

DESIGNING A NOVEL STRATEGY FOR EXPLORING LITERATURE CORPORA

Research paper

Gerit Wagner, University of Regensburg, and HEC Montréal, Canada, gerit.wagner@hec.ca

Philip Empl, University of Regensburg, Germany, philip.empl@stud.uni-regensburg.de

Guido Schryen, Paderborn University, Germany, guido.schryen@upb.de

Abstract

Understanding a new literature corpus can be a grueling experience for junior scholars. Nevertheless, corresponding guidelines have not been updated for decades. We contend that the traditional strategy of skimming all papers and reading selected papers afterwards needs to be revised. Therefore, we design a new strategy that guides the overall exploratory process by prioritizing influential papers for initial reading, followed by skimming the remaining papers. Consistent with schemata theory, starting with in-depth reading allows readers to acquire more substantial prior content schemata, which are representative for the literature corpus and useful in the following skimming process. To this end, we develop a prototype that identifies the influential papers from a set of PDFs, which is illustrated in a case study in the IT business value domain. With the new strategy, we envision a more efficient process of exploring unknown literature corpora.

Keywords: *Reading and skimming, exploring literature, review methodology, design science research, schemata theory.*

1 Introduction

Research is knit together by citations. Exploring it, we should not treat it as incoherent fabric.

When junior scholars explore new literature corpora, they are challenged to acquire a solid understanding of the body of knowledge that has been constructed in a domain. Skimming, classifying, reading, and analyzing papers, they learn about the important cornerstones of research, appreciate their connections, and ultimately see how they cumulatively form the fundamental structure of a domain. As scholars who are unfamiliar with a domain cannot be expected to distinguish cornerstone contributions, they need appropriate strategies that guide this exploratory process. Methodological guidelines on reviewing literature commonly suggest skimming and classifying all papers, and then selecting papers that require thorough reading (e.g., vom Brocke et al. 2015, Hart 1998, Ridley 2012, Blumberg et al. 2005).

Although these guidelines are appealing, the struggles of junior scholars are hard to overlook. For example, Kwan (2008, p.43) concludes from her survey that “*the process of reviewing is a grueling experience*” to many doctoral students, and Okoli (2015), considers the inefficiencies with which doctoral literature reviews are typically conducted to result in an incredible waste of time. Reasons thereof are manifold. First, junior scholars are inundated by an overwhelming amount of publications with literature corpora comprising several hundreds of papers (Kwan, 2008; Okoli, 2015). Second, they are challenged to assess the “*relevance and quality of research papers at the outset of a study*”, which is especially difficult for inexperienced researchers (vom Brocke et al., 2015, p.211). Finally, reviewing requires a considerable amount of time, with reading activities consuming most of it; this process may take as long as six months (Kwan, 2008; Leidner, 2018; Okoli, 2015). In light of these challenges, researchers have concluded that there is a lack of appropriate strategies for junior scholars to “*constrain their reading*

[...] within a reasonable time frame” and to guide the direction of reviewing (Kwan, 2008, p.53), that “*performance of reading, however, is usually left both unprobed and unaided*” (Van Pletzen, 2006, p.105), and that reading is “*often overlooked within the graduate education context*” with research into best practices being “*sorely needed*” (Urquhart et al., 2016, p.155). In a nutshell, literature corpora are overwhelming, distinguishing relevant papers is difficult, and time to read is scarce. This necessitates more strategic approaches for reviewing literature in a new domain.

The guiding research question of this paper is: *How can an overall reading-and-skimming strategy be designed to support readers in exploring new literature corpora efficiently?* To address this question, we propose a strategy that guides the overall order of reading and skimming throughout a literature corpus, which comprises multiple papers. While existing guidelines suggest skimming and classifying all papers, and then selecting papers that require thorough reading, we suggest reversing these phases. In a first phase, our strategy is based on a mechanism that prescribes the order in which influential papers are read. In a second phase, our strategy suggests skimming the remaining papers to efficiently achieve an understanding of the whole literature corpus. The hypothesized mechanism that makes our strategy effective is that in-depth reading of influential papers allows readers to acquire a better understanding of the foundation of a domain. Furthermore, this understanding makes it easier for scholars to classify the remaining papers afterwards because the remaining papers build on those influential cornerstone works, reuse their concepts, and share the same assumptions, for example. This mechanism is based on schemata theory (Anderson, 1984b), which describes the underlying cognitive comprehension processes as associating new pieces of information with existing content schemata, or prior knowledge. This notion of schemata, which serves as our underlying kernel theory, provides a theoretical rationale as to *why* the strategy is effective.

The remainder of this paper is structured as follows. The next section frames the problem of skimming and reading as part of the review process by presenting extant methodological guidelines and summarizing current practices. In Section 3, we design the strategy and discuss principles guiding its design. To demonstrate the usefulness of the proposed strategy, we instantiate the order-setting mechanism, apply it in an illustrative scenario, and propose how the underlying hypotheses can be tested in future research (Section 4). Section 5 concludes the paper.

2 Framing the Reading and Skimming Process

We frame the skimming and reading phases as part of reviewing literature, distinguish it from the preceding search and succeeding writing phase, and conceptualize its inputs and outputs. While *reading* refers to studying a paper in-depth and reading all of its content, *skimming* refers to a rapid and superficial examination of a paper, or its main ideas. We focus on skimming and reading as part of the literature review process, in which scholars aim at understanding a new literature corpus. Reading and skimming are critical when scholars focus on new domains of which they have limited prior knowledge.

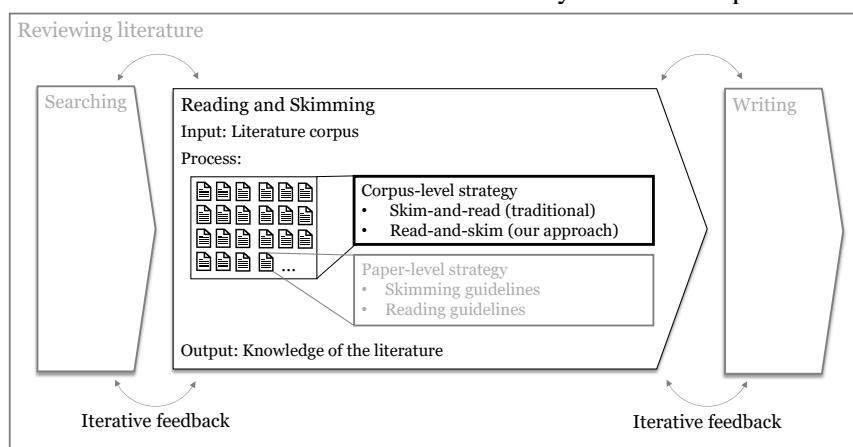


Figure 1. A framework for the reading and skimming process.

To clarify our focus on the skimming and reading phase, our framework (Figure 1) distinguishes it from the preceding phase of searching the literature and the succeeding phase of writing a synthesis. The framework borrows from extant literature (e.g., Boell and Cecez-Kecmanovic 2014, Paré et al. 2016) and serves the purpose of clarifying interactions with and iterations between preceding and succeeding phases and therefore does not depict the whole process of reviewing literature. Methods for searching the literature (e.g., vom Brocke et al. 2015), typologies (e.g., Schryen et al. 2020), standards for reporting the methodological approach (Templier and Paré, 2018), recommendations for iterating between those phases (e.g., Boell and Cecez-Kecmanovic 2014), and suggestions on how to incorporate feedback (e.g., Yang and Carless 2013) can be found in the literature. We specifically focus on the initial yet substantial skimming and reading phase that primarily aims at gaining an understanding of a relatively comprehensive set of papers. While standalone review papers may represent the most prevalent example, skimming and reading activities are also necessary for developing related work sections and situating research in extant literature (Schryen et al., 2017). This phase may be followed by a more structured classification of papers that are in scope for a subsequent literature review (e.g., Paré et al. 2016).

