least somewhat systematic, even though all reviews are not "systematic literature reviews."

Table 1. Characteristics of Trustworthy Literature Reviews

Characteristic	Definition(s)	Reference(s)
Systematicity	"a disposition towards organized, methodic, and orderly inquiry that uses various methods and processes to search, screen, assess, analyze and interpret relevant information with a view to achieving a set of specific research goals." (p. 596)	Paré et al. (2016); cf. Valanides & Charoula (2008) and Borko et al. (2007)
Transparency	"the extent to which the review process is made explicit." (p. 504) "the completeness with which a review is presented and whether important methodological aspects about its design and execution are clearly or explicitly reported." (p. 497)	Templier & Paré (2018) Paré et al. (2016); cf. Shea et al. (2009) and Liberati et al. (2009)
Reproducibility	The methods used in a literature review could be reproduced by an independent party. An existing study where the data are made available to others and reanalyzed by an independent party in order to duplicate the results.	Paré et al. (2016); Templier & Paré (2018) Bollen et al. (2015); Cassey & Blackburn (2006); Goodman et al. (2016); Peng (2011).

For example, vom Brocke et al. (2015) suggest that there is no one-size-fits-all approach to the literature search, that literature searches often lead to unexpected results, and that it is not easy to tell when a literature search is finished. Toward that end, the activities required to conduct a literature review have at least some degree of variability, which may be seen as conflicting with the concept of completely organized and orderly inquires (Boell & Cecez-Kecmanovic, 2015b; Leidner, 2018).

Some commentators, such as Paré et al. (2016), argue that systematicity can be embedded within each step (i.e., developing a plan, searching the literature, selecting studies, assessing the quality of studies, extracting data, and analyzing/interpreting/ synthesizing/formulating conclusions) of all review types. Others advocate that systematicity should be of varying importance depending on the review step or type of literature review. For example, Rowe (2014) argues that

> *systematicity, like perfect coverage, may not always be the most important quality elements of a literature review. In fact, higher systematicity does not help much "abstracting data" from papers and synthesizing it. Systematicity is more and more important for the assessment of the material in the collecting stages and to some extent for the "doing the review" stage, but it is more important for explaining and testing reviews rather than for understanding and viewing the landscape.* (p. 247)

2.1.2 Transparency

A second characteristic is that of *transparency*, which refers to "the extent to which the review process is made explicit" (Templier & Paré, 2018, p. 504). Similarly, Paré et al. (2016) (cf. Shea et al., 2009; Liberati et al., 2009) define transparency as "the completeness with which a review is presented and whether important methodological aspects about its design and execution are clearly or explicitly reported" (p. 497). From this perspective, trustworthiness is not only determined by completing the necessary literature review steps in an orderly way (i.e., systematicity), but by also clearly describing the literature review steps to the reader. Such details could include the activities undertaken to complete a thorough database search or assess the quality of the literature. Being transparent can allow the strengths and weaknesses of a study to be evaluated (Liberati et al., 2009; Rowe, 2014).

2.1.3 Reproducibility

Finally, a third characteristic that is occasionally recognized as part of the discussion on literature review trustworthiness is *reproducibility*. Paré et al. (2016) argue that reproducibility contributes to the credibility of a literature review by clarifying the reasonability of the research design. However, despite this recognition of importance, there are conflicting views as to what reproducibility actually means. From one perspective, reproducibility is closely connected with transparency in the sense that if a review is transparent (i.e., the review steps are explained), the authors' methodological steps, such as the literature search, could be "reproduced" in that they could be reperformed by an independent party (Paré et al., 2016; Templier & Paré, 2018). For example, Paré et al. (2015) note that "reliability describes the reproducibility of the review process, which may be facilitated by a comprehensive documentation of the literature search process, extraction, coding and analysis performed in the review" (p. 192). It is important to note that this approach to reproducibility orients itself around the reproducibility of the *methods*

only and not the *results* (Templier & Paré, 2018), on the basis that literature reviews "are a human-based activity, and the literature pool changes over time" (Paré et al., 2016, p. 497).

