

Comprehension of Procedural Visual Business Process Models

A Literature Review

Kathrin Figl

Received: 3 March 2016 / Accepted: 16 October 2016 / Published online: 2 January 2017
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Visual process models are meant to facilitate comprehension of business processes. However, in practice, process models can be difficult to understand. The main goal of this article is to clarify the sources of cognitive effort in comprehending process models. The article undertakes a comprehensive descriptive review of empirical and theoretical work in order to categorize and summarize systematically existing findings on the factors that influence comprehension of visual process models. Methodologically, the article builds on a review of forty empirical studies that measure objective comprehension of process models, seven studies that measure subjective comprehension and user preferences, and thirty-two articles that discuss the factors that influence the comprehension of process models. The article provides information systems researchers with an overview of the empirical state of the art of process model comprehension and provides recommendations for new research questions to be addressed and methods to be used in future experiments.

Keywords Process modeling · Comprehension · Human information behavior · Literature review

Accepted after three revisions by the editors of the special issue.

Electronic supplementary material The online version of this article (doi:10.1007/s12599-016-0460-2) contains supplementary material, which is available to authorized users.

Dr. MMag. K. Figl (✉)
Institute for Information Systems and New Media, Vienna
University of Economics and Business, Welthandelsplatz 1,
Building D2, 1020 Vienna, Austria
e-mail: kathrin.figl@wu.ac.at

1 Introduction

Process models visually represent the flow of an organization's business activities. One of the top tasks of process model applications is to help those involved understand the process (Indulska et al. 2009) in order to appreciate its benefits and enable organizations to profit fully from the positive impacts of process management (Škrinjar et al. 2008). Decisions made on the basis of process models tend to be better than those that are not, therefore process models can help to increase revenue, and the efficiency of managing and monitoring business processes is improved. Process models are instrumental in defining information system requirements and help to reveal errors during the requirements engineering phase, when it is comparatively easy and inexpensive to correct them (Charette 2005). Thus, improved comprehensibility of process models has a direct significance for the development, efficiency, and costs of information systems. Comprehensibility of models not only facilitates a common understanding of processes between users and system engineers but also helps improve the quality of models.

Prior contributions to the area of process-model comprehension examine a variety of influence factors in isolation, so a comprehensive body of knowledge that might provide an overview of the research field is lacking. Literature reviews are essential for progress in a field of study. Webster and Watson (2002, p. 14) note for the information systems (IS) field that “the literature review represents the foundation for research in IS. As such, review articles are critical to strengthening IS as a field of study.” In a similar vein, Recker and Mendling (2016) conclude for the business process management (BPM) discipline that literature reviews “are required in BPM that assist the development of novel theory about processes and their management.”

Therefore, the main objective of this article is to gain systematic insight into existing findings on what factors influence the intuitiveness and understandability of process models. In short, the article addresses the cognitive aspects of acquiring and interpreting information on business processes that are presented in process diagrams.

In the context of the special issue, the article's focus is on the use aspect of human information-seeking behavior, which is defined as the "totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use" (Wilson 2000, p. 49). Since the article looks at a specific source and channel of information – visual process models – which represent formal externalized knowledge of the kinds of enterprise processes that are available in most organizations (Patig et al. 2010), "information seeking" is considered in the narrow sense of seeking information inside a process model. The focus on comprehension is directly connected to the mental part of behavior related to using information, which is described as "the physical and mental acts involved in incorporating the information found into the person's existing knowledge base" (Wilson 2000, p. 50). Comprehension of process models is a type of intrapersonal information behavior in which the information is supplied in the form of a process model (Heinrich et al. 2014). In a narrower sense, behavior related to intrapersonal information encompasses tasks like the reception, selection, organization and use of information to solve tasks (Heinrich et al. 2014). As several factors that influence comprehension are considered in the article, it fits into the category of cognitivist information behavior research, which focuses on the individual user of information. However, it also considers how variations in the information artifact "process model" influence shared, intersubjective sense-making (Olsson 2005), so it extends the human information behavior research on information delivery through IS to the area of process modeling by looking at the visualization of process models and the cognitive fit between process models and tasks and users (Hemmer and Heinzl 2011).

Building on a thorough review, the article integrates findings related to theoretical perspectives and empirical data in the field into an overarching framework in order to categorize the factors that influence the comprehension of process models. The article also compares the variables that empirical research has addressed with the variables mentioned in theoretical discussions of process-model comprehension, including discussions of modeling guidelines. This endeavor is especially important because modeling guidelines have not been well tied to experimental findings (Mendling 2013).

There is a vast amount of literature on human comprehension of conceptual models in areas that range from

separate evaluations of modeling notations to reviews on how to evaluate conceptual models (e.g., Burton-Jones et al. 2009; Parsons and Cole 2005; Moody 2005; Gemino and Wand 2004). This article focuses on studies which investigate the factors that influence the comprehension of one specific type of models-business process models. In contrast to Houy et al. (2012, 2014), who focus on defining the dependent variable (model comprehension), measurement instruments, and the theoretical underpinning used in experimental studies, we focus on an overview of the independent variables (the sources of cognitive load and their relationship to the dependent variable of model comprehension).

The number of empirical studies on cognitive aspects of process models is increasing rapidly, and this topic includes a recent stream of work on the cognitive load involved in model creation (Pinggera et al. 2013; e.g., Claes et al. 2012). In contrast, the scope of the present article does not so much include the creation of process models but rather how they are understood.

This article is organized as follows: It begins with an introductory background on cognitive load in model comprehension. Then it describes how the literature search was conducted, articulates the selection criteria, and identifies the main works included in the review. Next, it presents a framework for influence factors, and based on this framework, analyzes research designs and types of variables and summarizes the results of empirical studies. After contrasting empirical studies with theoretical viewpoints and presenting research gaps, the article provides ideas for future directions in research methods and discusses the limitations of the review.

2 Process Model Comprehension and Cognitive Load

A visual model must be comprehensible if it is to be useful since, as Lindland et al. (1994, p. 47) put it, "not even the most brilliant solution to a problem would be of any use if no one could understand it." Therefore, model comprehension is a primary measure of pragmatic model quality, as distinguished from syntactic quality, which refers to how a model corresponds to a particular notation, and semantic quality, which refers to how a model corresponds to a domain (Lindland et al. 1994; Overhage et al. 2012). Research in the area of data models shows that comprehensibility is the most important influence factor in the assessment of a model's overall quality, outranking completeness, correctness, simplicity, and flexibility (Moody and Shanks 2003).

An important reference discipline for intrapersonal information-related behavior like process model comprehension is cognitive psychology (Heinrich et al.

2014). A basic precondition for comprehension is that a model does not overwhelm a reader's working memory. Working memory may become a bottleneck in comprehending complex models because it limits the amount of information that can be heeded at any one time (Baddeley 1992). The cognitive load theory (Sweller 1988), which provides a general framework for designing the presentation of instructional material to ease learning and comprehension, can also be applied to the field of process model comprehension. Overall, the working memory's maximum capacity should be available for "germane" cognitive load, which refers to the actual processing of the information and the construction of mental structures that organize elements of information into patterns (i.e., schema).

Intrinsic cognitive load is concerned with the "complexity of information that must be understood" (Sweller 2010, p. 124). Together the characteristics of the process model, such as model-based metrics, and the content of the labels and the characteristics of the comprehension task determine the intrinsic cognitive load of a comprehension task. Therefore, cognitive load is also influenced by how comprehension is measured, as comprehension performance in an experiment varies according to the questions asked (Figl and Laue 2015) and the kind and amount of assistance given to subjects (e.g., Soffer et al. 2015).

While it is difficult to change a process model's intrinsic cognitive load without changing the behavior and content of the process being modeled, the visual presentation can be changed and can have a significant impact on cognitive load without changing the modeled process. How a process is visualized relates to the "extraneous" cognitive load (Kirschner 2002). If the same process is modeled using different notations or another layout for the labels and the overall model, the resulting models will have comparable intrinsic cognitive loads but differ in their extraneous cognitive loads, affecting comprehension (Chandler and Sweller 1996). Moody (2009) identifies nine principles for designing notations so they do not cause more extraneous cognitive load than necessary: semiotic clarity, graphic economy, perceptual discriminability, visual expressiveness, dual coding, semantic transparency, cognitive fit, complexity management, and cognitive integration.

Moreover, individuals differ in their processing capacity. Cognitive load is higher for novices than for experts, because they lack the experience and have not yet developed and stored schemas in long-term memory to ease processing. Knowledge and experience with process models tends to facilitate better and faster comprehension, regardless of the cognitive load.

3 Research Method

While exhaustiveness can never be guaranteed for a literature review (vom Brocke et al. 2015), effort was made to choose criteria for reference selection that would maximize the comprehensiveness of the review. The following sections describe how the literature search was conducted and the references were selected.

3.1 Primary Search of English Literature

We collected a base of articles on the comprehension of process models from three sources: bibliographic databases; a forward search with Google Scholar, a citation-indexing service; and two review articles on process model comprehension from Houy et al. (2012, 2014).

3.1.1 Bibliographic Databases

Ending in May 2016, our systematic literature search used seven bibliographic databases (EBSCO Host, ProQuest, ISI Web of Science Core Collection, ScienceDirect, ACM Digital Library, IEEE Xplore Digital Library) and Google Scholar, a citation-indexing service, guided by four search criteria:

- Search fields: title, abstract, key words (metadata, anywhere except full text).
- Search string: ("quality" OR understand* OR "readability") AND Title = ("process") AND Title = (-model* OR representation* OR diagram*).
- Document types = conference publications, journals articles, books.
- Timespan = none.

The search string was adapted based on the database because only in some databases was it possible to limit search fields, topics (e.g., process models), or research areas (e.g., computer science or business economics). We included not only journal articles but also conference papers published in reputable conference proceedings because they are recommended as source material in the IS field (Webster and Watson 2002). We limited the literature selection to sources published in English.

This search yielded 2666 papers, and a manual scan for relevance performed by viewing titles and abstracts reduced the total to 137 articles. Eliminating duplicates resulted in a total of 108 articles. Then, we reviewed the 108 articles in close detail to determine whether they fulfilled the article selection criteria, as described below.

3.1.2 Forward Search

For each empirical article we selected that measures process model comprehension objectively – plus a few more, which were later discarded in the final selection because of missing details and other reasons – we conducted a forward search on Google Scholar of current works (“cited by”) to account for the most recent papers. We repeated the forward search for each empirical paper that was identified, performing forward search for a total of fifty articles in June 2016. There were as few as 0 and as many as 251 citing papers (mean = 40.80, median = 21.50) for the initially selected papers. By adding all of the references we found into a Google library, we avoided repeated screening of articles. Taken together, we scanned 1050 articles by viewing titles and abstracts and reading the paper if a decision could not be made on basis of the abstract to determine whether they fulfilled our selection criteria. After duplicates were eliminated, this search yielded an additional 79 articles.

3.1.3 Prior Review Articles

We cross-checked the references in Houy et al. (2012, 2014), which discuss how 42 articles measure conceptual model comprehension and investigate the theoretical foundations of 126 articles on model comprehension. Based on these two articles, we added 92 articles to the initial set.

3.2 Selection Criteria for Type of Process Model and Visualization

We excluded all studies that did not investigate visual, procedural process models as research objects. Although some general principles may apply to all conceptual models, specific frameworks for the quality of the various types of models (e.g., data models, process models) are needed because of fundamental differences between the types of models (Moody 2005). Therefore, we removed from consideration any articles that investigate the comprehension of conceptual models other than process models. For instance, among the discarded articles were graph drawings and ER diagrams. We included UML activity diagrams (UML AD) but discarded UML sequence, class, interaction and statechart diagrams.

We focus on *procedural* process models because they follow the same underlying representation paradigm. An increasing number of studies also investigate *declarative* process models (e.g., Haisjackl and Zugal 2014; Haisjackl et al. 2016; Zugal et al. 2015). In comparison to *procedural*

(or imperative) process models, which specify all possible alternatives for execution, *declarative* process models focus on modeling the constraints that prevent undesired alternatives for execution (Fahland et al. 2009). While articles on procedural and declarative process models share a discussion of similar constructs (e.g., comprehension of parallelism or exclusivity of process paths), the extensive differences in visual representation render these articles unusable for comparing study results.

One characteristic of the visualization of process models that contrasts with the characteristics of other conceptual models is their representation as node-link diagrams. Some studies on comparable representations (e.g., flowcharts) share this basic visualization paradigm of process models, so it made sense to include them in the review even though these studies did not use the term “process model.”

3.3 Selection Criteria for Articles

Our review contains articles that offer three types of contributions on process model comprehension:

- empirical studies that measure the comprehension of process models *objectively*.
- empirical studies that measure user preferences and the comprehension of process models *subjectively*.
- “theoretical” discussions on the comprehension of process models.

3.3.1 Empirical Studies that Measure the Comprehension of Process Models and User Preferences

We focus on empirical studies (experiments, questionnaire studies) with process models as their research objects, humans as participants, and comprehension as a dependent variable. Similar to the selection criteria Chen and Yu (2000) use, we checked every study for fulfillment of several criteria:

- Experimental design with
 - At least one experimental condition with a “visual process model,” as defined above.
 - At least one dependent variable on model comprehension that measures comprehension either
 - *objectively* (e.g., a multiple-choice test with correct/incorrect answers) OR
 - *subjectively* or by measuring user preferences (e.g., questionnaire scales like perceived ease of understanding, perceived usefulness, and preference ratings).

- A sufficient level of detail of results reported.

3.3.2 “Theoretical” Discussions on the Comprehension of Process Models

The literature analysis revealed a large number of articles that deal with process model comprehension (e.g., modeling guidelines) but do not present a study that measures model comprehension. These articles are also useful as a theoretical lens through which to draw a comprehensive map of what is known and what is not in the field, to build a framework for reviewing empirical research, and to uncover inconsistencies and gaps in the research. While these articles are diverse in nature, the fulfillment of three criteria was required if they were to be considered eligible as articles that offer “theoretical” discussions of process model comprehension:

- Identification of independent variables that may affect comprehension.
- Relationship to the comprehension of procedural process models.
- Sufficient level of detail.