The phase of reading and skimming draws on a literature corpus identified in the search phase. We focus on the construction of knowledge through skimming and reading activities, which are cognitive intellectual processes whose main, though intangible aim is an in-depth understanding of the literature in the respective domain (Boell and Cecez-Kecmanovic, 2014). As such, these activities are contingent on interactions with prior knowledge. Furthermore, insights gained from these activities may provide feedback to the previous phase and trigger additional search iterations.

We focus on corpus-level strategies that involve setting the overall order of reading and skimming activities. On this level, the prevalent recommendation is to “*skim everything and decide what to read afterwards*”, which we refer to as the traditional *skim-and-read strategy*. Consistent with the assumption that the reader has no knowledge of the domain and its literature, existing approaches commonly advise readers to move “*from the general to the particular*” (Hart, 1998, p.53) by skimming all papers first and deciding which of those should be read in detail (Blumberg et al., 2005, p.32, Boell and Cecez-Kecmanovic 2014). To identify relevant papers that should be read thoroughly, methodologists provide heuristics, such as those summarized by Blumberg et al. (2005, p.131): (1) prominence (citations), (2) recency, (3) methodological quality, (4) fit with arguments of own paper, and (5) uniqueness. Similarly, vom Brocke et al. (2015) suggest reviewing citation indices for assessing the relevance of a paper. Another common suggestion is to consult with a supervisor or external mentor who has more expertise in the domain (e.g., Cooper 1998), but evidence from practice suggests that this is not always possible (e.g., Kwan 2009, McAlpine 2012). In the absence of expert advice, junior scholars may find it difficult to decide on the overall order of reading and skimming by assessing how unique, methodologically sound, or prominent a given paper is compared to a literature corpus they have not yet explored.

On the paper-level, strategies and guidelines on reading and skimming have been published (Israel and Duffy, 2017; Renear and Palmer, 2009). This includes theorization and empirical evidence on the metacognition of reading (Grabe, 2009), aspects related to self-regulation (Paris and Myers, 1981), and negative effects of mind wandering and fatigue (Unsworth and McMillan, 2013). Insights from this literature have been taken up by methodologists of literature reviews (e.g., Díaz et al. 2017, Ridley 2012, vom Brocke et al. 2015, Hart 1998). For example, Hart (1998, p.54) advises readers to “*skim through the text, get information about the structure and the general idea, read the preface and introduction, and read the parts of the text that are marked as important by yourself*”. These activities can be augmented by technology for rapid scanning, filtering and ontology-based linking of documents, for example (Renear and Palmer, 2009). In contrast, in-depth reading involves critical reading, an assessment of strengths and weaknesses, and summarizing, for example (Boell and Cecez-Kecmanovic 2014). Detailed guidance can be found in the literature (e.g., Paris et al. 1983, Spivey 1990).

In summary, we contend that it is time to develop more effective alternatives to the traditional *skim-and-read strategy* at the level of a whole literature corpus. Furthermore, there is a lack of theoretical rationales that describe *why* these strategies are effective from a cognitive perspective. In the next section, we present our novel strategy that guides the overall order of reading and skimming activities.

3 Designing a Strategy for Exploring Literature Corpora

The presentation of our strategy (cf., Table 1) is structured according to the components of design theories (Walls et al., 1992) and their interrelations (Arazy et al., 2010). We specify the requirements which are derived from the problem context and the literature, and then propose the strategy design. The kernel theory informs the design and provides a rationale for the mechanisms that are reflected in our hypotheses. This theoretical ‘because of’ justification provided by the kernel theory backs the ‘how to’ design principles (Markus et al., 2002).

Component ^a	Description
Requirements	<p>R1: Readers cannot be assumed to have prior knowledge (content schemata) of the domain</p> <p>R2: The literature corpus is coherent, i.e., the included papers are connected with regard to structure (i.e., citations) and domain knowledge</p> <p>R3: The strategy operates on the whole literature corpus (the coverage must not be limited); extensions of the literature corpus are optional</p> <p>R4: The strategy provides automated suggestions for the overall order of reading and skimming of literature corpora</p>
Strategy Design	<ul style="list-style-type: none"> • An order-setting mechanism that directs readers to influential papers in the literature corpus • A process that strategically guides the overall order of the reading and skimming activities and supports readers in developing a substantive content schema early by drawing on the order-setting mechanism, involving two main phases: <ul style="list-style-type: none"> - Phase 1: Read influential papers - Phase 2: Skim remaining papers
Kernel Theory ^b	<p>Schemata theory:</p> <ul style="list-style-type: none"> • Understanding as a process of relating new information to prior knowledge (in our context, especially content schemata) (Anderson, 1984b) • Lack of validated measures because multiple interpretations might be equally valid, especially for advanced materials (Leslie and Caldwell, 2009) • Informal measure for substantiveness of reading comprehension: associations that indicate understanding or a lack thereof (Leslie and Caldwell, 2009, p.417)
Hypotheses	<p>Comparison of the <i>skim-and-read</i> and the novel <i>read-and-skim</i> strategy:</p> <p>H1: Comparing the reading phases, readers following <i>skim-and-read</i> do not develop more substantive content schemata than readers following <i>read-and-skim</i>.</p> <p>H2: Comparing the skimming phases, readers who follow <i>read-and-skim</i> are able to draw on more substantive prior content schemata and therefore make more associations, indicating understanding, than readers who follow <i>skim-and-read</i>.</p>

^a The strategy components are structured according to the framework of Walls et al. (1992).

^b Theoretical rationale describing how the design works and informing the hypotheses (Arazy et al., 2010).

Table 1. A strategy for exploring literature corpora.

3.1 Requirements

We derive four requirements from the goal of supporting readers in understanding new literature corpora. They are based both on the given problem context and on the literature. First, readers cannot be assumed to have substantive prior knowledge of the domain (**R1**), since, almost tautologically, the “*apprehension of meaning is the superordinate goal to nearly all reading tasks*” (Paris et al., 1983, p.301). The main point of R1 is that useful designs should be applicable by a reader who only has a set of keywords (such as *IT business value*, or *technology acceptance*) and does not feel confident to judge whether one paper in a domain is more relevant than another. With this requirement, we focus on populations of readers with limited domain knowledge. Adapted strategies for readers with substantial prior knowledge of the literature could be developed in future research. We deliberately focus on junior scholars and Ph.D. students, i.e., the group most actively reading research papers, as suggested by empirical

evidence from a large-scale study (Mohammadi et al., 2015). While a-priori knowledge in this population clearly varies, it typically covers knowledge of the English language and a broad understanding of the topics and basic concepts in the field, which were acquired in an undergraduate course or in preparation of the comprehensive examinations (which are not always a mandatory part of doctoral training). This requirement describes a class of problems that has also been recognized by Vaishnavi and Kuechler (2015, p.111), who state: "*A common problem when pursuing research in an interesting but new-to-you area is to become overwhelmed by the new information you have gathered, which, by definition, is only generally applicable to an area not well understood by anyone and especially not by you.*"

Second, the literature corpus needs to be coherent, i.e., the included papers need to be connected regarding structure (i.e., citations) and domain knowledge (**R2**). This is typically the case for papers that are identified by a keyword-based database search because scientific papers give credit to their intellectual foundation by appropriate citations (Merton, 1957). In contrast, our design does not apply to degenerated literature corpora, which can evolve when, for example, papers are selected because they have applied a particular research method regardless of the domain.