However, other perspectives, particularly those from fields outside of IS, such as the natural sciences and medicine, focus not on the reproduction of methods, but instead on the reproduction of analysis and results. From this perspective, reproducibility is achieved when data from an existing study are made available to an independent party for reanalysis in order to duplicate the study's results (Bollen et al., 2015; Cassey & Blackburn, 2006; Goodman et al., 2016; Peng, 2011). [3] This "reproducibility of results" approach diverges significantly from the "reproducibility of methods" approach advocated by Paré et al. (2016) and Templier and Paré (2018). This difference is further articulated by Boell & Cecez-Kecmanovic (2015a), who suggest that the rigor of IS literature reviews is, to a large part, assessed on the trustworthiness of the document search, while other fields are more concerned with how comprehensive the literature is and the methodology of the selected studies, but "not how they are identified or whether the process of identifying them is reproducible by others" (p. 163).

A second point of confusion regarding the notion of reproducibility is its ambiguous association with replicability. Since both concepts share elements associated with the duplication of process steps by an independent party, it remains unclear how the two concepts are distinct from one another. Although references to replicability exist in the recent guidance to authors of IS literature reviews, they predominantly relate to trustworthiness characteristics other than reproducibility (e.g., transparency). For example, Bandara et al. (2015) state that "the value of IS literature reviews and, indeed, literature reviews in any field can thus be significantly enhanced through greater accuracy and comprehensiveness in the review process and through better justification and legitimization of choices. The review becomes not only more useful to the field but also more replicable and transparent" (p. 155). Similarly, Paré et al. (2015) note that "the quality of a review is … reflected by the thoroughness of the documentation of the search and synthesis process, and the soundness in the choice of the approach used. At any point in time, a researcher interested in replicating a review should have all the information needed to complete the process" (p. 192).

A final point of contention is the division of opinions on the desirability and feasibility of achieving reproducibility. For example, Leidner (2018) suggests that writing a theoretical literature review is an iterative process, which may not be compatible with reproducibility. She notes that "I have seen authors obsess over reproducibility and have enjoyed some lively debates about creativity versus reproducibility. The process of conducting a review, to me, is as much an art as a science" (p. 562). Similarly, Boell & Cecez-Kecmanovic (2015a) argue that at least some forms of literature reviews contain too many subjective decisions (e.g., determining quality criteria) that can't be replicated or even adequately explained. They argue that even if a description of the tasks can be articulated (e.g., the quality of the collected manuscripts that were assessed), it is much more difficult to ensure that those tasks can actually be performed by others. This distinction seems to be somewhat consistent with past conceptualizations of the differences between explicit versus tacit knowledge (Nonaka, 1994).

In general, the concern raised with this argument is that the act of recording the steps required to make a literature review reproducible may be disruptive and distracting to the creative process (Boell & Cecez-Kecmanovic, 2015b; Leidner, 2018). For example, Chiasson (2015) notes that excessive focus on the mechanics of literature review methods "is a warning to pay attention to the increasing use of methodological checklists in IS research arising out of wholesale methodological absorption" (p. 175) and that such checklists can restrict "the ability of the authors to pursue alternative means and ends" (p. 175). From a feasibility perspective, the concept of reproducibility can also be questioned in terms of the continually changing pool of literature that is available for review. When reproducibility is used in this context, concerns have been raised in the recent commentaries regarding potential challenges in reperforming database searches (Boell & Cecez-Kecmanovic, 2015b).