The articles’ relationship to the comprehension of procedural process models (e.g., adapting theories of the overall field of conceptual model comprehension research to the specific field of process models) was an important criterion. While Moody’s (2009) seminal work on designing modeling notations, for instance, is highly cited, we included any article that introduces these design principles to process modeling (Figl et al. 2009, 2010; Genon et al. 2010).

3.4 Final Selection of Literature

Based on the initial search, we screened and read in detail 279 articles, choosing 76 papers (27%): 38 (50%) that fulfill the criteria for a study that objectively measures the comprehension of process models, 7 (9%) that fulfill the criteria for measuring subjective comprehension and user preferences, and 31 (41%) that fulfill the criteria for offering a “theoretical” discussion. Table 1 lists the number of articles we found for each category based on where we found it. Literature databases were the primary source, and Google Scholar was the secondary source.

Of the 203 articles that were not selected, 79 (39%) were not closely related to model comprehension, 64 (32%) did not address procedural process models, 27 (13%) were related to active modeling instead of model comprehension, 8 (4%) reported too little detail (e.g., no details on the tasks used to measure objective comprehension empirically), 5 (2%) that were conference versions of a journal paper published later, 11 (5%) that mentioned no

Table 1 Chosen articles by source

Contribution related to the comprehension of process models	Literature databases	Forward search (Google cited by)	Houy et al. (2012, 2014)	Total
Studies that measure objective comprehension				
Count	16	19	3	38
%	15%	24%	3%	14%
Studies that measure subjective comprehension or user preferences				
Count	3	4	0	7
%	3%	5%	0%	2%
“Theoretical” discussions of the factors that influence the comprehension of process models				
Count	12	14	5	31
%	11%	18%	6%	11%
Not chosen				
Count	77	42	84	203
%	71%	53%	91%	73%
Total				
Count	108	79	92	279

independent variable of interest (e.g., evaluating a tool or evaluating a single notation without a reference value), 5 (2%) that mentioned no dependent variable of interest (e.g., articles that measure only comprehension time and not comprehension accuracy), 3 (1%) that dealt with a modeling tool and 1 (2%) (Moher et al. 1993) that we could access only in part.

Based on personalized Google Scholar updates, two additional articles about studies that measure model comprehension objectively were added in October 2016, leading to a final size of 40 articles of this type.

3.5 Search of German Literature

As this field of research seems to be particularly prevalent in German-speaking areas – 80% of selected articles have at least one author who was employed by or had graduated from a German-speaking university – we performed an additional literature search in German. Repeating the literature search in the databases did not deliver adequate results with German search terms, so in September 2016 we followed three strategies to account for literature published in German:

- We scanned all sixty German references that were cited in the final list of selected articles.
- We searched in the Karlsruhe Virtual Catalog KVK (meta search for Germany, Austria, Switzerland) for combinations of the search terms “Prozessmodell*/Prozessdiagramm*/Geschäftsprozessmodell*” with the search terms “verständlich*/lesbar*” (774 search results).

- We searched the proceedings of the major German conference series “Wirtschaftsinformatik” in the AIS Electronic Library (139 search results) and screened the titles and abstracts of sixty-one German issues of the journal “Wirtschaftsinformatik/BISE” (1999–2008) in SpringerLink online.

Based on this literature search, we identified one reference in the German-language literature that fulfilled all of the criteria for offering a relevant theoretical discussion. The article describes the “clarity” aspect (including the goal of comprehensibility) of the “guidelines of modeling” (GOM) (Becker et al. 1995) in relation to process modeling. Therefore, there were thirty-two theoretical articles in the final sample.

3.6 Coding

We first coded the forty-seven empirical studies manually using coding tables in Excel, and later imported the coding tables to SPSS for further analysis. They are reproduced in a shortened version in Online Appendices B and C.

We selected a concept-centric approach with which to structure our descriptive literature review (Webster and Watson 2002). The first coding table is study-based, so each line represents an article and the study it describes, as none of the articles present more than one study (see Table 4 in Online Appendix B, available via springerlink.com). The second coding table is variable-based: each line represents an independent variable for which its effect on model comprehension and/or user preference is reported by a study (see Table 5 in Online Appendix C). The main concepts in our context are independent variables that cause variation in the dependent variables. For all empirical studies, measurement of variables and statistical results for main, relevant effects on model comprehension are reproduced in detail. We analyzed and compared the design, the participants, analysis methods, and publication outlets in detail.

Unfortunately, the statistics reported in many studies are neither sufficiently detailed to calculate effect sizes in order to combine findings in a meta-analysis nor are p-values consistently reported, which would be a requirement for using vote-counting formulas (King and He 2005). Therefore, we inductively developed a coding schema for the “level of evidence” based on the articles’ reporting of statistical results. These evidence ratings are meant to be interpreted only in relation to each other for this selection of empirical studies. Table 3 (in Online Appendix A) gives an overview of the categories, which we developed based on the result descriptions (the statistical reporting) in the empirical articles. We distinguished among five levels of evidence (no evidence, conflicting evidence, weak evidence,

moderate evidence, and strong evidence), so the variable-based overview in Online Appendix C provides not only a descriptive summary of the direction and significance of an effect of an independent variable on comprehension, but also provides an evidence rating for the effect.

In addition, we characterized sample sizes in relation to each other by dividing them into quartiles (small, medium, large, very large), as detailed in Table 2 (in Online Appendix A). The two indicators – level of evidence and quartile of sample size – are used to ease comparison of the studies’ results.

Table 6 (in Online Appendix D) provides an overview of all independent variables that are identified in the theoretical discussions, relevant to the comprehension of process models, and investigated in the studies. We derived the related influence factors inductively from papers that offer theoretical discussions and assigned category labels to the main thematic areas, as is done in a qualitative content analysis (Mayring 2003). We sorted all variables that the empirical studies include according to categories. This tabular representation allows us to tie together all of the variables that have been reviewed and to discuss differences among the key variables addressed in theoretical and empirical work. Table 7 (in Online Appendix D) is a condensed version of Table 6 (in Online Appendix D). We used this categorization to derive a framework for independent variables and to organize and classify the empirical material, as presented in Sect. 4.1 below.

4 Results

4.1 A Framework for Independent Variables

While existing research reports empirical results on various factors that influence the comprehension of process models, these insights remain scattered across multiple studies and articles. To categorize the main types of independent variables, we build on Mendling and Strembeck (2008), who distinguish personal factors, model factors, and content factors and add four other dimensions, so every variable in the empirical studies fits into one category: presentation medium, notation, secondary notation, characteristics of the process models, labels, the users, and the types of comprehension tasks (see Fig. 1 for a visual illustration of all types of variables except the presentation medium type).

We also need a lower-level categorization of influence factors in order to compare study results with each other. Table 2, a shortened overview of Tables 6 and 7 in Online Appendix D, provides a framework for the main categories and subcategories of influence factors. The factors can be categorized in the main categories as well as according to

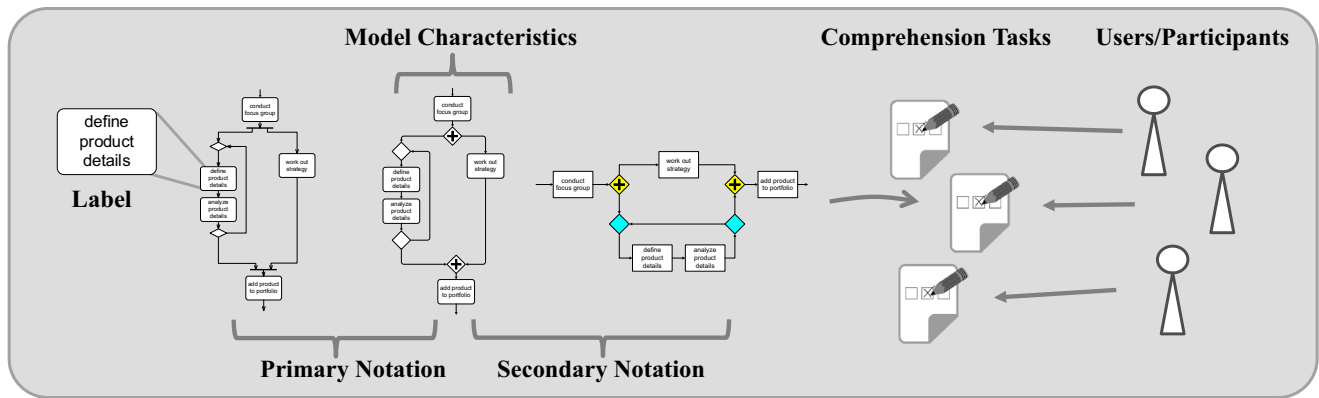


Fig. 1 Types of influence factors for process model comprehension

Table 2 Framework of relevant influence factors based on articles (italicized terms are used as categories)

Influence factors for process model comprehension (independent variables)		Cognitive load
Presentation medium	<i>Presentation medium</i> (paper versus computer) [empirical studies only]	Extraneous
Notation	<p><i>Representation paradigm</i> (e.g., text versus model, differing dialects and cognitive fit, declarative versus imperative process models, assigning domain semantic-oriented pictorial elements like icons and images to modeling elements, animation, narration and visualization techniques)</p> <p><i>Primary notation</i> (e.g., BPMN, UML AD, BPMN, vBPMN, YAWL, C-YAWL, EPCs, configurable EPCs, SBD)</p> <p><i>Notational characteristics</i> (e.g., semiotic clarity, perceptual discriminability, semantic transparency, visual expressiveness, graphic economy)</p>	
Secondary notation	<p><i>Decomposition</i> (use of decomposition/modularization, decomposition heuristics)</p> <p><i>Gestalt theory</i> (dual coding, highlighting, like using colors for control blocks)</p> <p><i>Layout</i> (edges like crossing edges, direction, shape and size, symmetry, alignment of elements and spacing, ending points)</p>	
Label	<p><i>Label design</i></p> <p><i>Naming conventions</i> (syntactic like using a verb-object label style for activities, semantic like using a domain-specific vocabulary, avoidance of synonyms and homonyms)</p>	Intrinsic
Model characteristics	<p><i>Size measures</i> (number of activities, events, gateways, diameter)</p> <p><i>Modularity</i> and <i>block structuredness</i> (corresponding split and join elements) and related metrics (separability, maximum nesting depth)</p> <p><i>Refactoring</i> (simplification without changing the process’s behavior) [theoretical discussions only]</p> <p>Removing <i>redundant elements</i> [theoretical discussions only]</p> <p><i>Gateway interplay/control structures</i> (e.g., XOR, cycles, OR, AND, concurrency) and related metrics (control flow complexity, sequentiality, cycles, heterogeneity of gateway types)</p> <p><i>Connection</i></p> <p><i>Syntax rules</i></p>	
Task	<i>Task</i> [empirical studies only]	
User	<p><i>Tailoring</i> of process models for personal factors [theoretical discussions only]</p> <p><i>Domain knowledge</i></p> <p><i>Experience and familiarity with modeling</i></p> <p><i>Modeling knowledge</i></p> <p><i>Education</i></p> <p><i>User characteristics</i></p>	

Table 3 Variable type and research design

	Type of independent variable							Research design (per variable)		
	Presentation medium	Notation	Secondary notation	Label	Model	Task	User	Between-subjects	Within-subjects	Mixed
Number of studies	3	21	9	5	11	10	21	32	21	10
% of studies	6%	45%	19%	11%	23%	21%	45%	68%	45%	21%

Table 4 Combinations of variable type and research design

	Type of independent variable							Total
	Presentation medium	Notation	Secondary notation	Label	Model	Task	User	
Research design (per variable)								
Between-subjects	2	15	6	3	0	4	43	73
Within-subjects	0	10	2	3	36	11	0	62
Mixed	1	2	2	1	3	1	2	12
Total	3	27	10	7	39	16	45	147

whether they add extraneous or intrinsic cognitive load to a comprehension task.

This framework allows us to gain systematic insight into existing empirical findings and to structure the discussion and summarization of all articles. Section 4.3 uses the framework to capture the current research status and synthesize empirical findings according to the similarity of the variables investigated. Then Sect. 4.4 maps the theoretical discussions of relevant influence factors to the results of the studies that measure comprehension in order to identify gaps to be addressed by future research.

4.2 Characteristics of Empirical Articles on Comprehension of Process Models

This section provides an overview of the characteristics of the forty empirical articles that measure objective process model comprehension and the seven articles that measure subjective comprehension and user preferences.

4.2.1 Independent Variables and Research Designs

The selected studies report results for up to 12 independent variables (median = 2.00, mean = 3.13, SD = 2.46), for a total of 147 independent variables for which influence on comprehension of process models was assessed. However, it is likely that the number of independent variables that the extant articles initially gathered or calculated is even higher, as we observed a gap in some studies between a higher number of variables (e.g., model metrics or control variables that were collected in the questionnaire) mentioned at the beginning of the article and a lower number of

variables for which results were reported. Perhaps in some cases the authors include only significant influence factors in the final statistical analyses. Table 3, which lists the number of studies that investigate a specific type of variable (based on the main categories of influence factors) and the research design, shows that almost half of the studies include notation (21, 45%) and/or user characteristics (21, 45%) as influence factors. Approximately a fifth of studies take either model-related variables (11, 23%) or task-related variables (10, 21%) into account. The studies investigate nine variables (19%) related to secondary notation and five (11%) related to labels. Sixty-eight percent of the studies used a between-subjects design, 45% a within-subjects design, and 21% used a mixed design to investigate the effect of any independent variables. When variables were investigated in “mixed” designs, researchers typically used counterbalanced designs in which group C for instance receives model A in version A and model B in version B, while group D receives model A in version B and model B in version A. The percentages in Table 3 do not add up to 100% because most studies consider more than one independent variable.