Third, the strategy operates on the whole literature corpus (the coverage must not be limited); extensions of the literature corpus are optional (**R3**). Literature corpora may not be covered completely when search results in a literature database (e.g., Google Scholar or Web of Science) are considered, for example. A particular limitation of database searches (and corresponding scores, such as *cited by* on Google Scholar) is that it may not be possible to limit the results to the exact literature corpus that has been selected. Searches typically include additional (irrelevant) papers that match the search terms, or they exclude (relevant) papers that have not been indexed (possibly due to a database embargo). Conversely, reading and skimming strategies can allow additional papers to be included in the literature corpus if this possibility is optional, i.e., if this extension can be reverted efficiently. Prohibiting extensions would arguably be equivalent to eliminating feedback from the reading phase to the search phase. In essence, exploratory strategies should allow for this feedback, but not strictly require it.

Fourth, the strategy should provide automated suggestions for the overall order of reading and skimming activities in literature corpora (**R4**). This is not only for reasons of efficiency and objectivity but also for providing reading priorities independent of domain experts, whose expertise might not be available to all readers.

3.2 Strategy design

The purpose of our design, as a methodological artefact (March and Smith, 1995), is to support a strategic approach to reading the literature, which "connotes intentionality and purpose on the part of the learner. It suggests that a person chooses one alternative action over others." (Paris et al., 1983, p.294) Consistent with this notion, we conceive methodological guidelines on reading and skimming literature corpora as (suggested) strategies. The design comprises two elements: An order-setting mechanism that identifies influential papers, and a process that guides the reading and skimming activities.

The first design element, the **order-setting mechanism**, determines the overall order of reading and skimming by prioritizing influential papers. This mechanism can draw on various measures of *intra-corpus influence*, which refers to influences between papers that are observable in the literature corpus. These measures exploit the cumulative structure of literature corpora in which authors build on the research of others, giving credit to influential ideas in the form of citations. They are *heuristic* because they deliberately ignore some information in the literature corpus to avoid the effort required for reading every paper and determining exactly those papers that provide the single best starting point. We specify the order-setting mechanism as a formula, calculating the order $i = 1, \dots, |P|$ of papers $p \in P$ recursively:

$$p_i = \arg \max_{q \in P^{(i-1)}} \sum_{r \in P} \text{influence}(q, r) \cdot w_r^{(i-1)}.$$

Note that $P^{(i-1)} := P \setminus \{p_1, \dots, p_{i-1}\}$ includes all papers p that have not been selected in previous iterations. When determining the i -th paper p_i of the reading order (i.e., after determining the first $i - 1$

papers p_1, \dots, p_{i-1}), the weights $w_r^{(i-1)}$ can be used to reduce the influence related to the $i - 1$ papers that have already been included in the reading order. To illustrate the role of these weights, consider a literature corpus that is characterized by a dominant cluster (this situation may occur in inter-disciplinary research). If the marginal utility of reading papers within this cluster drops below the utility of reading papers from other clusters, these weights can direct the reading order towards unexplored clusters.

The order-setting mechanism is generic. The prototype proposed in the next section instantiates it based on a measure of intra-corpus citations. Further possible measures could apply weights to citations based on information on (1) the quantitative extent to which paper r is used in paper q (e.g., number of in-text citations), or (2) the way in which paper r uses paper q (e.g., based on an automated qualitative analysis of citation context).

The second design element is a **three-phased process** (cf., Figure 2) guiding the exploratory process in such a way that it supports readers in developing a coherent content schema early on. It thereby reverses the traditional order of the two archetypal activities: in-depth reading and quick skimming. Before starting the reading and skimming activities, readers need to complete two initial steps: determining the time they want to invest and identifying influential papers (by calculating intra-corpus influence).

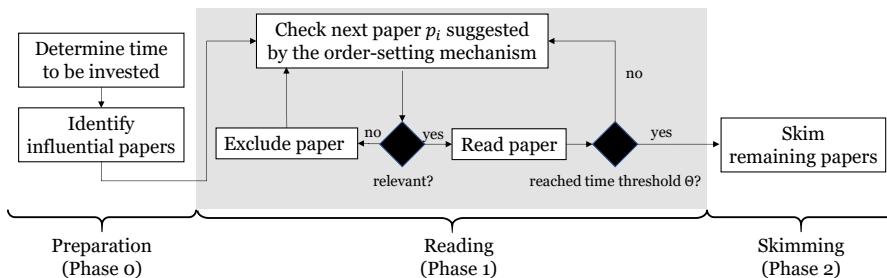


Figure 2. The exploratory process.

The initial **preparation phase** (Phase 0) requires readers to set a time limit for reading and skimming activities and to estimate the number of papers that should be read analytically. The threshold $\Theta = |R| * (\bar{t}_r - \bar{t}_s)$ for the papers that are read in detail corresponds to the summand of the following formula. Assuming that each paper needs to be skimmed or read, it mainly involves a decision on the amount of analytical reading that is to be allocated. Setting a target “level of understanding” and determining the required time is infeasible due to the problems of quantifying the “level of understanding” (cf., Subsection 3.3). Considering averaged times for reading \bar{t}_r and skimming \bar{t}_s , a reader can expect a total time of T for reading $R \in P$ papers (with $n = |R|$) and skimming all papers $S = P \setminus R$:

$$T = |R| * \underbrace{(\bar{t}_r - \bar{t}_s)}_{\text{average time for reading a paper}} + |S| * \underbrace{\bar{t}_s}_{\text{average time for skimming a paper}} \quad (\text{skimming the paper is no longer necessary})$$

The first summand serves as the threshold for transitioning from phase 1 (reading) to phase 2 (skimming). The times needed for reading and skimming are dependent on the individual reading speed and the length and complexity of papers; average readers may use 5 minutes to skim a paper and 120 minutes to read a paper of 30 pages analytically as a reasonable starting point and adjust the corresponding parameters once reading speed has been measured. To substantiate this decision, readers can inspect the distribution of the intra-corpus influence in the given literature corpus to determine whether there are distinctive inflection points. Note that the actual times needed for reading and skimming papers may vary. Nevertheless, these variations will balance out if the average reading time was estimated accurately. An instantiation of the order-setting mechanism can then be used to identify the most influential papers that are read in the first phase.

The reading phase (Phase 1) involves going through the most influential papers in an order of intra-corpus influence, checking the paper to ensure the actual fit to the research goal. Readers can apply the reading strategies mentioned in Section 2 (Framing the Problem). In terms of schemata theory (the

underlying kernel theory), readers develop a content schema that serves as prior knowledge in the following phase of skimming the remaining papers. When the overall reading efforts exceed the designated time threshold, readers can decide to invest more time or to transition to the next phase. When the time-threshold for in-depth reading is reached before all papers in R have been read, the reader can decide to allocate additional time for the remaining papers, or to move these papers from R to S (i.e., decide not to read the remaining papers, but to skim them briefly). In the latter case, the order-setting mechanism ensures that those papers with the lowest intra-corpus influence are moved to S .