2.2 (Re)considering the Concept of Reproducibility

As described above, the IS discipline currently faces three broad issues associated with literature review reproducibility: (1) competing perspectives on what reproducibility means, (2) ambiguity in distinguishing reproducibility from replicability, and (3) questions related to the desirability and feasibility of

[3] There remains some ambiguity in terms of whether the reproducibility of results is achieved through the *potential* of being reproducible (e.g., the data are made available and the results could be recreated, if desired) or as a consequence of the *actual* reproduction of the results (e.g., the data are made available, the analysis is reperformed, and the findings are confirmed to be accurate). In order to remain consistent with the reproducibility of methods definition (where an independent party could reperform the methodological steps but are not actually required to do so—see Table 1), we acknowledge that the reproducibility of results could be satisfied with the achievement of *potential reproducibility*.

reproducibility. As shown in Table 2, several avenues may be helpful in moving towards a resolution to the reproducibility concerns noted above. The first step is to clarify the definition of reproducibility. We propose adjusting the terminology pertaining to reproducibility to make it increasingly consistent with other scientific fields. Specifically, we suggest that the term *repeatability* be used to refer to an independent party that is able to reperform the methodological steps of an existing literature review, with the aim of generating trust in the methodological process used by the authors. The concept of repeatability is established in the scientific literature and can be achieved in cases where "from the information presented, a third party must be able to perform a study using identical methodological protocols and analyze the resulting data in an identical manner" (Cassey & Blackburn, 2006, p. 958). This use of the term repeatability would replace the currently ambiguous concept of methods reproducibility that is currently employed within the IS literature. By taking this approach, the term *reproducibility* could then be applied within IS in a manner consistent with other scientific fields, as the situation where existing data are made available to others and potentially reanalyzed by an independent party in order to duplicate the results, with the objective of generating trust in the data analysis process used by the authors (Bollen et al., 2015; Cassey & Blackburn, 2006; Goodman et al., 2016; Peng, 2011).

Table 2. Current Issues on Reproducibility in IS Literature Reviews

Current issue	Examples	Potential impact	Proposal
Competing perspectives on what reproducibility means	Paré et al. (2016) and Templier and Paré (2018) view reproducibility in terms of methods versus other scientific fields that view reproducibility in terms of findings, such as Bollen et al. (2015), Goodman et al. (2016), and Peng (2011).	An inconsistent reproducibility definition could lead to misunderstandings on the literature review methodology, leading to either a reviewer's/editor's false sense of trust in the paper or an underestimation of trustworthiness.	Modify IS terminology to be consistent with other fields. For example, adapt the "reproducibility of methods" terminology used by Paré et al. (2016) and Templier and Paré (2018) to *repeatability.* Cassey and Blackburn (2006) explain that repeatability is achieved when, "from the information presented, a third party must be able to perform a study using identical methodological protocols and analyze the resulting data in an identical manner" (p. 958).
Ambiguous application of the term reproducibility, compared to replicability	Reproducibility is not always clearly differentiated from replicability (Bandara et al., 2015; Paré et al., 2015; Templier & Paré, 2018).	Ambiguity on reproducibility as a distinct concept may lead authors to make incorrect trustworthiness claims.	Researchers should be increasingly mindful of the distinction between *reproducibility* (where an independent party duplicates the results of an existing study using the original dataset, in order to gain trust in the data analysis process used by the authors) and *replicability* (a completely new study that follows the methodological and analysis approach of a previous study, but collects its own data and aims to corroborate or refute the results of the earlier study) (Dennis & Valacich, 2014; Peng, 2011).
Disagreements on the desirability and feasibility of reproducibility	Competing opinions on when reproducibility is desirable and/or feasible (Boell & Cecez-Kecmanovic, 2015b; Chiasson, 2015; Leidner, 2018).	Without some consensus on the desirability and feasibility of reproducibility, authors may waste time pursuing a concept not valued by editors and reviewers. Alternatively, authors may conclude reproducibility is unimportant even though editors/reviewers desire it, contributing to grounds for manuscript rejections.	Highlight the cases in which reproducibility is actually valuable (e.g., when results can actually be reproduced and creativity is not impeded). For those that don't meet the criteria, reviewers and editors could focus on other trustworthiness criteria instead.

This proposal achieves two objectives. First, it recognizes the valuable observation made by Paré and colleagues (2016) related to the importance of being able to independently complete the methodological steps that the authors of a review have conducted (i.e., repeatability). Secondly, it distinguishes this activity from the concept of reproducibility, in terms of findings that could be duplicated by an independent party, based on existing data (e.g., Bollen et al., 2015; Cassey & Blackburn, 2006; Goodman et al., 2016; Peng, 2011). By adopting this proposal, prospective authors can be increasingly clear in specifying the trustworthiness of their work, while maintaining consistent terminology with scholars in other fields.