Table 4 shows the raw values of variables for all studies. User-related variables are typically between-subject variables and are the only variables that the researchers could not manipulate. (Some authors consider this type of variable an independent variable, and others consider it a control variable.) The only exception was the variable of domain-specific knowledge, which is additionally investigated as a within-subject factor if models in different domains are part of the experiment. Model- and task-related variables are typically realized as within-subject

factors, and (secondary) notation-related variables as between-subject factors.

4.2.2 Participants

The unit of analysis in the studies is typically the individual, but some studies use the process models themselves and their characteristics or labels as units on which to base statistical analyses.

In most cases, students are the participants used in the studies (29, or 62%), but domain experts (6, or 13%), process model experts from academia and practice (2, or 4%), and mixed participants' groups (10, or 21%) are also used.

4.2.3 Statistical Analysis Methods

The analysis methods used in the studies span a variety of statistical methods, with ANOVA (34%), regression analysis (15%), Spearman's correlation analysis (10%), and Pearson's correlation analysis (8%) used most often to analyze variables (see Table 1 in Online Appendix A for details.) Four studies take more than one statistical approach to investigate the relationships between the independent and dependent variables.

4.2.4 Publication Outlets

The publication outlets with more than one study included journals like *Decision Support Systems* (5 studies), *Journal of the Association for Information Systems* (3 studies), *Information Systems* (3 studies), *Communications of the Association for Information Systems* (2 studies), and *Information and Software Technology* (2 studies) and conferences like Business Process Management (BPM) and its workshops (6 studies) and the Conference on Advanced Information Systems Engineering (CAISE, 3 studies).

4.2.5 Measurement of the Dependent Variables

In general, measuring model comprehension is difficult because the outcome is “tacit understanding created in the model viewers' cognition” (Gemino and Wand 2004, p. 251). Houy et al. (2012) inductively develop a categorization of comprehension measures for conceptual models based on a literature review and distinguish among five objective measures (four measures related to effectiveness – “recalling model content,” “correctly answering questions about model content,” “problem solving based on the model content,” and “verification of model content,” – and one related to efficiency – “time needed to understand a model”), and one subjective measure (perceived ease of understanding a model). Gemino and Wand (2004) also

mention confidence in correctness of comprehension or recall tasks as a subjective measure.

Based on prior categorizations of comprehension measures, we analyzed the studies and report their results in terms of two objective indicators – comprehension accuracy and time taken – and a category for subjective measures:

- Objective comprehension accuracy: measured using comprehension questions; synonyms used in studies include (task) performance, accuracy, percentage of correct answers, interpretational fidelity, solution percentage, (comprehension) effectiveness, objective difficulty.
- Time taken: synonyms used in studies include speed, comprehension efficiency (referring to time taken), task completion time.
- Subjective comprehension difficulty and user preference: for example, the ease of understanding a model, perceived difficulty, subjective/perceived cognitive load, subjective difficulty of control flow comprehension. In contrast to prior categorizations of subjective comprehension measures, this category has a wider focus, as we included dependent variables like preferences and perceived usability. The studies use differing measurement scales for subjective comprehension and preferences, so there was no common ground on which to compare results in the “subjective” category directly.

Our primary interest lies in comprehension questions, which are characterized as surface-level understanding, that is, “the person's competence in understanding the constructs of the modeling formalism” (Moody 2004, p. 135), in contrast to deep-level understanding, which refers to applying domain understanding. We focus on surface-level understanding for two reasons: First, surface-level measurement with comprehension tasks is the most common measure, making it easier to compare results, and second, this measurement is more directly related to concrete models than is measurement via recall or problem-solving tasks, so relationships to independent variables should be strongest.

According to Houy et al.'s (2012) analysis of forty-two studies on the comprehension of conceptual models, recall of model content was measured only three times. Similarly, in our selection of empirical articles, in most cases the participants had the models available while performing the tasks. The situation differs in the area of data modeling; Parsons and Cole (2005) find that researchers in seven of thirteen studies took the models away before participants performed their tasks. The difference may be related to the stronger focus in the data-modeling area on measuring not only comprehension of a diagram but also the domain

understanding the model's user acquires (Gemino and Wand 2004).

All empirical studies that measure objective model comprehension include some kind of comprehension task. While many of the comprehension tasks are related to the control flow (e.g., “Can task A be processed more often than task B?”), they may also address other perspectives of the process (e.g., actors, resources, data). In principle, there are endless ways to construct these tasks. For instance, most studies with comprehension tasks on control flow include only a few activities, so they relate to a sub-part of the whole process model. The difficulty of comprehension tasks can also vary significantly, so one can draw conclusions concerning which aspects of a process model are difficult to understand. Therefore, researchers use not only comprehension tasks to measure comprehension but also construct them in a way to reveal the effects of specific model characteristics on model comprehension. Based on this observation, we discuss ten out of fifteen task-related variables in the context of model characteristics (see Table 4 in Online Appendix B).

4.3 Empirical Results

4.3.1 Presentation Medium

Turetken et al. (2016) are the first to have compared interactive model visualizations on a website that could be zoomed and navigated and that offered mouse-over pop up for sub-models with printed models. Their study revealed strong evidence that participants find models on paper easier to understand and more useful, perhaps because printed models reduced the effort entailed in information-seeking in their specific setting. However, they found no evidence that the representation medium had a significant effect on comprehension accuracy. This finding is in line with those of two other studies (Mendling et al. 2012b; Recker et al. 2014), both of which use a paper-based questionnaire and an online questionnaire and find no evidence of differing effects on comprehension. Twenty-two (47%) of the studies in our sample present models on paper, eight (17%) in online questionnaires, and four (9%) in modeling tools. Three (6%) studies use more than one presentation medium and ten (21%) did not mention the presentation medium.

4.3.2 Notation

Different notations “tend to emphasize diverse aspects of processes, such as task sequence, resource allocation, communications, and organizational responsibilities” (Soffer and Wand 2007, p. 176). In the context of this article, we follow Moody (2009, p. 756), who defines visual notation as “a set of graphical symbols (visual

vocabulary) [and] a set of compositional rules (visual grammar).” The literature also refers to the difference between notations and models using the terms “grammar” versus “script,” and researchers characterize empirical studies that contrast notations as “intergrammar comparison,” while “intragrammar comparisons” use only one notation and investigate other variables beyond notation (Gemino and Wand 2004).

Most process modeling notations share a basic set of concepts but use divergent symbols to represent them. An important distinction is that between *primary* and *secondary* notation: *Primary* notation defines the symbols and the rules for combining them, while *secondary* notation relates to “things which are not formally part of a notation which are nevertheless used to interpret it... (e.g., reading a ... diagram left-to-right and top-to-bottom, use of locality (i.e., placing logically related items near each other))” (Petre 2006, p. 293). Moody (2009, p. 760) similarly defines *secondary* notation as “the use of visual variables not formally specified in the notation to reinforce or clarify meaning.” Primary and secondary notation have some overlaps since some notations define rules on certain aspects of notation in their formal definitions, which other notations do not. Following our framework, we first discuss the factors that are broadly related to notation and then discuss those that are related to secondary notation.

4.3.2.1 Representation Paradigm First, we summarize studies that challenge the assumption concerning whether using a modeling notation for process descriptions instead of alternative representation paradigms to maximize model comprehension is always the best choice. Ottensooser et al. (2012) finds that process models improve comprehension accuracy more than written use cases and that they increase comprehension accuracy for users trained in process models, but not for users who have no prior training. In line with this result, Rodrigues et al. (2015) report that experienced users perform better in comprehension tasks with BPMN models than with textual process descriptions, while there was no difference among inexperienced users. In summary, there is a moderate level of evidence of the effect of representation in these two studies – that is, that process models are superior to textual descriptions – since the effect was not significant for all user groups. Users also seem to prefer BPMN diagrams over structured text and textual descriptions when the goal is to understand a process (Figl and Recker 2016).

Regarding other types of representation, users report a slight preference for diagrams with icons attached to activity symbols that express the semantic meaning of the process activities over diagrams without icons (Figl and Recker 2016). One other study, with relatively small sample size, shows that procedural (imperative) process

models aid comprehension more than declarative models do (Pichler et al. 2012), perhaps because procedural process models explicate the sequence of activities, while this information is hidden in declarative models. In addition, BPMN3D, a version of BPMN that uses dimensions for data objects is evaluated as best for comprehension, followed by Bubble, which visualizes process tasks as bubbles, and then other uncommon visualization concepts of process models (Hipp et al. 2014).

4.3.2.2 Primary Notation Several studies compare the composite effects of notations. BPMN, EPC, UML AD, and YAWL were investigated in more than one study, so results can be compared to some degree. These studies show that vBPMN, a configuration extension of BPMN, is easier to comprehend than C-YAWL, a configuration extension of YAWL (Döhning et al. 2014). This finding is also reflected in Figl et al. (2013a), who isolate the perceptual discriminability deficiencies of symbols in YAWL and demonstrate that these symbols lowered comprehension accuracy below that of UML AD and BPMN. Sarshar and Loos (2005) find that EPC scores better than Petrinets in helping users understand XOR, but their evidence of a difference in subjective difficulty is weak. Recker and Dreiling (2007, 2011) compare BPMN and EPC and find no evidence of higher comprehension accuracy of BPMN, although participants performed better with BPMN when the task was to fill in missing words in a cloze test about the process. Two other studies, each of which reports only descriptive statistics, thereby offering only weak evidence, compare EPC to other notations. Sandkuhl and Wiebring (2015) award eEPC – that is, extended EPC using additional symbols – the highest absolute score in subjective perception of notation and the second-highest score in comprehension, following UML AD; however, Weitlaner et al. (2013) find that eEPC is less well understood than either UML AD or BPMN. Two additional studies offer moderate evidence of the lower comprehension accuracy of EPC in comparison to other notations: In the study by Jošt et al. (2016) UML AD statistically significantly outperformed both EPC and BPMN in some cases, although results were not consistent and notations were not presented in consistent flow directions, compromising the study’s validity. This result is in line with Figl et al. (2013a), who show that semiotic clarity deficiencies in EPC reduce comprehension accuracy below that of UML AD and BPMN.

Results on communication-oriented flow diagrams and functional flowcharts are mixed. One study reports that communication-oriented flow diagrams have higher perceived ease of understanding (Kock et al. 2009), but this result is not confirmed in a second study (Kock et al. 2008).

Several notations have been investigated only in a single study, so results can only provide an overview of the

notations that have been the subjects of empirical evaluation. Natschläger (2011) provide weak evidence in the form of descriptive statistics for higher comprehension of “deontic” BPMN, which expresses deontic logic like obligations, alternatives and permissions, in comparison to regular BPMN. Similarly, Recker et al. (2005) find weak evidence without using statistical tests that configurable EPCs (c-EPCs) are perceived as more useful than standard EPCs. In Stitzlein et al. (2013), the domain-specific health process notation (HPN) outperforms BPMN in complex tasks, but the effect is reversed for simple tasks.

4.3.2.3 Notational Characteristics Some studies investigate notational characteristics like aesthetic symbol design, semantic transparency, perceptual discriminability, and the use of gateways¹ as isolated factors, so their results can be generalized beyond specific notations. These studies do not always adhere to exact syntactic restrictions of modeling notations in their experimental material but focus instead on varying specific notational characteristics. This kind of isolation makes it easier to achieve internal validity and to determine what causes an effect on model comprehension. When comparing notations as a whole, models differ based on many variables (e.g., different numbers of symbols), so it is difficult to suggest how to improve a notation.

For instance, Recker (2013) demonstrates that the use of gateway constructs benefits understanding and explains this effect as resulting from higher perceptual discriminability. Perceptual pop-out and discriminability show their relevance for comprehension accuracy and perceived cognitive load in (Figl et al. 2013a, b). In contrast, characteristics like semantic transparency and aesthetics, which relate to later stages of perceptual processing, affect perceived cognitive load but not comprehension accuracy directly (Figl et al. 2013a, b).

Kock et al. (2009) report that subjectively experienced “ease of generating the models” in a notation is also positively associated with model comprehension.

4.3.3 Secondary Notation

4.3.3.1 Decomposition Reijers et al. (2011b, p. 10) show that modularization is positively related to model comprehension, explaining that modularization in large models “shields the reader from unnecessary information.” In contrast, Turetken et al. (2016) report strong evidence of higher comprehension accuracy for fully flattened models in local comprehension tasks that can be completed based on information in a sub-model. The type of comprehension

¹ We use the term “gateway” to ease comprehension, although some authors of papers cited herein use “connector”.

task seems central to the investigation of decomposition, as no evidence of differences in other tasks have been found. In addition, Johannsen et al. (2014) find that low levels of violated decomposition principles, as described in Wand and Weber's decomposition model (Wand and Weber 1995), positively influence comprehension.

4.3.3.2 Gestalt Theory Another stream of research on secondary notation is concerned with how to incorporate Gestalt theory in model design to make it easier for humans to recognize related elements as belonging together. Gestalt theory deals with principles associated with how humans perceive whole figures instead of simpler, separate elements (Wagemans et al. 2012). Several visual variables can be employed in this context; studies have indirectly investigated the principle of "common region" (Palmer 1992) by researching the use of swim lanes or visual grouping of sub-processes and the principle of "similarity" using colors and syntax highlighting. Although, according to the principle of "common region," swim lanes are hypothesized to benefit comprehension, Bera (2012) removes the lanes in the "no swim lane" condition and finds evidence of an effect on time taken and performance in problem-solving but none on comprehension. Jeyaraj and Sauter (2014) investigate a "no swim lane" condition in which actors are completely cut from the side of the diagram and redundantly inserted in each activity symbol and find no evidence of a cognitive advantage of swimlanes for process models. In fact, some items on external actors were easier to answer in the "no swim lane" condition.