The **skimming phase** (Phase 2) involves skimming the remaining papers and understanding how they relate to the developed cognitive content schema of the domain. Useful recommendations are provided in the literature (e.g., Boell and Cecez-Kecmanovic 2014).

3.3 Kernel theory

As our strategy aims at supporting readers in acquiring an understanding of the literature corpus, its design implements principles that are consistent with an appropriate (kernel) theory of the underlying cognitive processes. Specifically, it is informed by schemata theory, which explains comprehension as associating new information with prior knowledge, or schemata (Anderson, 1984b; Schmidt, 1975). This theory has replaced the conventional view that comprehension occurs by aggregating the meaning of unconnected pieces of information (Anderson, 1984a, p.598). Schemata play an essential role when interpreting different kinds of information, which may be perceived through reading text, hearing stories, or looking at sketches, for example (Anderson, 1984b). A corollary of schemata theory is that a text does not carry meaning by itself, but that it “*provides directions for readers as to how they should retrieve or construct meaning from their own previously acquired knowledge*” (An, 2013, p.130). Readers draw on different schemata, such as *formal*, *content*, and *cultural schemata* (Urquhart and Weir, 1998, p.71). In our context, we focus on developing content schemata, which refers to “background knowledge of the content area” (Carrell and Eisterhold, 1983, p.560) and “*conceptual knowledge or information about what usually happens within a certain topic, and how these happenings relate to each other to form a coherent whole*” (An, 2013, p.130).

Schemata theory is recognized as an established theory for the process of comprehending text (Anderson, 1984a), describing several functions (Anderson, 1978; Anderson and Pichert, 1978) that are useful for reading and skimming literature. Schemata offer an ideational scaffold which allows readers to associate new information with existing knowledge. They allow readers to search their memories systematically and recognize information that can be recalled. By enabling readers to distinguish important information, and to allocate their attention accordingly, schemata also facilitate summarizing activities. Furthermore, schemata enable readers to make inferential elaborations when information provided in a text (e.g., an abstract) is incomplete or not completely explicit (Anderson, 1984a, pp.598–599). Empirical research has confirmed that readers with different content schemata recall information differently after completing reading activities (e.g., Steffensen et al. 1979). Comprehension processes therefore result in different interpretations of the same text depending on the readers perspective on the domain (Bernstein, 2011, p.139). This explains why research on reading comprehension faces challenges in deriving validated formal and informal measures for reading comprehension, especially for advanced material (Leslie and Caldwell, 2009).

3.4 Hypotheses

To inform future validation of our strategy (Schuster et al., 2018; Walls et al., 1992), we derive testable hypotheses that enable comparisons of how the prototype (cf., the following section) performs in comparison to alternative approaches. In our context, the critical mechanism of our strategy, according to schemata theory, is to facilitate the development of cognitive content schemata by the reader. In this regard, empirical validation raises the question of how the effectiveness of our design should be investigated. Does one of the alternative strategies better support the development of content schemata? How does the strategy enable readers to make associations when a new paper is read or skimmed? As

mentioned in the previous subsection, there is a lack of validated formal measures for the process of understanding. Nevertheless, informal measures, such as think-aloud associations that indicate understanding or a lack thereof (Leslie and Caldwell, 2009, p.417), can provide insights into how the strategies differ in terms of their comprehension activities. These criteria provide a basis for comparing our *read-and-skim* strategy with the traditional *skim-and-read* strategy. In the following, we hypothesize on mechanisms that make our strategy more useful for understanding a given literature corpus (cf., Figure 3).

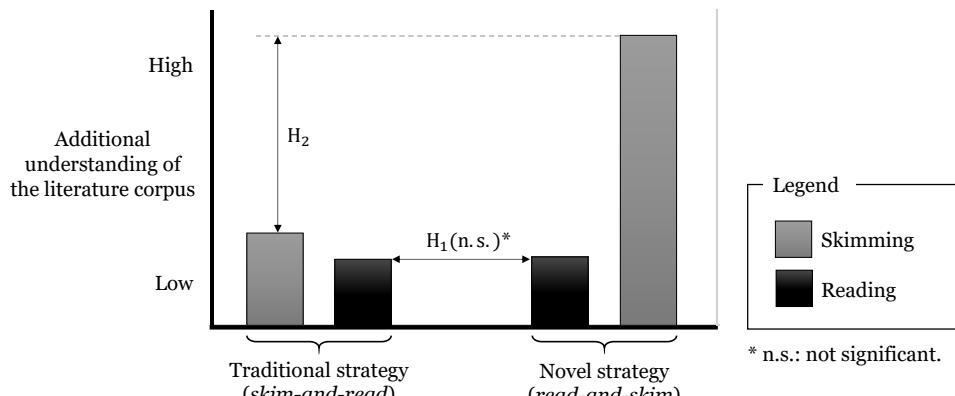


Figure 3. Hypotheses comparing the strategies.

The hypotheses are derived from the mechanisms described by the kernel theory and incorporated in our design. In our context, schemata theory suggests that understanding of new papers is contingent on existing content schemata, which are established by previous reading or skimming activities in our context. Our hypotheses deliberately compare the comprehension process between the reading phases (**H1**) and between the skimming phases (**H2**). This allows for a fairer evaluation than comparisons within phase 1 (skimming in the traditional strategy and reading in the novel strategy) and within phase 2 (reading in the traditional strategy and skimming in the novel strategy), since skimming and reading activities entail different time investments and levels of immersion (superficial vs. in-depth). Consistent with this kernel theory and similar to Vitharana et al. (2016), we propose two hypotheses. First, comparing the reading phases, we hypothesize that readers who follow the *skim-and-read* strategy do not develop more substantive content schemata than readers who follow the *read-and-skim* strategy. Second, comparing the skimming phases, we hypothesize that readers who follow the *read-and-skim* strategy are able to draw on more substantive prior content schemata and therefore make better associations than readers who follow the *skim-and-read* strategy. Figure 3 illustrates that the main difference between both strategies is hypothesized to exist between the effectiveness of the skimming activities (**H2**).

The reading phases cover the primary activities of in-depth engagement with selected papers that allow readers to construct substantive content schemata. In this phase, readers construct content schemata that allow them to understand the respective paper(s) at hand. This understanding pertains to various aspects of a paper, for example to the focal phenomenon, the constructs, the theoretical relationships, the research designs, the empirical evidence, and the implications. Naturally, this engagement with the complexity of individual research papers coincides with a focus on understanding the selected papers as opposed to understanding the rest of the literature corpus. This focus on a limited set of papers provides a straightforward rationale for expecting a limited understanding of the whole literature corpus after the reading activities for both strategies (cf., Figure 3). Developing more substantive content schemata in the *skim-and-read* strategy would require retrospective associations, i.e., readers could memorize papers they have skimmed and retrospectively associate them with a content schema that is constructed during the reading phase. We expect the effect of retrospective associations to be negligible since readers have a limited capacity to memorize (unconnected) information (Daneman and Carpenter, 1980), and since these associations would have to pertain to large quantities of papers that have been skimmed rapidly, superficially, and in an arbitrary order. We therefore propose the following hypothesis:

H1: Comparing the reading phases, readers who follow the *skim-and-read* strategy do not develop more substantive content schemata than readers who follow the *read-and-skim* strategy.