The second area of concern is to clarify how the concept of reproducibility is distinct from replicability. In general, we suggest that in order for a literature review to be reproducible, it must satisfy two conditions. First, the review must utilize an analysis approach that is objective (rather than subjective; see Boell & Cecez-Kecmanovic, 2015a), including the existence of observable and measurable data. Second, the independent party performing the reproduction must have the necessary research capabilities (e.g., analysis skills, tools) to derive the same results from the existing dataset. Without these contributing elements, results from the existing study could not be consistently reproduced. However, we note the importance of distinguishing reproducibility from replicability, which also has an established definition in other fields (e.g., natural sciences, medicine). Specifically, rather than using the existing data (as is done with reproducibility), replicability infers that a separate, stand-alone study would independently collect a similar dataset and then analyze the data in order to verify if the initial study's results can be duplicated (Peng 2011).[4] That being said, we believe that both repeatability and reproducibility are necessary (but insufficient) conditions for replicability. Here, not only does a third party require a suitable analysis approach and capabilities to transform a dataset into the same results (reproducibility), but they also require the capabilities to undertake an entirely separate study using the same methodological steps as the original authors, in order to duplicate the original data collection, the corresponding dataset, the analysis, and the results.

The third and final issue is concerned with establishing when reproducibility (and the related concepts of repeatability and replicability) can add value to the trustworthiness of a literature review process and when it cannot. For instance, in what scenarios might the achievement of reproducibility be valuable versus impractical? In order to address this question, we considered past work by Paré et al. (2015; 2016) in terms of the different steps undertaken when conducting a literature review (Table 3) and the different types of literature reviews (Table 4). When considering the six generic review steps that comprise a literature review, the elements that would appear to pertain to our proposed application of the repeatability definition include the core methodological steps: developing a review plan, searching the literature, selecting studies, assessing the quality of studies, and extracting data or key aspects from the included studies. It would also include the initial analysis of the data, but would not go so far as to require the repeatability of findings or conclusions (see definition in Table 2).

Table 3. Repeatability, Reproducibility, and Replicability in Literature Review Steps

Literature review step	Repeatability	Reproducibility	Replicability
1. Developing a review plan	X		X
2. Searching the literature	X		X
3. Selecting studies	X		X
4. Assessing the quality of studies	X		X
5. Extracting data or key aspects from included studies	X		X
6a. Analyzing data	X	X	X
6b. Interpreting and/or synthesizing data, and formulating conclusions		X	X

[4] We recognize that Peng's (2011) definition of replicability is consistent with the concept of exact replications proposed by Dennis & Valacich (2014). Although two other forms of replications are also outlined by Dennis & Valacich (methodological and conceptual), we restrict our focus in this essay to exact replications only.

Table 4. Repeatability, Reproducibility, and Replicability in Literature Review Types

IS Literature review type	Repeatability	Reproducibility	Replicability
Narrative review			
Descriptive review	X	X	
Scoping review	X	X	
Critical review	X		
Meta-analysis	X	X	X
Qualitative systematic review	X	X	X
Umbrella review	X	X	X
Theory development review			
Realist review	X		

In comparison, under the proposed definition of reproducibility, the only step that would be relevant is the one related to analysis, interpretation, and/or data synthesis, including the formulation of conclusions, as reproducibility is only concerned with the duplication of results based on provided data. Finally, from the perspective of replicability, all six steps would need to be sufficiently articulated to allow an independent party to conduct a separate, stand-alone study that both collects the data and then duplicates the results.