Turetken et al. (2016) use an experimental condition in which sub-processes are visually grouped in a common region by means of background colors but find that this representation does not significantly enhance comprehension accuracy. On the contrary, the version without visual grouping was rated as easier to understand. Perhaps this result was due to participants' not needing information about how the process was separated into sub-processes to answer the comprehension tasks.

Color coding with bright colors lowers perceived difficulty for participants from a Confucian culture, but not for participants from a Germanic culture; this result may be related to a culture-dependent preference for bright colors in Asia, although color in general is not related to model comprehension in this study (Kummer et al. 2016). In contrast, Reijers et al. (2011a) find that syntax highlighting with colors for matching gateway pairs is positively related to novices' model comprehension, as it helped them identify relevant patterns of matching gateways, but the study finds no evidence that syntax highlighting improves experts' performance. The inconsistency between the findings of Kummer et al. (2016) and Reijers et al. (2011a) may occur because of differences in the kinds of

information that the studies highlight with color. For instance, Kummer et al. (2016) color activity symbols, while Reijers et al. (2011a) color-code gateway pairs that belong to each other, information that is more relevant to comprehension tasks. Petrusel et al. (2016) use task-specific color highlighting of regions of model elements that are relevant to a comprehension task and find no evidence of reduced comprehension accuracy, although color highlighting lowers mental effort (measured by fixations and fixation durations in eye-tracking results) and time taken.

4.3.4 Layout

Only two of the selected empirical studies conduct empirical evaluations of explicating factors related to model layout. Figl and Strembeck (2015) investigate flow direction in models and find no evidence of the hypothesized superiority of left-to-right flow direction. The authors speculate that this result may be explained in part by humans' ability to adapt quickly to uncommon reading directions (e.g., right-to-left). Petrusel et al. (2016) investigate a task-specific layout of relevant model elements, which were made larger and were repositioned, and find no evidence of an effect on comprehension accuracy, although mental effort, measured by eye-tracking, is reduced.

4.3.5 Label Characteristics

Empirical studies that are related to labels focus on labels for activities. Mendling et al. (2012b) reveal strong evidence that comprehension accuracy is higher for abstract labels like letters than it is for concrete textual labels. One explanation for this result is that process models with abstract labels offer no semantic content, so information processing can focus on understanding the syntactical model structure. However, two studies with smaller sample sizes find no significant effect of abstract versus concrete labels (Figl and Strembeck 2015; Mendling and Strembeck 2008). Research also reports that, the longer the labels, the lower the comprehension accuracy, perhaps because of the increased effort required to find and read longer labels (Mendling and Strembeck 2008).

Mendling et al. (2010b) find that the verb-object label style is rated highest in perceived usefulness (followed by the action-noun label style), as this style is least ambiguous in, for example, helping the user to infer the type of action required in a process task. The lower ambiguity of this kind of label is also positively related to perceived usefulness of label (Mendling et al. 2010b). Users also rate linguistically revised labels as easier to understand than unrevised labels in the case of non-domain-specific vocabulary, although the opposite effect occurs in domain-specific vocabulary,

perhaps because of the use of a standard language dictionary instead of a domain ontology in the experiment (Koschmider et al. 2015b).

4.3.6 Process Model Characteristics

Researchers use a variety of metrics to measure and operationalize the structural complexity and properties of process models (Mendling 2013). Aguilar et al. (2008) distinguish between “base” measures, which count the business process model’s most significant elements, and “derived” measures, which provide the proportions between a model’s elements. Combined metrics like the control-flow-complexity measure are also used (Cardoso 2006).

Mendling et al. (2012a) categorizes process metrics into five categories: size measures, connection, modularity, gateway interplay, and complex behavior. We build on this categorization and integrate it with five other terms mentioned in our selected articles: size measures, connection, modularity/structuredness, gateway interplay/control structures, and syntax rules.

The empirical studies use two main approaches to measuring the effect of process models’ characteristics: relating global model metrics to model comprehension and relating the difficulty of the comprehension task (e.g., comprehension tasks that consider control structures as loops or concurrency) to model structures.

4.3.6.1 Size Measures Size measures relate to the number of elements, including arcs, gateway nodes, event nodes, and task nodes (Mendling et al. 2012a).

Two studies with large sample sizes find that model size operationalized as the number of nodes is negatively related to model comprehension accuracy (Recker 2013; Sánchez-González et al. 2010), but Mendling and Strembeck (2008) report no effect of the number of nodes on comprehension accuracy. Sánchez-González et al. (2010) find that another measure of model size, the length of the longest path from a start node to an end node, lowers comprehension, while Mendling and Strembeck (2008) find no evidence of such an effect. Relationships between higher size and lower comprehension accuracy are in line with cognitive load theory, as these metrics reflect higher intrinsic cognitive load. Mendling and Strembeck’s (2008) non-significant results may also be due to a lower variance in the numbers of nodes and diameters in their set of models, as they do not mention a systematic variation of these variables, and all models fit on A4-size pages. Another possible interpretation is that only a few elements are relevant for each comprehension task, not the whole model, so the choice of comprehension tasks could influence the results.

Aguilar et al. (2008) report weak evidence of negative correlations between model comprehension and metrics that are related to events (number of events, intermediate events, and number of sequence flows from events). There was further weak evidence for the relevance of the metrics of number of gateways (Sánchez-González et al. 2012) and exclusive decisions (Aguilar et al. 2008), while the number of OR joins (Reijers and Mendling 2011) did not have a significant effect.

4.3.6.2 Modularity/Structuredness Several studies investigate the metrics related to models’ structuredness. Structuredness denotes that, “for every node with multiple outgoing arcs (a split) there is a corresponding node with multiple incoming arcs (a join) such that the subgraph between the split and the join forms a single-entry-single exit (SESE) region” (Dumas et al. 2012, p. 33). Mendling and Strembeck (2008, p. 147), who calculate the metric as “one minus the number of nodes in structured blocks divided by the number of nodes,” find no significant correlation between it and comprehension accuracy. Dumas et al. (2012) find that a comparable metric results in inconsistent results, as structuredness reduces comprehension for two models and heightens comprehension in two other models, which leads the authors to conclude that structuring might be beneficial only when it does not increase the number of gateways.

Research has also identified three other metrics related to model structure of the model as potentially relevant factors. First, Sánchez-González et al. (2010) find that the maximum nesting of structured blocks in a process model is a negative influence factor, which finding is supported by research showing that interactivity of the process activities included in a comprehension task, measured via the process structure’s tree distance metric, is negatively related to comprehension (Figl and Laue 2015). Second, Mendling and Strembeck (2008) find strong evidence that the relationship between a model’s comprehensibility and its cut-vertices, the absence of which would separate the model into two parts, is such that higher separability as a global metric is associated with higher overall comprehension accuracy. However, experiments that measure the presence of cut-vertices for each comprehension task separately find no evidence of such an effect (Figl and Laue 2015). Third, there is weak evidence of the ability of the degree to which a model is constructed out of pure sequences of tasks to heighten comprehensibility (Sánchez-González et al. 2010).

4.3.6.3 Gateway Interplay/Control Structures Many studies investigate control structures using measurements beyond simple counting of gateway elements. Studies find weak evidence of the relevance of the metrics of control-

flow complexity (Sánchez-González et al. 2012) and sequence flows looping (Aguilar et al. 2008), but no significant effect of concurrency (Mendling and Strembeck 2008). We also rated as weak the evidence of a negative effect on the comprehensibility of gateway mismatch (measured as the sum of gateway pairs that do not match with each other, such as when an AND-split is followed up by an XOR-join), based on Sánchez-González et al. (2010, 2012). Reijers and Mendling (2011) report a non-significant correlation between gateway mismatch and model comprehension. The frequency with which different types of gateways are used in a model (gateway heterogeneity) is investigated in four studies, with two finding no evidence of a correlation (Reijers and Mendling 2011; Mendling and Strembeck 2008), one finding weak evidence of a correlation (Sánchez-González et al. 2012), and the fourth reporting strong evidence of a significant correlation (Sánchez-González et al. 2010).

Seven studies investigate comprehension of control structures by comparing the difficulty of comprehension tasks. At first view, results seem contradictory, as some authors report strong evidence that order/sequence tasks are easiest and repetition/loops tasks are most difficult (Figl and Laue 2011, 2015; Laue and Gadatsch 2011), while Melcher et al. (2010; Melcher and Seese 2008) find that order tasks are most difficult and repetition tasks are easier. While the counterintuitive result from Melcher and Seese (2008) could be explained by the low number of participants (9) for the order task, the second study's sample was large, and the authors (Melcher et al. 2010) themselves note that order's being the most difficult is "not directly intuitive" and that it is the only normally distributed variable in their study. According to Weitlaner et al. (2013), order and repetition are easier to comprehend than concurrency, but they present only descriptive statistics. Another study reveals that tasks related to "OR" routing symbols are more difficult than those related to "AND" and "XOR" (Sarshar and Loos 2005), without giving exact numbers. It is not possible to assess whether different studies' "rankings" of control structures differ statistically, but the differences among studies may also be related to question wording as an influence factor in comprehension accuracy (Laue and Gadatsch 2011). Drawing upon these insights, Figl and Laue (2015) base their analysis on a qualitative coding of the control-flow patterns that must be understood in order to answer a comprehension task correctly, rather than on the task's wording. Their study finds that order is easiest and repetition most difficult in terms of comprehension accuracy, while concurrency and exclusiveness patterns have a medium level of difficulty.

4.3.6.4 Connection Some studies look at metrics that relate routing paths (arcs) to model elements. A higher average

number of a gateway node's incoming and outgoing arcs and a higher ratio between the number of arcs in a process model and the theoretically maximum number of arcs both have a negative effect on comprehension accuracy (Sánchez-González et al. 2010, 2012). These results are supported by Reijers and Mendling (2011, p. 9), who comment that "the two factors which most convincingly relate to model understandability both relate to the number of connections in a process model, rather than, for example, the generated state space." There is additional weak evidence that a decision node's higher maximum number of incoming and outgoing arcs is related to lower comprehension, but the extent to which all nodes in a model are connected to each other is not reported to be a significant influence factor (Reijers and Mendling 2011).

4.3.6.5 Syntax Rules Only one study, Heggset et al. (2015), has been published on syntax rules. The authors report that improving models by means of syntactic guidelines increases comprehension accuracy. As the guidelines include more than twenty steps, such as those related to symbol choice and routing behavior, and the study reports results for only one model, it is not possible to derive from this study which syntactic guidelines are especially important.

Mendling and Strembeck (2008) also examine soundness in EPCs, which can be violated by, for instance, incorrect insertion of OR-joins, but find no evidence for a relationship between such soundness and comprehension.

4.3.7 Task Characteristics

The tasks characteristics that capture to which control structures and elements in a model a task is related were discussed in the respective sections about model characteristics.

Pichler et al. (2012) report strong evidence that comprehension accuracy is higher for sequential tasks that relate to local parts of the models (e.g., how input conditions lead to a certain outcome) than it is for more global circumstantial tasks (e.g., what combination of circumstances will lead to a particular outcome). Soffer et al. (2015) find that providing a catalog of routing possibilities is helpful in increasing comprehension accuracy because of the direct availability of cases in the catalog and reduced need for cognitive integration. In addition, invalid statements on a process model are easier to identify than valid statements (Figl and Laue 2015), probably because only one falsifying argument must be found to complete such a comprehension task correctly.

4.3.8 User Characteristics

Modeler expertise – that is, the "required skills, knowledge and experience the modeler ought to have" (Bandara

et al. 2005, p. 353) – is an individual-level variable that affects the success of process modeling. Petre (1995, p. 34) claims that “experts ‘see’ differently and use different strategies.” Against this background, it is not surprising that several studies include user characteristics ranging from general education, individual cognitive abilities, and styles to experience with process models as independent variables.

4.3.8.1 Domain Knowledge Studies have found no evidence of the effect of domain knowledge on model comprehension (Bera 2012; Turetken et al. 2016; Recker and Dreiling 2007; Recker et al. 2014). There are several possible explanations for this influence factor’s lack of statistical significance. For example, researchers have held domain knowledge constant and have chosen homogenous groups of participants (e.g., students) to avoid bias from high levels of familiarity with the domain in their experiments (e.g., Recker and Dreiling 2007), making an effect on model comprehension unlikely because of the low variation in the variable. Studies have also used self-reported scales to measure “perceived” domain knowledge, which might not be able to capture knowledge in the domain, or the “domain-specificity” of the models could have been low.

4.3.8.2 Experience and Familiarity with Modeling Measures of modeling experience and familiarity vary significantly in the studies. Categorizing these measures is difficult because some authors distinguish between the place where experience was acquired (university versus practice) and the type of experience (formal training versus use in practice), while others use measures that overlap with these distinctions (e.g., self-rated familiarity with a notation/process modeling in general, self-rated number of processes modeled, knowledge of a notation). Therefore, we distinguish only between self-assessed experience (or familiarity) and modeling knowledge that is measured objectively.

Of the eighteen measures of modeling experience, we categorize four as showing moderate/strong evidence of an effect on model comprehension, while the rest do not report a statistically significant effect. Studies find no evidence of an effect of self-assessment of previous modeling knowledge (Johannsen et al. 2014; Reijers and Mendling 2011; Weitlaner et al. 2013; Recker and Dreiling 2007), duration of involvement with business process modeling (Mendling and Strembeck 2008), self-assessment of process modeling experience (Reijers and Mendling 2011; Turetken et al. 2016), intensity of work with process models (Mendling and Strembeck 2008), or modeling familiarity (Ottenssooser et al. 2012; Recker 2013; Kummer et al. 2016) on model comprehension.