The skimming phase refers to activities of rapidly examining large quantities of literature, often involving several hundreds of papers, each of which is typically skimmed in less than 5 or 10 minutes. Due to frequent switching between papers, this process is highly contingent upon effective associations with prior content schemata, which serve as an ideational scaffold, or framework of the domain. We expect content schemata to be particularly useful when they are shaped by the influential papers in a literature corpus (*read-and-skim* strategy). This intra-corpus influence increases the likelihood that skimmed papers cumulatively build on or extend the influential papers that have been read in the previous phase (novel strategy). We therefore contend that skimming remaining papers leads to more associations, indicating understanding, when prior content schemata have been constructed from influential papers. In contrast, skimming all papers in an arbitrary order (*skim-and-read* strategy) makes it challenging to understand the content of a paper and organize it in a coherent content schema. In a typical skimming phase, this construction of the content schema may even lack an overview of theoretical models or concepts. Furthermore, papers focusing on details of a domain's knowledge, e.g., methodological problems, may remain unconnected in the overall content schema. We therefore propose the following hypothesis:

- H2:** Comparing the skimming phases, readers who follow the *read-and-skim* strategy are able to draw on more substantive prior content schemata and therefore make more effective associations, indicating understanding, than readers who follow the *skim-and-read* strategy.

4 Instantiation and Proposed Evaluation

To illustrate our strategy, we instantiate the order-setting mechanism by developing a prototype and evaluating it in an illustrative case study. The recommended exploratory process is provided in Figure 2 in the previous section. We conclude this section by summarizing how the hypotheses on the overall cognitive comprehension process can be evaluated.

4.1 Instantiation of the prototype

To instantiate the order-setting mechanism, we developed *ENLIT* (Exploring New LITERature), a prototype that determines the aggregated influence of papers in the literature corpus by matching citations from the reference sections. *ENLIT* is available online under the MIT open-source license (Empl and Wagner, 2018). The prototype is used in our illustrative case study in the following section. As stated in the requirements, the input of the prototype is a set of PDFs. To process the input, the prototype draws on GROBID for extracting information from scholarly documents. GROBID processes the PDFs and provides an XML output that conforms to the TEI standard and represents a document model. In particular, references are annotated and linked to in-text citations. The accuracy of the output was convincing and is expected to improve even further as GROBID is updated frequently. To determine the aggregated influence of papers on the corpus, our prototype matches reference sections automatically (cf., R4). The matching process is consistent with recommendations in the literature (e.g., Fellegi and Sunter 1969, and Lee et al. 2007) and the implementation draws on the matching algorithm of Jabref (published under an MIT license). Matching accuracy was substantial with minor problems related to journal abbreviations. We addressed this problem by replacing abbreviated journals before matching the references.

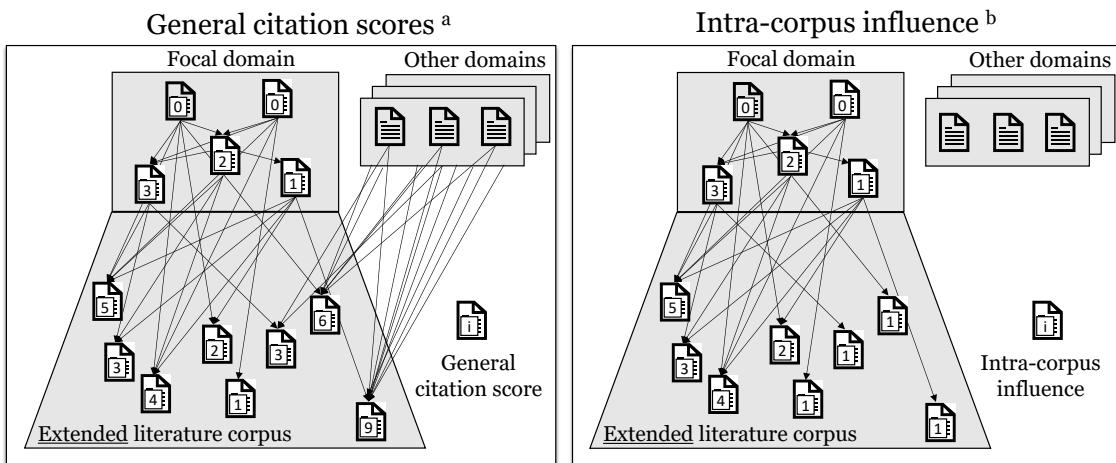
We instantiate the order-setting mechanism based on the number of intra-corpus citations as follows:

$$\text{influence}(i, p) = \begin{cases} 1 & \text{if paper } p \text{ cites paper } i \\ 0 & \text{otherwise} \end{cases}$$

Although the existence of a citation is only a rough heuristic for influence (cf., R2), it is consistent with the normative theory of citing behavior (Merton, 1957). The annotations provided by GROBID enable the development of more refined measures that consider the number of in-text citations or the semantics of the citation contexts. To keep our illustrative scenario simple, we deliberately assume additive effects (and set $w_r^{(i-1)} = 0$). Thus, the instantiated order-setting mechanism reduces to:

$$p_i = \arg \max_{q \in P^{(i-1)}} \sum_{r \in P} \text{influence}(q, r).$$

To demonstrate how intra-corpus influence differs from general citation scores, we apply both measures in a setting which is characterized by several of the requirements outlined above. Specifically, we apply the influence measures after a keyword-search has been completed and after the data of the corresponding citation network has been extracted. As this citation network includes additional (cited) papers, we refer to the resulting set as the *extended literature corpus*, which comprises both the papers available in full-text and their references¹. This is consistent with considering citations by papers contained in the literature corpus as suggesting that the cited paper is somehow relevant and cannot be excluded a-priori – a notion that readers who have limited prior knowledge (content schemata) of the domain (R1) may not dismiss easily. Our setting is also consistent with the option to extend the literature corpus (cf., R3). The exemplary setting described in the following shows that intra-corpus influence identifies papers that are more relevant to the domain because in contrast to general citation scores, it excludes citations from other domains (as illustrated in Figure 4).



Notes. ^a Recommended in the literature. ^b Proposed in this paper.

Figure 4. Comparison of order-setting mechanisms.

In the exemplary setting, we deliberately selected a literature search with less than perfect coverage. Showing that our strategy performs well on a literature corpus crafted by a senior scholar would not show convincingly that it is also useful for scholars who do not have prior knowledge in the domain (cf., R1). Specifically, we focus on the IT business value domain. We conducted a simple database search for the keyword “IT business value” and included all papers published in the AIS Senior Scholars’ Basket of journals between 2005 and 2015. This scope can barely be considered to meet the minimum requirements of a scholarly literature search. The search yielded 160 papers, for which intra-corpus influences were calculated (cf., Table 3). To assess the appropriateness of the papers recommended for in-depth reading, and thereby the effectiveness of the order-setting mechanism, we evaluate the papers identified by the intra-corpus influence score (Table 3) and compare them to the ranking of papers according to a general citation score (Table 2). The binary assessment of *relevance to the domain* (last column) was conducted by one of the authors who has published multiple papers on IT business value, including review papers (Schryen, 2010; Schryen, 2013). The assessment is evident from the titles of the respective papers.

To show how intra-corpus influence differs from the total amount of citations for intra-corpus papers, Table 2 provides general citation scores from Web of Science. It shows that, in the extended literature corpus, general citation scores do not identify papers relevant to IT-business value. Instead, the identified papers focus mostly on research methodology. Hence, one lesson is that while papers influential to the

¹ Note that the extended literature corpus contains additional papers whose relevance to the domain varies considerably. In contrast to the original literature corpus, in which the relevance of the papers to the domain varies only slightly, this variance allows us to clarify the difference between intra-corpus influence and general citation scores. The extended literature corpus does not require us to make arbitrary inclusion decisions (such as randomly adding irrelevant papers to the literature corpus) to evaluate the influence measures.

domain (Table 3) may be cited frequently, highly cited papers (Table 2) may not necessarily be relevant to the domain (knowledge) as such. This distinction has not been discussed by methodologists who recommend general citation scores as an indicator of paper relevance (e.g., Blumberg et al. 2005, p.131).