Extending this line of thinking to the broader context of the literature review types proposed by Paré et al. (2015), we considered the feasibility and value that would be derived through a review that was repeatable, reproducible, and/or replicable (see Table 4). For *repeatability*, our view is consistent with that of Paré et al. (2015) that narrative reviews are recognized for having shortcomings related to explanations of how the review process was conducted, which would present difficulties. Similarly, theory development reviews, such as the broad theorizing reviews and specific theorizing reviews proposed by Leidner (2018), introduce fundamental challenges in achieving repeatability because of the difficulty of theorizing in a structured, consistent way that could be reperformed by others. However, each of the remaining genres has the potential to sufficiently detail their methodological procedures to allow for a third party to reperform the steps. By satisfying the criteria that would be required for repeatability, a review's authors provide a higher level of trustworthiness than they could with either systematicity or transparency alone. For *reproducibility*, the meta-analysis genre represents the most obvious candidate that would benefit from the capability of having an independent party duplicate its results, because of its reliance on quantitative data and standardized statistical techniques. Achieving reproducibility could provide confidence to a reader that the calculations were performed accurately and no errors were present in the results. Additionally, qualitative systematic reviews may also be able to satisfy reproducibility criteria on the basis that they rely on quantitative data derived from empirical studies.

Similarly, because of the high level of method structure employed with descriptive reviews (Paré et al., 2015; Pickering & Byrne, 2014) and scoping reviews (Arksey & O'Malley, 2005; Levac, Colquhoun, & O'Brien, 2010), these two genres also have the potential for reproducibility. Finally, although umbrella reviews remain an emerging genre, those that use quantitative data could also be expected to satisfy the criteria of reproducibility. Finally, we believe that only the literature review types oriented around theory testing—meta-analysis, qualitative systematic reviews, and umbrella reviews—have the potential to be *replicated* by a third party in a stand-alone study because of the data collection structure that each has in place, which would be required for a third party to recollect the data.

3 Future Considerations

Several promising paths remain unaddressed that pertain to the three related concepts of repeatability, reproducibility, and replicability. The first relates to the practical challenges of IS literature reviews. Whereas repeatability is currently being achieved within the existing journal format and structure, an increased focus on reproducibility introduces complexities related to data management and associated software tools that would need to be provided by the authors. Goodman et al. (2016) note that reproducibility requires "at minimum, the sharing of analytical datasets (original raw or processed data), relevant metadata, analytical code, and related software" (p. 1). Many journals in the natural sciences are equipped to receive, store, and distribute such resources and these capabilities are in place at some IS journals as well. For example, the *Journal of the*

Association for Information Systems has data policies in place, including the requirement that "all authors using empirical datasets have to make them available on request for checking by senior editors or reviewers after care has been taken to anonymize the data."[5] However, it remains to be seen whether the benefits of submitting data alongside a manuscript will be viewed by authors as too onerous or invasive.

Additional challenges also exist, such as in the case of a reviewer for a meta-analysis manuscript who wishes to access not only the data, but also the tools used by the authors, in order to recalculate the results that appear in the paper. In this situation, there are both technical (e.g. infrastructure) and logistical (e.g., copyright restrictions for software) challenges that could arise (Peng, 2011). Although some authors may design their own analysis tools to perform calculations, such as a meta-analysis based on accepted statistical procedures (e.g., Lipsey & Wilson, 2001) that could be easily shared, others use tools that are difficult to obtain, such as the customized software for the Hunter-Schmidt methods (Schmidt & Le, 2014) or expensive commercial tools such as Comprehensive Meta-Analysis. Difficulty in supplying broad access to these tools could impede the ability of a journal to confirm reproducibility.

Another area of concern relates to the volatility of literature databases and several commentators have raised issues with the database search process (e.g., Boell & Cecez-Kecmanovic, 2015b; vom Brocke et al., 2015). As new journals and conferences are added or removed from databases and the search algorithms are adjusted over time, an identical search conducted at two separate times may not always generate the exact same output. Although repeatability may still be feasible even with this volatility if authors clearly articulate the search terms, justify search decisions, and test search parameters (vom Brocke et al., 2015), it may not always be possible at a later date to replicate the same search results. Similarly, a common technique in literature reviews, particularly meta-analyses, is to collect unpublished studies in order to minimize the risk of publication bias (Rothstein, Sutton, & Borenstein, 2005; Schmidt & Hunter, 2015). This activity may involve the authors contacting a large pool of researchers to determine whether they have any unpublished studies that they are willing to privately share, but the results of such an activity in terms of who responds are highly variable. Although authors can be transparent in explaining the process they employed, repeating that process at a future time would be much more challenging.