However, the research shows significant positive effects on model comprehension of the frequency of use of “flow charts” (Ottenssooser et al. 2012), process model working experience (Recker and Dreiling 2011), training on modeling basics at a university or a school (Figl et al. 2013a), and model training at different universities (Reijers and Mendling 2011). Mendling et al. (2012b) report an inverse u-shaped curve as describing the relationship between modeling intensity and duration of modeling experience, on one hand, and comprehension accuracy, in which medium intensity/experience was best, on the other. This finding may explain the mixed results for experience-related variables. Another likely explanation for the low amount of evidence regarding experience measures’ effect on comprehension is that researchers are often more interested in keeping the effect of these variables constant in their study design and focusing on other influence factors instead of selecting samples with larger variance in experience and familiarity with process modeling. This explanation is also reflected in Gemino and Wand’s (2004, p. 258) warning that choosing participants with a higher level of experience, “while seemingly providing more realistic conditions, might create substantial difficulties in an experimental study.”

4.3.8.3 Modeling Knowledge Eight studies use adapted versions of the process modeling knowledge test Mendling et al. (2012b) propose; of these eight studies, six find a strong positive effect on comprehension accuracy (Figl and Laue 2015; Figl and Strembeck 2015; Kummer et al. 2016; Figl et al. 2013b; Mendling and Strembeck 2008; Recker 2013), while the effect was not significant in two (Figl et al. 2013a; Recker et al. 2014). Additional research on user preferences regarding representations to understand a process shows that knowledge of conceptual modeling heightens the preference for process diagrams over structured text (Figl and Recker 2016).

4.3.8.4 Education and User Characteristics Other individual factors that positively affect comprehension accuracy include higher education in general (Weitlaner et al. 2013), as academics and high-school graduates perform better than apprenticeship graduates do. Döhring et al.’s (2014) study compares students with post-docs and industry employees and finds no evidence of a difference, possibly because students had more modeling training, while seniors had more practical experience. Recker and Dreiling (2011) show that the use of participants’ native language in labels is significantly positively related to performance in filling out a cloze test on a process but find no evidence of an effect on comprehension accuracy. Cultural background (Germanic versus Confucian) also has no effect on comprehension accuracy, although participants from Germanic

cultures rate the models as more difficult to understand (Kummer et al. 2016).

Research also addresses the influence of model readers' learning styles and cognitive styles on comprehension. The sensing learning style, which characterizes learners who "prefer learning and memorizing facts from a process model bit-by-bit" (Recker et al. 2014, p. 204), and the surface learning strategy, which indicates learning by memorization, are positively related to comprehension accuracy, while abstraction ability and surface learning motive, the last of which indicates extrinsic (instead of intrinsic motivation) and low learning intensity, are negatively associated with comprehension accuracy (Recker et al. 2014). Moreover, the spatial cognitive style heightens the preference for diagrams over text, while the verbal cognitive style lowers it (Figl and Recker 2016).

4.4 Theoretical Discussions Versus Empirical Studies

The articles that offer "theoretical" viewpoints on process model comprehension are diverse in terms of research approaches, but all have in common that they discuss one or more potential influence factors for comprehension without measuring model comprehension. This is not to say that they do not use other forms of empirical research to support their claims. For instance, some use expert surveys to evaluate proposed quality marks (Overhage et al. 2012) or patterns (Rosa et al. 2011) for process models, while others ask users to rate the visual similarity of models to identify visual layout features (Bernstein and Soffer 2015) or to use proposed decomposition guidelines in modeling sessions (Milani et al. 2016). Some articles perform an extensive literature search, such as one on decomposition in process models (Milani et al. 2016) or look at existing notations and modeling tools (La Rosa et al. 2011) or large process model repositories (Weber et al. 2011) to infer heuristics or patterns for modeling practice. Others build on reference theories from fields like cognitive psychology and adapt them to the process-modeling domain (Zugal et al. 2012; Figl and Strembeck 2014) or use generic frameworks such as those for the quality of modeling notations (Moody 2009) to assess process modeling notations (Figl et al. 2009; Genon et al. 2010). While the first proposals for modeling guidelines provide no empirical evidence on which to build (e.g., Becker et al. 1995), later guidelines (Mendling et al. 2010a; Leopold et al. 2016; Mendling et al. 2012a) refer to empirical data. For instance, these guidelines interpret the occurrence of errors related to correctness in large, natural collections of process models (e.g., Mendling 2007) or violations of modeling style guides written for practitioners (Leopold et al. 2016) as indicators of comprehension difficulties, incorporating also selected

empirical studies on model comprehension that are also part of this review.

Empirical studies like controlled experiments, in which researchers actively manipulate factors in order to observe their effects, can provide insights into cause-and-effect, so they can offer evidence on whether influence factors and practical guidelines derived theoretically not only contribute to overall model quality but actually ease model comprehension. The following sections contrast theoretical discussions and empirical studies according to the type of influence factor they address.

4.4.1 Presentation Medium

Only one study compares interactive web-based visualizations of process models to models on paper (Turetken et al. 2016). In other studies, comparisons between models on paper versus those on computer screens are only a side issue used to combine datasets. Theoretical discussions do not comment on the representation medium per se but do suggest options for how modeling tools and visualizations could ease comprehension in relation to other influence factors. Additional experiments are recommended in order to determine the effect of the choice of information channel and the potential benefits of interactive navigation and interaction strategies in tools for human information-acquiring behavior in process comprehension.

4.4.2 Notation

4.4.2.1 Representation Paradigm Comparison of theoretical and empirical work shows that the research discusses the representation paradigm (e.g., text versus model) theoretically and subjects it to empirical evaluation. The main conclusion is that users prefer diagrammatic representations, but prior experience and training is an important precondition if they are to benefit from diagrams more than from textual representations in comprehension tasks.

The research also compares the procedural process paradigm with declarative models both conceptually and empirically.

In addition, the research validates a variety of alternative visualizations empirically, but such research might be pursued in more directions. For instance, animation and narration techniques used to increase the intuitiveness of process representation (Aysolmaz and Reijers 2016) have not been addressed empirically. Newly proposed visualization opportunities like augmenting process tasks in models with storyboard-like shots of a 3D virtual world to simulate the process (Kathleen et al. 2014) or associating activities with user stories (Trkman et al. 2016) could illustrate business activities vividly while improving comprehension. Furthermore, the effect on comprehension of

using semantically oriented pictorial elements like icons and images could be assessed empirically in more detail using objective comprehension tests since existing research investigates icons only from a subjective point of view.

4.4.2.2 Primary Notation and Notational Characteristics A high number of studies in all three article categories use or discuss BPMN as a notation for modeling processes, which reflects the establishment of BPMN as the de-facto standard for business process modeling (Kocbek et al. 2015).

The effect on comprehension of such process modeling notations as BPMN, UML AD, YAWL, EPCs, and Petri Nets is analyzed theoretically and in empirical studies, the latter of which also evaluate a variety of less common notations. While there is still room for future research to clarify their interplay, the extant empirical work makes several attempts to investigate such notational characteristics as semiotic clarity, perceptual discriminability, semantic transparency, and visual expressiveness. For example, Recker (2013) demonstrates that using gateways aids in model comprehension more than implicit splits and joins in a process do because of a perceptual discriminability effect. This finding is in agreement with hints from error analyses of process model repositories that indicate that implicit representation is often misunderstood, an indication that is also explained by low semiotic clarity because BPMN allows more than one way to model the same concept (Leopold et al. 2016). Concerning semantic transparency of BPMN, Genon et al. (2010) propose new symbols with higher levels of intuitiveness, and Leopold et al. (2016) hypothesize that throwing message events are misunderstood as passive instead of active, based on the event symbol. Future research that measures comprehension could empirically evaluate such hypotheses.

Such criteria as semiotic clarity and visual expressiveness can also be determined in expert evaluations like ontological analyses that compare potentially relevant semantic concepts and symbols, in the case of semiotic clarity, or by identifying visual variables used, in the case of visual expressiveness. However, determining the degree to which they affect comprehension is best done in user studies. There are opportunities for scholars to examine such other notational characteristics as graphic economy and restriction and the extension of a notation's syntax and semantics.

4.4.3 Secondary Notation

Many questions in the area of secondary notation of process models remain in need of empirical investigation.

4.4.3.1 Decomposition Still more empirical work can be done on the decomposition of models and hierarchical

structuring. Zugal et al. (2012) describe on theoretical grounds how decomposition could lead to two opposing effects: abstraction, which would aid comprehension, and a potential split-attention effect, which would lower comprehension. Research questions like this are central to understanding human behavior related to information acquisition in the context of process models to determine benefits of “hiding” irrelevant information in sub-processes. Turetken et al. (2016) report no evidence of increased comprehensibility from using such abstraction; on the contrary, tasks that require information from sub-processes are answered better when this information is not hidden (and, thus, no split-attention effect could occur). Future empirical research should specify a tradeoff curve between the effect of abstraction and split-attention on comprehension and determine how interactive model visualizations can support human information-seeking in larger process models. While Johannsen et al. (2014) are first to have investigated decomposition heuristics experimentally for EPCs, the literature also contains other proposals for decomposition heuristics that have not yet been looked at from an empirical point of view (e.g., Milani et al. 2016).

4.4.3.2 Gestalt Theory Some studies address dual coding and highlighting. Reijers et al. (2011a) demonstrates the potential of color use for highlighting elements or syntax structures to influence novices' information-search behavior, as novices lack the task-specific experience to identify matching gateway patterns. Color use seems to depend on the type of information for which it is used, as it is not always beneficial (Kummer et al. 2016).

Turetken et al. (2016) investigate visual enclosure to highlight elements that belong to a sub-model, and others investigate swim lanes to group activities visually according to their actors (Bera 2012; Jeyaraj and Sauter 2014). Model readers have to use different information-search behaviors that apply to identifying actors, with and without swim lanes, both of which seem to satisfy this particular information need. When they are not directly relevant to the information need, the additional visual information on elements belonging together may even lower perceived ease of understanding (Turetken et al. 2016).

Additional studies on other aspects of dual coding, such as alignment of symbols and textual annotations and highlighting of elements, may provide additional valuable insights. (See, e.g., La Rosa et al. (2011) for ideas on highlighting.)

4.4.3.3 Layout Empirical studies address few variables related to process models' layouts. Figl and Strembeck (2015) investigate the overall model direction, but this experiment should be replicated in combination with layout

options in order to shed light on why no significant differences are found related to flow direction. Future research could assess whether insights from related fields, such as graph drawing, can be generalized to process models. (For example, Purchase et al. (2000) show that minimizing the number of crossings and maximizing symmetry are related to comprehension.) Some conceptual work has been done: Bernstein and Soffer (2015) identify process models' key layout features (e.g., symmetry, angles, overall shape and size, alignment of elements) based on users' perceptions of models' readability and define measurable metrics to characterize the layout of process models. Schrepfer et al. (2009) propose how model layout might relate to the model's comprehensibility. Effinger et al. (2011) perform a user evaluation of the criteria for process models' layout (e.g., bending and crossing of edges, arrangement, overlapping and size of elements, coloring), but did not measure the effect of these criteria on the comprehension of process models. Additional investigation into and experimentation regarding factors related to the model's layout are required to determine the degree to which these factors are of only aesthetic value or in which cases they hinder information reception from process models.

4.4.4 Label Characteristics

Label characteristics carry the meaning of the process – that is, the semantic “information.” Fettke et al. (2012), Mendling (2013) and Overhage et al. (2012) analyze label design theoretically, making evident that empirical research already addresses most of the topics mentioned (syntactic and semantic naming conventions and label design), but this handful of studies focuses only on activity labels and does not always measure comprehension objectively. While user preference studies provide valuable insights, it would be practically relevant to test objectively whether naming conventions like the verb-object style for activities are superior to others because many organizations adhere to existing modeling guidelines that suggest this labeling style and to extend research to labels of gateways and events. A variety of challenges that are associated with automatically improving labels have also been studied extensively (Mendling et al. 2014). Research deals with designing automatic tools to improve model quality, such as for detecting violations of naming conventions (Leopold et al. 2013), automatically relabeling (Leopold et al. 2010) or proposing reuse options of existing labels based on glossaries (e.g., Peters and Weidlich 2009). Therefore, the underlying foundations of design decisions for automatic tools should build on a strong empirical basis.

Theoretical work also contains analyses of the factors of label design that influence model comprehension

(Koschmider et al. 2015a), but these factors are not yet empirically evaluated.

4.4.5 Process Model Characteristics

Several empirical studies measure metrics and model comprehension. Overall, it seems that model characteristics regarding size, structuredness, difficulty of control structures, events, and routing paths are studied exhaustively. Future research could concentrate on explaining the inconsistencies in prior results. The visualization of metrics related to comprehension can be helpful to future research in this area (see e.g., Storch et al. 2013).

4.4.5.1 Size Measures Articles that provide modeling guidelines give concrete, practical hints for some size measures, such as “do not use more than 31 [nodes]” or “use no more than 2 start and end events” (Mendling et al. 2012a, p. 1195), which are grounded on analyses of errors in large collections of process models. Experimental studies on the comprehension of process models do not compare models beyond or below specific thresholds but find an overall negative correlation between larger size and comprehension. Similarly, Sánchez-González et al. (2010, 2012) calculate thresholds for the levels of comprehension based on data sets of models whose comprehension scores come from user studies. While information about the size at which a process model begins to be difficult to understand is useful for practitioners, more complex characteristics should also be taken into account. These are detailed below.

4.4.5.2 Modularity/Structuredness While block structuredness is advised in modeling guidelines (Mendling et al. 2010a; Rosa et al. 2011), the theoretical literature also warns of drawbacks like the duplication of model elements and gives advice on when to prefer unstructured modeling (Gruhn and Laue 2007). These ideas are in line with Dumas et al.'s (2012) empirical results, which show that structuredness is not always superior.