Reference from the (extended) literature corpus	Citations	Relevance to the domain (expert assessment)
Cohen (1988)	48,813	No
Baron and Kenny (1986)	31,672	No
Cox (1972)	31,438	No
Strauss (1967)	26,336	No
Landis and Koch (1977)	23,430	No
Hu and Bentler (1999)	21,242	No
Miles and Huberman (1994)	21,185	No
Dempster et al. (1977)	21,139	No
Aiken et al. (1991)	19,905	No
Nunnally (1978)	18,961	No

Table 2. Priorities based on general citation scores provided by Web-of-Science (Top-10).

In contrast to general citation scores (Table 2), intra-corpus influence (Table 3) identifies papers that are relevant to the IT business values domain. The papers that have introduced the resource-based view in IS can also be found in Table 3, most noteworthy the paper of Bharadwaj (2000), which has paved the way for many other empirical papers, and the review of Melville et al. (2004), which constitutes the theoretical foundation of research on IT business value. All papers listed in Table 3 are authored by prolific scholars; eight papers are published in main IS journals and two papers in general management journals. In position four, readers find the influential paper of Barney (2011), published in the *Journal of Management*. The paper of Teece et al. (1997), published in the *Strategic Management Journal*, and further papers following the Top-10 (e.g., Henderson and Venkatraman 1999) are relevant to IS journals but have not been published in typical IS journals. This suggests that intra-corpus influence captures associations related to domain knowledge without necessarily restricting them to top-journals, such as the AIS Senior Scholars' basket of journals. In this regard, considering the extended literature corpus, similar to a backward search, may help Ph.D. students to include further relevant papers.

Reference from the (extended) literature corpus	Intra-corpus influence	Relevance to the domain (expert assessment)
Bharadwaj (2000)	56	Yes
Sambamurthy et al. (2003)	44	Yes
Melville et al. (2004)	43	Yes
Barney (1991)	42	Yes
Mata et al. (1995)	36	Yes
Barua et al. (1995)	32	Yes
Hitt and Brynjolfsson (1996)	29	Yes
Kohli and Grover (2008)	28	Yes
Tallon et al. (2000)	27	Yes
Teece et al. (1997)	26	Yes

Table 3. Priorities based on intra-corpus influence, i.e., papers prioritized for in-depth reading (Top-10).

While the measure of intra-corpus citations has certainly identified influential papers, one may speculate whether other papers, such as the review of Wade and Hulland (2004) which is not included in the top-10 list, should be ranked higher. While a definitive judgement on the relative influence of single papers may be elusive, the order-setting mechanism appears to be effective in identifying influential papers that are relevant to the IT-business value domain. While its effectiveness in other domains remains to be evaluated, it seems likely that scholars who read the identified papers are well equipped to understand how the remaining IT-business value literature relates to this foundation.

4.2 Proposed evaluation of the hypotheses

In addition to the hypotheses and the prototype (expository instantiation), we outline how the design theory can be tested empirically (cf., Schuster et al. 2018). Considering parsimony and richness as the dual objectives of design (Baskerville et al., 2015; Briggs and Schwabe, 2011), we emphasize the need to assess the effectiveness of our strategy before refining it and introducing additional complexity. We propose testing the hypotheses through quasi-experiments in class-room environments in which two groups of participants are assigned either to the new *read-and-skim* strategy or to the traditional *skim-and-read* strategy (control group). Consistent with previous empirical research on schemata theory (e.g., Tynjälä 1999), both groups should complete the assignments using identical literature corpora to eliminate variation specific to the literature corpus. The following experimental procedures should be implemented. To control for possible confounds, participants should be required to complete a survey on their prior knowledge (content schemata) of the selected domain, their current research foci and expertise, as well as their language proficiency (Lundeberg, 1987). In the course of the experiment, participants should be provided with a literature corpus on a predefined topic, which requires at least a few days to read and skim. A description of the respective steps including guidelines should be provided; however the hypothesized mechanisms should not be disclosed to avoid that expectations regarding the effectiveness of either strategy give rise to corresponding confirmation biases (Campbell, 1957).

Concerning hypothesis **H1**, it needs to be ascertained that, comparing the reading phases, readers who follow the *skim-and-read* strategy do not develop more substantive content schemata than readers who follow the *read-and-skim* strategy. To assess the substantiveness of content schemata after the respective reading phases, participants should be asked to summarize their knowledge of the domain (cmp. Leslie and Caldwell 2009). To ensure a fair assessment, both groups should be interviewed after the same amount of time has passed. The summaries can be analyzed qualitatively by an expert in this domain, who gauges the internal consistency and completeness of each summary. Through consistent elaborations or distorted recalls, these answers can provide further insights into the appropriateness of the respective content schema (Anderson, 1984a; Spiro et al., 2017). In a final step, the assessment of the content schemata can be compared between groups, while considering the control variables collected in the survey (e.g., level of expertise in neighboring fields).

Concerning hypothesis **H2**, it needs to be tested whether, comparing the skimming phases, readers who follow the *read-and-skim* strategy are able to draw on more substantive prior content schemata and therefore make more associations, indicating understanding, than readers who follow the *skim-and-read* strategy. Due to the amount of papers that need to be skimmed, readers should be required to keep a reading diary in which they describe their experiences when skimming (or reading) papers. To measure reading comprehension for advanced materials, like a whole literature corpus, readers' comments could be classified as associations that indicate understanding or a lack thereof (Leslie and Caldwell, 2009). Exemplary items could include "I understood the paper/abstract.", "I know how the paper/abstract relates to other papers.", "Some information was implicit; however, I was able to understand it nonetheless.", or "When reading/skimming the paper, I immediately thought of another paper" (cmp. Spiro et al. 2017).

5 Conclusion

Our novel strategy for exploring literature corpora makes three contributions to extant research. First, it adds to the methodological literature on literature reviews, which has arguably neglected the reading phase and methods that guide the overall order of the reading and skimming activities for too long. Specifically, we suggest that the traditional approach of *skim-and-read* should be replaced by the new *read-and-skim* strategy. With this, we advocate for explicitly documenting corresponding guidelines. Similar to other research methods, guidance on the exploratory process should, in principle, follow explicitly codified methodological procedures that do not depend on tacit and experiential knowledge (e.g., provided by thesis supervisors). Second, the design of our strategy makes a theoretical contribution by backing the ‘how to’ design principles with theoretically grounded ‘because of’ justifications, explaining *why* the design works (Markus et al., 2002; Walls et al., 1992). By drawing on schemata theory, the novel exploratory strategy aligns with established theories of comprehension and advances the use of cognitive process theories as informing IS design science theories. Third, the literature on comprehension, learning, and memory offers a plethora of research in teaching contexts, traditionally focusing on primary and secondary education. The academic context has received less attention, with research focusing primarily on the comprehension of individual papers, for example, by providing reading strategies that take into account the structure of a paper (e.g., Paris et al. 1983). We contend that the mechanisms described by schemata theory apply to bigger literature corpora, but that corresponding implications and instructions need to be adapted. In doing so, we suggest that studying the comprehension of whole literature corpora – although requiring more extensive experimental designs – represents significant opportunities for research that is grounded in cognitive theories.