Future research opportunities also exist for other characteristics of trustworthiness in literature reviews. Past commentaries have lauded reviews that are viewed as objective, comprehensive, and unbiased (Boell & Cecez-Kecmanovic, 2015b; Rowe, 2014; Templier & Paré, 2015; vom Brocke et al., 2015), but little substantive guidance has yet been generated to guide IS authors on how to achieve these goals. Part of the solution here may be to more intensively investigate how other fields achieve trustworthiness in their reviews and determine whether these techniques could also be adopted by IS researchers. This could include review techniques, such as those used in biology, where care is taken to mitigate potential bias in the literature search process and to equally consider the evidence as a whole, rather than to focus on specific studies (Haddaway et al., 2015). It may also be worthwhile to create supporting software tools, such as the Systematic Literature unified Review program (SLuRp), which was developed for use in software engineering reviews to manage large numbers of papers, coordinate tasks among team members, and automate some methodological steps, such as quality checks (Bowes, Hall, & Beecham, 2012).

4 Conclusion

The objective of this essay was to clarify the terminology associated with reproducibility in the context of literature reviews. We highlighted three main concerns: competing perspectives on what reproducibility means, a lack of clarity in how reproducibility differs from replicability, and when achieving reproducibility is desirable and feasible. In response, we suggest that IS scholars adopt three different terms in order to distinguish between methods that can be duplicated by an independent party (repeatability), results that can be duplicated from an existing dataset by an independent party (reproducibility), and new, stand-alone reviews that seek to corroborate or refute a prior study's results by collecting and analyzing new data (replicability). Furthermore, we indicate that theory-testing reviews (e.g., meta-analysis, qualitative systematic reviews, and umbrella reviews) are the most applicable literature review types to benefit from these characteristics, while other review forms, such as theory development reviews, are less feasible. Overall, we hope that this essay aids in continuing the rich discussion in our field, and beyond, about how authors can continue to contribute valuable, meaningful, and trustworthy insights through their review articles.

[5] Refer to https://aisel.aisnet.org/jais/policies.html#data

References

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

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., & Beekhuyzen, J. (2015). Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support. *Communications of the Association for Information Systems, 34*(8), 154-204.

Boell, S. K., & Cecez-Kecmanovic, D. (2015a). Debating systematic literature reviews (SLR) and their ramifications for IS: a rejoinder to Mike Chiasson, Briony Oates, Ulrike Schultze, and Richard Watson. *Journal of Information Technology, 30*(2), 188-193.

Boell, S. K., & Cecez-Kecmanovic, D. (2015b). On being 'systematic' in literature reviews in IS. *Journal of Information Technology, 30*(2), 161-173.

Bollen, K., Cacioppo, J. T., Kaplan, R., Krosnick, J., & Olds, J. L. (2015). *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

Borko, H., Liston, D., & Whitcomb, J. A. (2007). Genres of empirical research in teacher education. *Journal of Teacher Education, 58*(1), 3-11.

Bowes, D., Hall, T., & Beecham, S. (2012). *SLuRp: A tool to help large complex systematic literature reviews deliver valid and rigorous results.* Paper presented at the 2nd International Workshop on Evidential Assessment of Software Technologies.

Cassey, P., & Blackburn, T. M. (2006). Reproducibility and Repeatability in Ecology. *BioScience, 56*(12), 958-959.

Chiasson, M. W. (2015). Avoiding methodological overdose: a declaration for independent ends. *Journal of Information Technology, 30*(2), 174-176.

Dennis, A. R., & Valacich, J. S. (2014). A Replication Manifesto. *AIS Transactions on Replication Research, 1*(1), 1-4.

Fink, A. (2010). *Conducting research literature reviews: From the internet to paper* (3rd ed.). SAGE.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*(341), 1-6.