The metric nesting depth's tendency to heighten cognitive load is addressed in theoretical discussions (Azim et al. 2008; Storch et al. 2013) and supported by empirical evidence (Figl and Laue 2015; Sánchez-González et al. 2010).

4.4.5.3 Refactoring More theoretical and practical discussions than empirical studies address model refactoring – that is, replacing “process model fragments [with] semantically equivalent ones” to achieve higher quality and understandability of models (Fernández-Roperero et al. 2013, p. 1397). In principle, model refactoring lowers the intrinsic cognitive load in understanding a process model without

changing the process's behavior. Weber et al. (2011) collect a catalogue of indicators of bad process model quality to identify refactoring opportunities and provide a set of behavior-preserving techniques for refactoring to avoid redundancies and unnecessary increases in the model's complexity. Automatic tools can help to identify "syntactical anti-patterns" (Laue and Awad 2011) and exchange them. Overall, the potential of automatic model refactoring and patterns to reduce cognitive load remains to be determined, so there are opportunities for scholars to examine these patterns in comprehension experiments.

4.4.5.4 Redundant Elements While theoretical discussions mention the importance of removing redundant elements (Rosa et al. 2011) and frame rules for how to deal with redundancies (Becker et al. 1995), no empirical studies investigate the potential negative effect of redundancy on comprehension.

4.4.5.5 Gateway Interplay/Control Structures Although theoretical discussions advise keeping gateway heterogeneity and gateway mismatch low in models (Mendling 2013), evidence in empirical studies is missing or weak. One problem with relating the global metrics of models to overall model comprehension scores is that, if characteristics are not isolated, they can be confounded with other characteristics of the models. Constructing informationally equivalent models while varying gateway heterogeneity and gateway mismatch is difficult, if not impossible, so empirical studies also use specifically designed comprehension tasks that relate to various control structures. Overall, empirical studies support the theoretical hypotheses that order is easy to understand, while repetition is more difficult to understand (Gruhn and Laue 2006). Gruhn and Laue (2006) also theorize that exclusiveness is easier to understand than concurrency is, but research does not yet report evidence of differences in comprehensibility. Guidelines for avoiding OR gateway elements (Mendling et al. 2010a) might have caused researchers to refrain from including OR in their comprehension studies at all, as only one study includes OR, and it reports weak evidence of OR's being more difficult to understand than AND and XOR are (Sarshar and Loos 2005).

4.4.5.6 Connection The recommendation to minimize the number of routing paths per element (Mendling et al. 2010a) is consistent with the results of empirical studies (Reijers and Mendling 2011).

4.4.5.7 Syntax Rules Few empirical studies investigate the effect on model comprehension of adhering to syntax rules; additional research must be undertaken before the association between those variables is more clearly understood.

4.4.6 Task Characteristics

Some empirical studies investigate aspects of comprehension tasks (e.g., wording of tasks (Laue and Gadatsch 2011), validity of statements (Figl and Laue 2015), sequential versus circumstantial tasks (Pichler et al. 2012)) that seem to be more relevant to measuring comprehension than they are to improving process models' comprehensibility. One study looks at how additional material provided by a catalog of routing possibilities (Soffer et al. 2015) can help users complete comprehension tasks on the basis of a process model. There is no direct equivalent of these variables in theoretical discussions.

4.4.7 Education and User Characteristics

An important contribution of the literature review on user characteristics is to ease the selection of variables for future researchers. The overview provides hints on which measures have been found to correlate with the comprehension of process models. With the exception of knowledge and experience in the process modeling domain, a variety of user characteristics could be generalized from process models to the parent class of conceptual models (and vice versa). This potential is also reflected in the theoretical discussions on process model comprehension. For instance, Schrepfer et al. (2009) present a framework for various aspects of expertise that is not grounded in cognitive theories but on how other modeling studies measure various areas of expertise. Most user characteristics that theoretical work mentions are included in studies, but they are typically collected as control variables, in addition to other variables of interest.

4.4.7.1 Tailoring Aysolmaz and Reijers (2016) mention the idea of tailoring process models to personal factors, an idea that has not been empirically assessed but seems practically and theoretically important. In a similar vein, Becker et al. (2004) present "configurative" process models, which can automatically adapt their information to user groups. Future work may address the extent to which this idea can ease comprehension, for example, by tailoring process models to aspects of cultural context, expert-novice differences, or cognitive styles, all of which are variables for which empirical studies report interaction effects with various forms of representation.

4.4.7.2 Domain Knowledge As theoretical discussions have highlighted the importance of domain knowledge (Schrepfer et al. 2009), it was surprising that variables that measure domain knowledge show no significant correlations with model comprehension in empirical studies. As

these results may be influenced by variance in the variable domain knowledge in the studies, future work could compare domain experts with users who are unfamiliar with a domain to determine the magnitude of a possible effect of domain knowledge on comprehension.

4.4.7.3 Experience and Familiarity with Modeling Despite the wealth of measures used to assess experience and familiarity, the research reports statistical evidence of an effect on comprehension for only a few. The possible reasons for this result have already been discussed. Future researchers must be particularly cautious in interpreting self-rated measures of modeling experience since Mendling et al. (2012b) report that their relationship with comprehension is u-shaped and not linear.

4.4.7.4 Modeling Knowledge In contrast to self-reported familiarity and knowledge, the process modeling knowledge test Mendling et al. (2012b) propose correlates with comprehension in six of eight studies, demonstrating its content validity. Therefore, this instrument can be recommended to control for individual variation in human information behavior related to the comprehension of process models.

4.4.7.5 Education and User Characteristics The study of Figl and Recker (2016), which focuses on cognitive styles, and that of Recker et al. (2014), which focuses on learning style, are unique in that user characteristics are the main variables of interest. Kummer et al. (2016) provide the first experimental analysis of the cultural dependency of human information behavior related to process comprehension but find no evidence of an effect.

The theoretical discussions mention additional variables, such as perceptual expertise, ability to recognize patterns (Schrepfer et al. 2009), user attitudes, and self-efficacy (Reijers et al. 2010), none of which are addressed in process modeling research and all of which future studies could take into account.

5 Reflections on Research Approaches and Methods

Section 4 provided an overview of the current state of the research field of process model comprehension. To deepen insight into the landscape of process model comprehension research, this section provides a view from the perspective of the wider research fields of IS and human information behavior. This section also explicates the current boundaries of existing empirical studies on process model comprehension in relation to research approaches, variable selection, and measurement of comprehension in order to

inspire future researchers with ideas on how to move the field forward.

5.1 Research Approaches and Questions

The dominant research approaches to producing new knowledge in the IS field and contributing to practice and theory have been controversially discussed in recent years (see e.g., Bichler et al. 2016; Grover and Lyytinen 2015; Hevner et al. 2004). Against this background, we discuss and position studies in the field of process model comprehension. Many of the empirical studies in our review borrow reference theories from fields like cognitive psychology and linguistics to generate hypotheses on comprehension and explain their results (see Houy et al. (2014) for an overview of the theories used.) However, the selection of independent variables makes evident that the selected studies' research goals go beyond generalization of reference theories to the field of process modeling (e.g., generalizing abstract cognitive theories on cognitive styles or cognitive load to the specific task of comprehending models, generalizing linguistic theories to the specific field of model labels) to answer practical questions related to modeling. By addressing questions like which notation to prefer, which label style to use, and how to decompose models or by calculating the thresholds for model metrics for various levels of cognitive difficulty, the research field seems to be closer to practice and more tangible than "typical" IS research as Grover and Lyytinen (2015) characterized it. Innovative design artifacts as new visualizations, layout, and labeling strategies become subject to empirical evaluation. Thus, the ongoing research on the comprehension of process models complements design-oriented research in process modeling with behavioral science and ensures its effectiveness. Some authors go as far as to categorize their experimental research as "design science research" because they evaluate artifacts (e.g., Bera 2012, p. 67).

Grover and Lyytinen (2015) call for new ways of conducting behavior research in the IS field. We identified examples of such "new ways" in our set of empirical studies that Grover and Lyytinen (2015) argue are used insufficiently in the IS field and demonstrate the variety of existing research approaches in the field of process model comprehension, including data-driven research instead of theory-driven research (e.g., research that interprets model characteristics in relation to comprehension) and the development of new measurement instruments (e.g., for perceived ambiguity in labels (Mendling et al. 2010b) or notational characteristics (Figl et al. 2013b)).

5.2 Measurement of Comprehension

This section discusses how researchers can improve the measurement of process model comprehension.

Measures of subjective comprehension are diverse in the extant research and include such measures as perceived ease of use and perceived usability, which are widely used in the IS field.

The “objective” measurement method of comprehension used predominantly in the extant empirical studies is the multiple-choice comprehension task. Low comprehension scores offer hints about the factors that might cause the failure of human information-acquiring behavior when people use process models. By constructing comprehension tasks, researchers define participants’ information needs that the researchers see as representative of the real-life use of process models and observe whether these needs are satisfied. Comprehension tasks determine and shape how well users take up information from a process model. In this context, researchers who use such tasks should be aware of the “principle of the least effort,” which indicates that humans take up only as much information as is necessary to solve a particular task (Simon 1955; Heinrich et al. 2014). Future research could also investigate the “cognitive stopping rules” used to minimize the effort required in searching for information, as investigated in other areas of human information-seeking behavior (Hemmer and Heinzl 2011; Browne and Pitts 2004; Browne et al. 2007). Studies in the comprehension of process models report fatigue effects in experiments, as participants took less time and scored worse in later comprehension tasks in the experiments (Recker 2013). Such results could hint that participants’ “stopping behavior” changes over the course of the experiment. The time a participant takes to complete a comprehension task might be a key to determining whether the participants stopped working too soon to complete a task correctly.

The emerging research field of neuroIS, which applies neuroscience theories and tools to IS research, has the potential to provide improved ways to clarify process model comprehension and evaluate visual notations in this context (Riedl et al. 2010; Riedl 2009; Dimoka et al. 2010). Neuroimaging tools that measure brain activation (e.g., functional magnetic resonance imaging (fMRI, EEG)) could be used to measure process model comprehension more objectively and in a more fine-grained way than can the rating scales and multiple-choice tasks that many comprehension studies currently use. Research in the related area of software program comprehension successfully employs fMRI to measure comprehension (Siegmond et al. 2014). Vom Brocke et al. (2013) give examples of how neuroscience theories on visual

perception can be applied in research on visual modeling notations; for instance, fMRI studies demonstrate a preference for curved objects and an association between sharp visual objects and threat, which could hamper cognitive processing. EEG has been applied to the measurement of cognitive workload related to working memory, which is central to model comprehension [see Müller-Putz et al. (2015) for an introduction to how to apply this technique to measure cognitive load in IS research].

Eye-tracking is the only comparatively simple neurophysiological tool that research has already applied successfully to the area of process model comprehension (Pinggera et al. 2013; Petrusel et al. 2016). Experiments demonstrate that eye-tracking metrics work, as the fixation of eyes on a relevant region can predict process model comprehension task scores up to 70% (Petrusel and Mendling 2013). One empirical study reports that the mental effort related to human information processing could be captured well via eye-tracking but that traditional comprehension tasks did not detect differences between the experimental groups they used (Petrusel et al. 2016). This finding may also provide a hint that eye-tracking is a preferred method in investigations of early stages of human information processing (e.g., Gestalt laws of perceptual organization).

Future research is advised to adopt these neuroscientific methods in order to provide more definitive evidence of the factors that influence model comprehension.

6 Limitations and Future Research Directions

As with all descriptive reviews, the current work has limitations.

6.1 Selection of Literature

To mitigate the degree of subjectivity, we reported detailed search criteria and inclusion criteria for the selected articles, but limitations to generalizability lie in the choice of these selection criteria. Because of the focus on individual process models, we did not take into account the literature on process model repositories (e.g., on different visualization opportunities for larger model collections) and on connections to other types of models.

In addition, we considered only those articles that have a strong relationship to process models. There is considerable cognitive research in the larger field of conceptual modeling and visualization in general that may offer insights and perspectives that are relevant to the comprehension of process models.

6.2 Categories of the Framework

The dimensions of the categories and subcategories of the framework differ widely. For instance, many user characteristics are stable and cannot be changed (e.g., native language, cognitive style), so we can adapt the models to the individual users, but not the other way around. (Modeling training and experience could change over time, but they are given at one point in time.) Moreover, it is possible to change a model's primary and secondary notational factors (e.g., layout, symbols, highlighting) while preserving informational equivalence, but models whose characteristics (e.g., size metrics) differ are typically not informationally equivalent. The review discusses such distinctions in terms of intrinsic and extraneous cognitive load, but caution must be applied, as the relative importance of findings in different categories cannot be compared with each other directly, and different numbers of studies were selected in each category. In addition, the influence factors of different categories might interact in a variety of ways, but a full discussion of all possible interaction effects lies beyond the scope of this review, which focuses on the main effects of the independent variables, and would also be difficult because only a few articles report interaction effects.

6.3 Reporting of Studies

To move the research field forward, this article encourages scholars to report exact p-values and effect sizes when describing the statistical results of future experiments in order to combine probability values in a meta-analysis. Such details would ease the management of variation in results among experiments and the determination of the effects' consistency and strength. To deal with the existing shortcomings of the descriptions of the results, we developed a coding schema with which to categorize the level of evidence regarding the effects. One limitation of this categorization is that, if a study investigates, for example, more than one type of user or model, its chance of reporting "conflicting" evidence about an effect is higher than it is for simpler studies that look at the effect of only one independent variable on comprehension.

We refrained from taking the quality of study designs into account because a wide range of study designs other than randomized, controlled experiments have been used that would not fit into existing coding schemas, such as those used for clinical trials. By not taking all study characteristics into account, our level-of-evidence rating, which is based primarily on descriptions of statistical results, might differ from the statements of more cautious authors. For instance, we rated the level of evidence of the effects of two variables on comprehension as "strong"

based on the reported statistical results, while the authors themselves argue that "the number of models is too small to make strong claims" (Reijers and Mendling 2011, p. 9).