This paper contributes to (research) practice by designing a novel strategy that guides the overall order of reading and skimming. The importance of such design-oriented research that supports readers, and especially Ph.D. students and junior scholars, in familiarizing themselves with new research domains has been highlighted in extant literature (Vaishnavi and Kuechler, 2015, p.111). A distinctive feature of our strategy is that it emphasizes the hermeneutic value of in-depth reading, complementing approaches that emphasize (partly) automated processing of large quantities of papers. We encourage teachers to recommend our method to (Ph.D.) students and methodologists to consider our insights when textbooks are revised. It is time to update existing guidelines, which have neither been challenged nor revised in the light of improved technological capabilities or theories about the cognitive processes involved in understanding new literatures. With our novel exploratory strategy, we envision a reading and skimming process that is less of a grueling experience and more congruent with the underlying cognitive processes involved in understanding new literature corpora.

Acknowledgments

We are grateful for the feedback provided by the associate editor and the anonymous reviewers, which helped us to clarify the paper. We would also like to thank Osamu Takeuchi for recommending schemata theory as a foundation, and Julian Prester, Richard Schuster, and Gerhard Raucheker for their feedback on earlier versions of this paper. The research is supported by a grant of the German Science Foundation (DFG) for the research project “Epistemological Advances Through Qualitative Literature Reviews in Information Systems Research” (EPIQUALIS) (<http://gepris.dfg.de/gepris/projekt/315925033?language=en>).

References

- Aiken, L.S., S.G. West and R.R. Reno (1991). *Multiple Regression: Testing and Interpreting Interactions*, Sage.
- An, S. (2013). "Schema Theory in Reading." *Theory and Practice in Language Studies* 3(1), pp.130–134.
- Anderson, R.C. (1984a). "Role of the Reader's Schema in Comprehension, Learning, and Memory." In: R. C. Anderson, J. Osborn, & R. J. Tierney, eds. *Learning to Read in American Schools: Basal Readers and Content Texts*. Hillsdale, New Jersey, USA, pp. 243–257.
- Anderson, R.C. (1978). "Schema-directed Processes in Language Comprehension." In: A. M. Lesgold et al., eds. *Cognitive Psychology and Instruction*. Boston, MA, pp. 67–82.
- Anderson, R.C. (1984b). "The Notion of Schemata and the Educational Enterprise: General Discussion of the Conference." In: R. C. Anderson, S. R. J., & W. E. Montague, eds. *Schooling and the Acquisition of Knowledge*. Hillsdale, New Jersey, USA, pp. 415–431.
- Anderson, R.C. and J.W. Pichert (1978). "Recall of Previously Unrecallable Information Following a Shift in Perspective." *Journal of Verbal Learning and Verbal Behavior* 17(1), pp.1–12.
- Arazy, O., N. Kumar and B. Shapira (2010). "A Theory-driven Design Framework for Social Recommender Systems." *Journal of the Association for Information Systems* 11(9), pp.455–490.
- Barney, J. (1991). "Firm Resources and Sustained Competitive Advantage." *Journal of Management* 17(1), pp.99–120.
- Baron, R.M. and D.A. Kenny (1986). "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6), p.1173.
- Barua, A., C.H. Kriebel and T. Mukhopadhyay (1995). "Information Technologies and Business Value: An Analytic and Empirical Investigation." *Information Systems Research* 6(1), pp.3–23.
- Baskerville, R.L., M. Kaul and V.C. Storey (2015). "Genres of Inquiry in Design-science Research: Justification and Evaluation of Knowledge Production." *MIS Quarterly* 39(3), pp.541–564.
- Bernstein, R.J. (2011). *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*, University of Pennsylvania Press.
- Bharadwaj, A. (2000). "A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation." *MIS Quarterly* 24(1), pp.169–196.
- Blumberg, B., D.R. Cooper and P.S. Schindler (2005). *Business Research Methods* 2nd ed., New York: McGraw-Hill.
- Boell, S.K. and D. Cecez-Kecmanovic (2014). "A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches." *Communications of the Association for Information Systems* 34(12), pp.257–286.
- Briggs, R.O. and G. Schwabe (2011). "On Expanding the Scope of Design Science in IS Research." In: H. Jain, A. P. Sinha, & P. Vitharana, eds. *International Conference on Design Science Research in Information Systems*. Milwaukee, WI, USA, pp. 92–106.
- Campbell, D.T. (1957). "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54(4), pp.297–312.
- Carrell, P.L. and J.C. Eisterhold (1983). "Schema Theory and ESL Reading Pedagogy." *TESOL Quarterly* 17(4), pp.553–573.
- Cohen, J. (1988). "Statistical Power Analysis for the Behavioral Sciences." *Hilsdale*. Lawrence Earlbauim Associates.
- Cooper, H.M. (1998). *Synthesizing Research: A Guide for Literature Reviews*, New York, NY, USA: Sage.
- Cox, D.R. (1972). "Regression Models and Life Tables (With Discussion)." *Journal of the Royal Statistical Society, Series B* 34, pp.187–220.
- Daneman, M. and P.A. Carpenter (1980). "Individual Differences in Working Memory and Reading." *Journal of Verbal Learning and Verbal Behavior* 19(4), pp.450–466.

- Dempster, A.P., N.M. Laird and D.B. Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.1–38.
- Díaz, O., J.P. Contell and J.R. Venable (2017). "Strategic Reading in Design Science: Let Root-cause Analysis Guide your Readings." In: A. Maedche, J. vom Brocke, & A. Hevner, eds. *International Conference on Design Science Research in Information Systems*. Karlsruhe, Germany, pp. 231–246.
- Empl, P. and G. Wagner (2018). "ENLIT (Exploring New Literature) Software, Version 1.0." Available at: 10.5281/zenodo.1226650.
- Fellegi, I.P. and A.B. Sunter (1969). "A Theory for Record Linkage." *Journal of the American Statistical Association* 64(328), pp.1183–1210.
- Grabe, W. (2009). *Reading in a Second Language: Moving from Theory to Practice*, Cambridge University Press.
- Hart, C. (1998). *Doing a Literature Review: Releasing the Social Science Research Imagination*, London: Sage.
- Henderson, J.C. and H. Venkatraman (1999). "Strategic Alignment: Leveraging Information Technology for Transforming Organizations." *IBM Systems Journal* 38(2.3), pp.472–484.
- Hitt, L.M. and E. Brynjolfsson (1996). "Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value." *MIS Quarterly* 20(2), pp.121–142.
- Hu, L. and P.M. Bentler (1999). "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus new Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6(1), pp.1–55.
- Israel, S.E. and G.G. Duffy (2017). *Handbook of Research on Reading Comprehension* 2nd ed., New York, USA: The Guilford Press.
- Kohli, R. and V. Grover (2008). "Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times." *Journal of the Association for Information Systems* 9(1), pp.23–39.
- Kwan, B.S.C. (2009). "Reading in Preparation for Writing a Phd Thesis: Case Studies of Experiences." *Journal of English for Academic Purposes* 8(3), pp.180–191.
- Kwan, B.S.C. (2008). "The Nexus of Reading, Writing and Researching in the Doctoral Undertaking of Humanities and Social Sciences: Implications for Literature Reviewing." *English for Specific Purposes* 27(1), pp.42–56.
- Landis, J.R. and G.G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33(1), pp.159–174.
- Lee, D. et al. (2007). "Are your Citations Clean?" *Communications of the ACM* 50(12), pp.33–38.
- Leidner, D.E. (2018). "Review and Theory Symbiosis: An Introspective Retrospective." *Journal of the Association for Information Systems* 19(6), pp.552–567.
- Leslie, L. and J. Caldwell (2009). "Formal and Informal Measures of Reading Comprehension." In: S. E. Israel & G. G. Duffy, eds. *Handbook of Research on Reading Comprehension*. New York, NY, USA, pp. 403–427.
- Lundeberg, M.A. (1987). "Metacognitive Aspects of Reading Comprehension: Studying Understanding in Legal Case Analysis." *Reading Research Quarterly* 22(4), pp.407–432.
- March, S.T. and G.F. Smith (1995). "Design and Natural Science Research on Information Technology." *Decision Support Systems* 15(4), pp.251–266.
- Markus, M.L., A. Majchrzak and L. Gasser (2002). "A Design Theory for Systems that Support Emergent Knowledge Processes." *MIS Quarterly* 26(3), pp.179–212.
- Mata, F.J., W.L. Fuerst and J.B. Barney (1995). "Information Technology and Sustained Competitive Advantage: A Resource-based Analysis." *MIS Quarterly* 19(4), pp.487–505.
- McAlpine, L. (2012). "Shining a Light on Doctoral Reading: Implications for Doctoral Identities and Pedagogies." *Innovations in Education and Teaching International* 49(4), pp.351–361.
- Melville, N., K. Kraemer and V. Gurbaxani (2004). "Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value." *MIS Quarterly* 28(2), pp.283–322.