Haddaway, N. R., Woodcock, P., Macura, B., & Collins, A. (2015). Making literature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology, 29*(6), 1596-1605.

Leidner, D. E. (2018). Review and theory symbiosis: An introspective retrospective. *Journal for the Association for Information Systems, 19*(6), 552-567.

Levac, D., Colquhoun, H., & O'Brien, K. K., 5: 69. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(69), 1-9.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7), 1-28.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14-37.

Oates, B. J. (2015). On systematic reviews for evidence-based practice. *Journal of Information Technology, 30*(2), 177-179.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems, 37*(43), 879-910.

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems, 25*(6), 493-508.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(2), 183-199.

Peng, R. D. (2011). Reproducible research in computational science. *Science, 334*(6060), 126-1227.

Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development, 33*(3), 534-548.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication Bias in Meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments* (pp. 1-7). Wiley.

Rowe, F. (2014). What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems, 23*(3), 241-255.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). SAGE.

Schmidt, F. L., & Le, H. (2014). Software for the Hunter-Schmidt Meta-Analysis Methods, Version 2.0. from University of Iowa, Department of Management & Organizations.

Schultze, U. (2015). Skirting SLR's language trap: Reframing the "systematic" vs "traditional" literature review opposition as a continuum. *Journal of Information Technology, 30*(2), 180-184.

Schwarz, A., Mehta, M., Johnson, N., & Chin, W. W. (2007). Understanding frameworks and reviews: A commentary to assist us in moving our field forward by analyzing our past. *The Database for Advances in Information Systems, 38*(3), 29-50.

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology, 62*(10), 1013-1020.

Sylvester, A., Tate, M., & Johnstone, D. (2013). Beyond synthesis: Re-presenting heterogeneous research literature. *Behaviour & Information Technology, 32*(12), 1199-1215.

Tate, M., Furtmueller, E., Evermann, J., & Bandara, W. (2015). Introduction to the special issue: The literature review in information systems. *Communications of the Association for Information Systems, 37*(5), 103-111.

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems, 37*(6), 112-137.

Templier, M., & Paré, G. (2018). Transparency in literature reviews: An assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems, 27*(5), 503-550.

Valanides, N., & Charoula, A. (2008). An exploratory study about the role of epistemological beliefs and dispositions on learners' thinking about an ill-defined issue in solo and duo problem solving contexts. In M. S. Khine (Ed.), *Knowing, knowledge and beliefs* (pp. 197-218). Springer.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., & Plattfaut, R. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems, 37*(9), 205-224.

Watson, R. T. (2001). Introducing MISQ Review: A New Department in MIS Quarterly. *MIS Quarterly, 25*(1), 103-106.

Watson, R. T. (2015). Beyond being systematic in literature reviews in IS. *Journal of Information Technology, 30*(2), 185-187.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii-xxiii.

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*(1), 45-55.

About the Authors

W. Alec Cram is an assistant professor in the School of Accounting and Finance at the University of Waterloo. His research focuses on how information systems control initiatives can contribute to improving the performance of organizational processes, including systems development and cybersecurity management. His work has been published in outlets including *Information Systems Journal, Information & Management, European Journal of Information Systems, MIS Quarterly, Communications of the Association for Information Systems, Information Systems Research,* and *Journal of the Association for Information Systems.*

Mathieu Templier is associate professor of management information systems in the Faculty of Business Administration at Université Laval. His research interests focus on research methods, literature reviews, the adoption and implementation of information systems, as well as open and collaborative innovation. His work has appeared in the *European Journal of Information Systems, Information & Management, the Communications of the Association for Information Systems,* and *IEEE Software*.

Guy Paré is the director of the administration PhD program and a professor of information technology at HEC Montréal. His current research focuses on the adoption and effectiveness of digital health interventions as well as on literature review methods. His work has appeared in several top journals including *MIS Quarterly, Journal of the Association for Information Systems, European Journal of Information Systems,* and *Journal of Information Technology.* He was elected a member of the Social Sciences Academy of the Royal Society of Canada in 2012.