Future research might standardize ways to describe the research design and the experimental manipulations, which would ease comparisons of the quality of studies in this research area.

6.4 Type of Contribution

Concerning the nature of the theoretical contribution based on the five theory types in IS research (Gregor 2006), this review article provides *descriptions* of the current state of research in the field regarding the independent variables that affect comprehension. Moreover, the article contributes to theory-building by *explaining* and *predicting* the comprehensibility of process models. The review provides an overview of why and how some variables affect comprehension. However, although the review extracts the relevant variables mentioned in existing modeling guidelines, it is beyond its scope to *prescribe* how to construct process models.

Given that the preconditions for reporting mentioned in the last section are satisfied, future work could offer a precise *prediction* theory on how certain changes in a process model, its notation, the user group, or the task type will affect model comprehension, thereby laying the foundation for future prescriptive modeling guidelines.

7 Conclusion

The descriptive review presented here provides a deepened and contextualized body of knowledge on process model comprehension. By developing a comprehensive framework of the factors that influence comprehension, the article adds substantially to the creation of a cumulative tradition of empirical work on model comprehension and provides a basis to be systematically updated by new studies. Thus, the article sheds light on the types of variables that influence intrapersonal information-acquiring behavior as it relates to process models.

From a research perspective, the article provides a foundation for future process modeling research, thereby establishing a potential to be adapted to other areas of conceptual modeling. Literature reviews like this one give scholars a quick overview of existing literature and motivate them to investigate knowledge gaps and formulate new research hypotheses on influence variables that have not been addressed (Webster and Watson 2002).

Moreover, the article contributes to advancing the field of human information behavior research. While other reviews provide overviews of empirical studies on human

information behavior in computer-mediated contexts (e.g., Hemmer and Heinzl 2011), this review is the first to address human information behavior in the conceptual modeling field, another central research topic in the IS field. Process modeling is at the core of designing information systems, so the article sheds light on an important facet of information-use behavior in the IS discipline: how the information in process models is incorporated into readers' existing knowledge.

From a practical perspective, the article helps to explain how to develop useful and understandable process models and how best to exploit process modeling as a cognitive tool for a variety of users. In the long run, the article can also contribute to the development of process models that are optimized for human understanding and problem-solving. While the article focuses on model *product* quality – that is, the comprehensibility of a finished model – its identification of influence factors can help to guide and improve the quality of the modeling *process* by helping model designers to focus on the criteria that can prevent comprehension problems. Thus, the article's insights can also be used in process modeling training and in educational texts.

The article also has direct and immediate significance for business process modeling practice, as its findings can inform ongoing revisions of process modeling notations and tool development. Future research can refine and evaluate existing practical guidelines for process modeling based on the review of empirical insights presented here.

Acknowledgements Open access funding provided by Vienna University of Economics and Business (WU).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aguilar ER, García F, Ruiz F, Piattini M, Visaggio CA, Canfora G (2008) Evaluation of BPMN models quality—a family of experiments. In: 3rd International conference on evaluation of novel approaches to software engineering, Funchal, pp 56–63
- Aysolmaz B, Reijers HA (2016) Towards an integrated framework for invigorating process models: a research agenda. In: Reichert M, Reijers HA (eds) Business process management workshops 2015, LNBIP 256, Innsbruck, Austria. Springer International Publishing, pp 552–558. doi:10.1007/978-3-319-42887-1_44
- Azim A, Ghani A, Koh Tieng WG, Muketha M, Wen WP (2008) Complexity metrics for measuring the understandability and maintainability of business process models using goal-question-metric. *Int J Comput Sci Netw Secur* 8(5):219–225
- Baddeley AD (1992) Working memory. *Science* 255(5044):556–559
- Bandara W, Gable GG, Rosemann M (2005) Factors and measures of business process modelling: model building through a multiple case study. *Eur J Inf Syst* 14(4):347–360. doi:10.1057/palgrave.ejis.3000546
- Becker J, Rosemann M, Schütte R (1995) Grundsätze ordnungsmäßiger Modellierung. *WIRTSCHAFTSINFORMATIK* 37(5):435–445
- Becker J, Delfmann P, Dreiling A, Knackstedt R, Kuropka D (2004) Configurative process modeling—outlining an approach to increased business process model usability. In: Khosrow-Pour M (ed) 14th Information resources management association international conference. IRM, Hershey, pp 615–619
- Bera P (2012) Does cognitive overload matter in understanding BPMN models? *J Comput Inf Syst* 52(4):59–69
- Bernstein V, Soffer P (2015) Identifying and quantifying visual layout features of business process models. In: Enterprise, business-process and information systems modeling. Springer, Heidelberg, pp 200–213
- Bichler M, Frank U, Avison D, Malaurent J, Fettek P, Hovorka D, Krämer J, Schnurr D, Müller B, Suhl L (2016) Theories in business and information systems engineering. *Bus Inf Syst Eng* 58(4):291–319
- Browne GJ, Pitts MG (2004) Stopping rule use during information search in design problems. *Organ Behav Hum Decis Process* 95(2):208–224
- Browne GJ, Pitts MG, Wetherbe JC (2007) Cognitive stopping rules for terminating information search in online tasks. *MIS Q* 31(1):89–104
- Burton-Jones A, Wand Y, Weber R (2009) Guidelines for empirical evaluations of conceptual modeling grammars. *J Assoc Inf Syst* 10(6):495–532
- Cardoso J (2006) Process control-flow complexity metric: an empirical validation. In: IEEE International conference on services computing, SCC'06, pp 167–173
- Chandler P, Sweller J (1996) Cognitive load while learning to use a computer program. *Appl Cognit Psychol* 10(2):151–170
- Charette RN (2005) Why software fails. *IEEE Spectr* 42(9):42–49. doi:10.1109/MSPEC.2005.1502528
- Chen C, Yu Y (2000) Empirical studies of information visualization: a meta-analysis. *Int J Hum Comput Stud* 53(5):851–866
- Claes J, Vanderfeesten I, Reijers HA, Pinggera J, Weidlich M, Zugal S, Fahland D, Weber B, Mendling J, Poels G (2012) Tying process model quality to the modeling process: the impact of structuring, movement, and speed. In: Barros A, Gal A, Kindler E (eds) Business process management, LNCS, vol 7481. Springer, Heidelberg, pp 33–48
- Dimoka A, Banker RD, Benbasat I, Davis FD, Dennis AR, Gefen D, Gupta A, Ischebeck A, Kenning P, Pavlou PA (2010) On the use of neurophysiological tools in IS research: developing a research agenda for NeuroIS. *MIS Q* 36(3):679–702
- Döhring M, Reijers HA, Smirnov S (2014) Configuration vs. adaptation for business process variant maintenance: an empirical study. *Inf Syst* 39:108–133. doi:10.1016/j.is.2013.06.002
- Dumas M, La Rosa M, Mendling J, Mäesalu R, Reijers H, Semenenko N (2012) Understanding business process models: the costs and benefits of structuredness. In: Ralyté J, Franch X, Brinkkemper S, Wrycza S (eds) Advanced information systems engineering, LNCS, vol 7328. Springer, Heidelberg, pp 31–46. doi:10.1007/978-3-642-31095-9_3
- Effinger P, Jogsch N, Seiz S (2011) On a study of layout aesthetics for business process models using BPMN. In: Mendling J, Weidlich M, Weske M (eds) Business process modeling notation, LNBIP, vol 67. Springer, Heidelberg, pp 31–45. doi:10.1007/978-3-642-16298-5_5
- Fahland D, Lübke D, Mendling J, Reijers HA, Weber B, Weidlich M, Zugal S (2009) Declarative versus imperative process modeling

- languages: the issue of understandability. In: Halpin T, Krogstie J, Nurcan S et al (eds) *Enterprise, business-process and information systems modeling*. Lecture Notes in Business Information Processing, vol 29. Springer, Berlin, Heidelberg, pp 353–366. doi:[10.1007/978-3-642-01862-6_29](https://doi.org/10.1007/978-3-642-01862-6_29)
- Fernández-Ropero M, Pérez-Castillo R, Cruz-Lemus JA, Piattini M (2013) Assessing the best-order for business process model refactoring. In: 28th Annual ACM symposium on applied computing, Coimbra
- Fettke P, Vella AL, Loos P (2012) From measuring the quality of labels in process models to a discourse on process model quality: a case study. In: 45th Hawaii international conference on system sciences (HICSS), pp 197–206. doi:[10.1109/HICSS.2012.290](https://doi.org/10.1109/HICSS.2012.290)
- Figl K, Laue R (2011) Cognitive complexity in business process modeling. In: Mouratidis H, Rolland C (eds) *Advanced information systems engineering*, LNCS, vol 6741. Springer, Heidelberg, pp 452–466. doi:[10.1007/978-3-642-21640-4_34](https://doi.org/10.1007/978-3-642-21640-4_34)
- Figl K, Laue R (2015) Influence factors for local comprehensibility of process models. *Int J Hum Comput Stud* 82:96–110. doi:[10.1016/j.ijhcs.2015.05.007](https://doi.org/10.1016/j.ijhcs.2015.05.007)
- Figl K, Recker J (2016) Exploring cognitive style and task-specific preferences for process representations. *Requir Eng* 21(1):63–85. doi:[10.1007/s00766-014-0210-2](https://doi.org/10.1007/s00766-014-0210-2)
- Figl K, Strembeck M (2014) On the importance of flow direction in business process models. In: 9th International conference on software engineering and applications (ICSOFT-EA). SCITE-PRESS, Vienna
- Figl K, Strembeck M (2015) Findings from an experiment on flow direction of business process models. In: International workshop on enterprise modelling and information systems architectures (EMISA), Innsbruck
- Figl K, Mendling J, Strembeck M (2009) Towards a usability assessment of process modeling languages. In: 8th GI-Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK), CEUR-WS, vol 554. Berlin, pp 138–156
- Figl K, Mendling J, Strembeck M, Recker J (2010) On the cognitive effectiveness of routing symbols in process modeling languages. In: 13th international conference on business information systems (BIS), Berlin
- Figl K, Mendling J, Strembeck M (2013a) The influence of notational deficiencies on process model comprehension. *J Assoc Inf Syst* 14(6):312–338
- Figl K, Recker J, Mendling J (2013b) A study on the effects of routing symbol design on process model comprehension. *Decis Support Syst* 54(2):1104–1118. doi:[10.1016/j.dss.2012.10.037](https://doi.org/10.1016/j.dss.2012.10.037)
- Gemino A, Wand Y (2004) A framework for empirical evaluation of conceptual modeling techniques. *Requir Eng* 9(4):248–260
- Genon N, Heymans P, Amyot D (2010) Analysing the cognitive effectiveness of the BPMN 2.0 visual syntax. In: *Software language engineering*, LNCS. Springer, Heidelberg, pp 377–396
- Gregor S (2006) The nature of theory in information systems. *MIS Q* 30(3):611–642
- Grover V, Lyytinen K (2015) New state of play in information systems research: the push to the edges. *MIS Q* 39(2):271–296
- Gruhn V, Laue R (2006) Adopting the cognitive complexity measure for business process models. In: 5th IEEE International conference on cognitive informatics, pp 236–241. doi:[10.1109/COGINF.2006.365702](https://doi.org/10.1109/COGINF.2006.365702)
- Gruhn V, Laue R (2007) Good and bad excuses for unstructured business process models. In: 12th European conference on pattern languages of programs (EuroPLOP)
- Haisjackl C, Zugal S (2014) Investigating differences between graphical and textual declarative process models. In: *Advanced information systems engineering workshops*. Springer, Heidelberg, pp 194–206
- Haisjackl C, Barba I, Zugal S, Soffer P, Hadar I, Reichert M, Pinggera J, Weber B (2016) Understanding declare models: strategies, pitfalls, empirical results. *Softw Syst Model* 15(2):325–352
- Heggset M, Krogstie J, Wesenberg H (2015) The influence of syntactic quality of enterprise process models on model comprehension. CAiSE forum. CEUR, Stockholm
- Heinrich LJ, Riedl R, Stelzer D (2014) *Informationsmanagement: Grundlagen, Aufgaben, Methoden*. De Gruyter, Berlin
- Hemmer E, Heinzl A (2011) Where is the “I” in “IS research”? The quest for a coherent research stream in the context of human information behavior. In: *Theory-guided modeling and empiricism in information systems research*. Springer, Heidelberg, pp 223–246
- Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. *MIS Q* 28(1):75–105
- Hipp M, Strauss A, Michelberger B, Mutschler B, Reichert M (2014) Enabling a user-friendly visualization of business process models. *Business process management workshops*, LNBP. Springer, Heidelberg, pp 395–407
- Houy C, Fettke P, Loos P (2012) Understanding understandability of conceptual models—what are we actually talking about? In: Atzeni P, Cheung D, Ram S (eds) *Conceptual modeling*, LNCS, vol 7532. Springer, Heidelberg, pp 64–77. doi:[10.1007/978-3-642-34002-4_5](https://doi.org/10.1007/978-3-642-34002-4_5)
- Houy C, Fettke P, Loos P (2014) On the theoretical foundations of research into the understandability of business process models. In: 22nd European conference on information systems, Tel Aviv
- Indulska M, Green P, Recker J, Rosemann M (2009) Business process modeling: perceived benefits. In: Castano S, Dayal U, Laender AHF (eds) *Conceptual modeling—ER 2009*, LNCS. Springer, Heidelberg, pp 458–471
- Jeyaraj A, Sauter VL (2014) Validation of business process models using swimlane diagrams. *J Inf Technol Manag* 25(4):27
- Johannsen F, Leist S, Braunnagel D (2014) Testing the impact of wand and weber’s decomposition model on process model understandability. In: *International conference on information systems*, Auckland
- Jošt G, Huber J, Heričko M, Polančič G (2016) An empirical investigation of intuitive understandability of process diagrams. *Comput Stand Interfac* 48:90–111. doi:[10.1016/j.csi.2016.04.006](https://doi.org/10.1016/j.csi.2016.04.006)
- Kathleen N, Ross B, Kriglstein S (2014) Storyboard augmentation of process model grammars for stakeholder communication. In: *IEEE International conference on information visualization theory and applications (IVAPPI)*, pp 114–121
- King WR, He J (2005) Understanding the role and methods of meta-analysis in IS research. *Commun Assoc Inf Syst* 16(1):32
- Kirschner PA (2002) Cognitive load theory: implications of cognitive load theory on the design of learning. *Learn Instr* 12(1):1–10
- Kocbek M, Jošt G, Heričko M, Polančič G (2015) Business process model and notation: the current state of affairs. *Comput Sci Inf Syst* 12(2):509–539. doi:[10.2298/CSIS140610006K](https://doi.org/10.2298/CSIS140610006K)
- Kock N, Danesh A, Komiak P (2008) A discussion and test of a communication flow optimization approach for business process redesign. *Knowl Process Manag* 15(1):72
- Kock N, Verville J, Danesh-pajou A, Deluca D (2009) Communication flow orientation in business process modeling and its effect on redesign success: results from a field study. *Decis Support Syst* 46(2):562–575. doi:[10.1016/j.dss.2008.10.002](https://doi.org/10.1016/j.dss.2008.10.002)
- Koschmider A, Figl K, Schoknecht A (2015a) A comprehensive overview of visual design of process model element labels. In: *Business process management workshops*, Innsbruck
- Koschmider A, Ullrich M, Heine A, Oberweis A (2015b) Revising the vocabulary of business process element labels. In: Zdravkovic J, Kirikova M, Johannesson P (eds) *Advanced information systems engineering*, LNCS, vol 9097. Springer, Heidelberg, pp 69–83. doi:[10.1007/978-3-319-19069-3_5](https://doi.org/10.1007/978-3-319-19069-3_5)