- Merton, R.K. (1957). "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22(6), pp.635–659.
- Miles, M.B. and A.M. Huberman (1994). *Qualitative Data Analysis: An Expanded Sourcebook*, Sage.
- Mohammadi, E. et al. (2015). "Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories." *Journal of the Association for Information Science and Technology* 66(9), pp.1832–1846.
- Nunnally, J.C. (1978). "Psychometric Theory. New York." NY McGraw-Hill.
- Okoli, C. (2015). "A Guide to Conducting a Standalone Systematic Literature Review." *Communications of the Association for Information Systems* 37(1), pp.879–910.
- Paré, G. et al. (2016). "Contextualizing the Twin Concepts of Systematicity and Transparency in Information Systems Literature Reviews." *European Journal of Information Systems* 25(6), pp.493–508.
- Paris, S.G., M.Y. Lipson and K.K. Wixson (1983). "Becoming a Strategic Reader." *Contemporary Educational Psychology* 8(3), pp.293–316.
- Paris, S.G. and M. Myers (1981). "Comprehension Monitoring, Memory, and Study Strategies of Good and Poor Readers." *Journal of Reading Behavior* 13(1), pp.5–22.
- Van Pletzen, E. (2006). "A Body of Reading: Making 'visible' the Reading Experiences of First-year Medical Students." In: L. Thesen & E. van Pletzen, eds. *Academic Literacy and the Languages of Change*. A&C Black, pp. 104–129.
- Renear, A.H. and C.L. Palmer (2009). "Strategic Reading, Ontologies, and the Future of Scientific Publishing." *Science* 325(5942), pp.828–832.
- Ridley, D. (2012). *The Literature Review: A Step-by-step Guide for Students* K. Metzler, ed., London, UK: Sage.
- Sambamurthy, V., A. Bharadwaj and V. Grover (2003). "Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms." *MIS Quarterly* 27(2), pp.237–263.
- Schmidt, R.A. (1975). "A Schema Theory of Discrete Motor Skill Learning." *Psychological Review* 82(4), pp.225–260.
- Schryen, G. et al. (2020). "A Knowledge Development Perspective on Literature Reviews: Validation of a new Typology in the IS Field." *Communications of the Association for Information Systems* 46, pp.134–168.
- Schryen, G. et al. (2017). "Literature Reviews in IS Research: What Can Be Learnt from the Past and Other Fields?" *Communications of the Association for Information Systems* 41, pp.759–774.
- Schryen, G. (2010). "Preserving Knowledge on IS Business Value: What Literature Reviews Have Done." *Business & Information Systems Engineering* 2(4), pp.233–244.
- Schryen, G. (2013). "Revisiting IS Business Value Research: What We Already Know, What We Still Need to Know, and How We Can Get There." *European Journal of Information Systems* 22(2), pp.139–169.
- Schuster, R., G. Wagner and G. Schryen (2018). "Information Systems Design Science Research and Cumulative Knowledge Development: An Exploratory Study." In: *Proceedings of the International Conference on Information Systems*.
- Spiro, R.J., B.C. Bruce and W.F. Brewer (2017). *Theoretical Issues in Reading Comprehension: Perspectives From Cognitive Psychology, Linguistics, Artificial Intelligence and Education*, CRC Press.
- Spivey, N.N. (1990). "Transforming Texts: Constructive Processes in Reading and Writing." *Written Communication* 7(2), pp.256–287.
- Steffensen, M.S., C. Joag-Dev and R.C. Anderson (1979). "A Cross-cultural Perspective on Reading Comprehension." *Reading Research Quarterly* 15(1), pp.10–29.
- Strauss, A.L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*, New York: Aldine.
- Tallon, P.P., K.L. Kraemer and V. Gurbaxani (2000). "Executives' Perceptions of the Business Value of Information Technology: A Process-oriented Approach." *Journal of Management Information Systems* 16(4), pp.145–173.

- Teece, D.J., G. Pisano and A. Shuen (1997). "Dynamic Capabilities and Strategic Management." *Strategic Management Journal*, pp.509–533.
- Templier, M. and G. Paré (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), pp.503–550.
- Tynjälä, P. (1999). "Towards Expert Knowledge? A Comparison Between a Constructivist and a Traditional Learning Environment in the University." *International Journal of Educational Research* 31(5), pp.357–442.
- Unsworth, N. and B.D. McMillan (2013). "Mind Wandering and Reading Comprehension: Examining the Roles of Working Memory Capacity, Interest, Motivation, and Topic Experience." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3), pp.832–842.
- Urquhart, A.H. and C.J. Weir (1998). *Reading in a Second Language: Process, Product, and Practice* 1st ed., London, UK: Longman.
- Urquhart, S.M. et al. (2016). "Factors Associated with Novice Graduate Student Researchers' Engagement with Primary Literature." *International Journal for Researcher Development* 7(2), pp.141–158.
- Vaishnavi, V.K. and W. Kuechler (2015). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, Boca Raton, Florida, USA: Taylor & Francis.
- Vitharana, P., F. Zahedi and H.K. Jain (2016). "Enhancing Analysts' Mental Models for Improving Requirements Elicitation: A Two-stage Theoretical Framework and Empirical Results." *Journal of the Association for Information Systems* 17(12), pp.804–840.
- vom Brocke, J. et al. (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), pp.205–224.
- Wade, M. and J. Hulland (2004). "Review: The Resource-based View and Information Systems Research: Review, Extension, and Suggestions for Future Research." *MIS Quarterly* 28(1), pp.107–142.
- Walls, J.G., G.R. Widmeyer and O.A. El Sawy (1992). "Building an Information System Design Theory for Vigilant EIS." *Information Systems Research* 3(1), pp.36–59.
- Yang, M. and D. Carless (2013). "The Feedback Triangle and the Enhancement of Dialogic Feedback Processes." *Teaching in Higher Education* 18(3), pp.285–297.