- Kummer T-F, Recker J, Mendling J (2016) Enhancing understandability of process models through cultural-dependent color adjustments. *Decis Support Syst* 87:1–12. doi:[10.1016/j.dss.2016.04.004](https://doi.org/10.1016/j.dss.2016.04.004)
- La Rosa M, ter Hofstede AHM, Wohed P, Reijers HA, Mendling J, van der Aalst WMP (2011) Managing process model complexity via concrete syntax modifications. *IEEE Trans Ind Inform* 7(2):255–265
- Laue R, Awad A (2011) Visual suggestions for improvements in business process diagrams. *J Vis Lang Comput* 22(5):385–399. doi:[10.1016/j.jvlc.2011.04.003](https://doi.org/10.1016/j.jvlc.2011.04.003)
- Laue R, Gadatsch A (2011) Measuring the understandability of business process models—are we asking the right questions? In: *Business process management workshops, LNBIP, vol 66*. Springer, Heidelberg, pp 37–48
- Leopold H, Smirnov S, Mendling J (2010) Refactoring of process model activity labels. *Natural language processing and information systems*. Springer, Heidelberg, pp 268–276
- Leopold H, Eid-Sabbagh R-H, Mendling J, Azevedo LG, Baião FA (2013) Detection of naming convention violations in process models for different languages. *Decis Support Syst* 56:310–325. doi:[10.1016/j.dss.2013.06.014](https://doi.org/10.1016/j.dss.2013.06.014)
- Leopold H, Mendling J, Günther O (2016) Learning from quality issues of BPMN models from industry. *IEEE Softw* 33(4):26–33. doi:[10.1109/MS.2015.81](https://doi.org/10.1109/MS.2015.81)
- Lindland OI, Sindre G, Solvberg A (1994) Understanding quality in conceptual modeling. *IEEE Softw* 11(2):42–49
- Mayring P (2003) *Qualitative Inhaltsanalyse (Qualitative Content Analysis)*, 8th edn. Beltz, Weinheim
- Melcher J, Seese D (2008) Towards validating prediction systems for process understandability: measuring process understandability. In: *10th International symposium on symbolic and numeric algorithms for scientific computing (SYNASC'08)*, pp 564–571
- Melcher J, Mendling J, Reijers HA, Seese D (2010) On measuring the understandability of process models. In: *Business process management workshops, LNBIP, vol 43*. Springer, Heidelberg, pp 465–476
- Mendling J (2007) Detection and prediction of errors in EPC business process models. *Vienna University of Economics and Business Administration*
- Mendling J (2013) Managing structural and textual quality of business process models. *Data-driven process discovery and analysis*. Springer, Heidelberg, pp 100–111
- Mendling J, Strembeck M (2008) Influence factors of understanding business process models. In: *Abramowicz W, Fensel D (eds) Business information systems (BIS), vol 7*. Springer, Heidelberg, pp 142–153
- Mendling J, Reijers H, van der Aalst WMP (2010a) Seven process modeling guidelines (7PMG). *Inf Softw Technol* 52(2):127–136
- Mendling J, Reijers HA, Recker J (2010b) Activity labeling in process modeling: empirical insights and recommendations. *Inf Syst* 35(4):467–482
- Mendling J, Sánchez-González L, García F, La Rosa M (2012a) Thresholds for error probability measures of business process models. *J Syst Softw* 85(5):1188–1197. doi:[10.1016/j.jss.2012.01.017](https://doi.org/10.1016/j.jss.2012.01.017)
- Mendling J, Strembeck M, Recker J (2012b) Factors of process model comprehension—findings from a series of experiments. *Decis Support Syst* 53(1):195–206
- Mendling J, Leopold H, Pittke F (2014) 25 challenges of semantic process modeling. *IJISEBC* 1(1):78–94
- Milani F, Dumas M, Matulevičius R, Ahmed N, Kasela S (2016) Criteria and heuristics for business process model decomposition. *Bus Inf Syst Eng* 58(1):7–17
- Moher TG, Mak DC, Blumenthal B, Leventhal LM (1993) Comparing the comprehensibility of textual and graphical programs: the case of petri nets. In: *Cook CR, Scholtz JC, Spohrer JC (eds) Empirical studies of programmers: fifth workshop*, pp 137–161
- Moody DL (2004) Cognitive load effects on end user understanding of conceptual models: an experimental analysis. In: *8th East European conference on advances in databases and information systems*. Springer, Heidelberg, pp 129–143
- Moody DL (2005) Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data Knowl Eng* 15(3):243–276
- Moody DL (2009) The “physics” of notations: towards a scientific basis for constructing visual notations in software engineering. *IEEE Transact Softw Eng* 35(5):756–779
- Moody DL, Shanks G (2003) Improving the quality of data models: empirical validation of a quality management framework. *Inf Syst* 28(6):619–650
- Müller-Putz GR, Riedl R, Wriessnegger SC (2015) Electroencephalography (EEG) as a research tool in the information systems discipline: foundations, measurement, and applications. *Commun Assoc Inf Syst* 37(1):46
- Natschläger C (2011) Deontic BPMN. In: *Hameurlain A, Liddle S, Schewe K-D, Zhou X (eds) Database and expert systems applications, LNCS, vol 6861*. Springer, Heidelberg, pp 264–278. doi:[10.1007/978-3-642-23091-2_23](https://doi.org/10.1007/978-3-642-23091-2_23)
- Olsson M (2005) Beyond ‘needy’ individuals: conceptualizing information behavior. *Proc Am Soc Inf Sci Technol* 42(1). doi:[10.1002/meet.1450420161](https://doi.org/10.1002/meet.1450420161)
- Ottensosser A, Fekete A, Reijers HA, Mendling J, Menictas C (2012) Making sense of business process descriptions: an experimental comparison of graphical and textual notations. *J Syst Softw* 85(3):596–606
- Overhage S, Birkmeier DQ, Schlauderer S (2012) Quality marks, metrics, and measurement procedures for business process models. *Bus Inf Syst Eng* 54(5):217–235. doi:[10.1007/s11576-012-0335-1](https://doi.org/10.1007/s11576-012-0335-1)
- Palmer SE (1992) Common region: a new principle of perceptual grouping. *Cognit Psychol* 24(3):436–447
- Parsons J, Cole L (2005) What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data Knowl Eng* 55(3):327–342
- Patig S, Casanova-Brito V, Vögeli B (2010) IT Requirements of business process management in practice—an empirical study. In: *Hull R, Mendling J, Tai S (eds) Business process management, LNCS, vol 6336*. Springer, Heidelberg, pp 13–28. doi:[10.1007/978-3-642-15618-2_4](https://doi.org/10.1007/978-3-642-15618-2_4)
- Peters N, Weidlich M (2009) Using glossaries to enhance the label quality in business process models. In: *8th GI-Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK)*, vol 554. CEUR, Berlin, pp 75–90
- Petre M (1995) Why looking isn’t always seeing: readership skills and graphical programming. *Commun ACM* 38(6):33–44. doi:[10.1145/203241.203251](https://doi.org/10.1145/203241.203251)
- Petre M (2006) Cognitive dimensions ‘beyond the notation’. *J Vis Lang Comput* 17(4):292–301
- Petrusel R, Mendling J (2013) Eye-tracking the factors of process model comprehension tasks. In: *Salinesi C, Norrie MC, Pastor Ó (eds) Advanced information systems engineering: 25th international conference, CAiSE 2013, Valencia*. Springer, Heidelberg, pp 224–239. doi:[10.1007/978-3-642-38709-8_15](https://doi.org/10.1007/978-3-642-38709-8_15)
- Petrusel R, Mendling J, Reijers HA (2016) Task-specific visual cues for improving process model understanding. *Inf Softw Technol* 79:63–78
- Pichler P, Weber B, Zugal S, Pinggera J, Mendling J, Reijers HA (2012) Imperative versus declarative process modeling languages: an empirical investigation. In: *Business process management workshops, LNBIP, vol 99*, pp 383–394

- Pinggera J, Furtner M, Martini M, Sachse P, Reiter K, Zugal S, Weber B (2013) Investigating the process of process modeling with eye movement analysis. *Business Process Management Workshops*. Springer, Heidelberg, pp 438–450
- Purchase HC, Carrington D, Allder J-A (2000) Experimenting with aesthetics-based graph layout. In: Anderson M, Cheng P, Haarslev V (eds) *Theory and application of diagrams: 1st international conference, Diagrams, Edinburgh*. Springer, Berlin, pp 498–501. doi:[10.1007/3-540-44590-0_46](https://doi.org/10.1007/3-540-44590-0_46)
- Recker J (2013) Empirical investigation of the usefulness of gateway constructs in process models. *Eur J Inf Syst* 22(6):673–689
- Recker J, Dreiling A (2007) Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In: Toleman M, Cater-Steel A, Roberts D (eds) *Australasian conference on information systems*. The University of Southern Queensland, pp 356–366
- Recker J, Dreiling A (2011) The effects of content presentation format and user characteristics on novice developers' understanding of process models. *Commun Assoc Inf Syst* 28(1):65–84
- Recker J, Mendling J (2016) The state of the art of business process management research as published in the BPM conference. *Bus Inf Syst Eng* 58(1):55–72
- Recker J, Rosemann M, van der Aalst WMP (2005) On the user perception of configurable reference process models—initial insights. In: *16th Australasian Conference on Information Systems*, Paper 66
- Recker J, Reijers HA, van de Wouw SG (2014) Process model comprehension: the effects of cognitive abilities, learning style and strategy. *Commun Assoc Inf Syst* 34(9):199–222
- Reijers HA, Mendling J (2011) A study into the factors that influence the understandability of business process models. *IEEE Trans Sys Man Cybern Part A* 41(3):449–462
- Reijers HA, Recker JC, van de Wouw SG (2010) An integrative framework of the factors affecting process model understanding: a learning perspective. In: *16th Americas conference on information systems: sustainable IT collaboration around the globe*. Association for Information Systems, Paper 184
- Reijers HA, Freytag T, Mendling J, Eckleder A (2011a) Syntax highlighting in business process models. *Decis Support Syst* 51:339–349
- Reijers HA, Mendling J, Dijkman RM (2011b) Human and automatic modularizations of process models to enhance their comprehension. *Inf Syst* 36(5):881–897
- Riedl R (2009) Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen. *NeuroPsychoEconomics* 4(1):32–44
- Riedl R, Banker RD, Benbasat I, Davis FD, Dennis AR, Dimoka A, Gefen D, Gupta A, Ischebeck A, Kenning P (2010) On the foundations of NeuroIS: reflections on the Gmunden Retreat 2009. *Commun Assoc Inf Syst* 27(1):15
- Rodrigues RDA, Barros MDO, Revoredo K, Azevedo LG, Leopold H (2015) An experiment on process model understandability using textual work instructions and BPMN models. In: *29th Brazilian symposium on software engineering (SBES)*, pp 41–50. doi:[10.1109/SBES.2015.12](https://doi.org/10.1109/SBES.2015.12)
- Rosa ML, Wohed P, Mendling J, ter Hofstede AHM, Reijers HA, van der Aalst WMP (2011) Managing process model complexity via abstract syntax modifications. *IEEE Trans Ind Inform* 7(4):614–629. doi:[10.1109/TII.2011.2166795](https://doi.org/10.1109/TII.2011.2166795)
- Sánchez-González L, García F, Mendling J, Ruiz F (2010) Quality assessment of business process models based on thresholds. In: Meersman R, Dillon T, Herrero P (eds) *On the move to meaningful internet systems: OTM 2010, LNCS*, vol 6426. Springer, Heidelberg, pp 78–95. doi:[10.1007/978-3-642-16934-2_9](https://doi.org/10.1007/978-3-642-16934-2_9)
- Sánchez-González L, García F, Ruiz F, Mendling J (2012) Quality indicators for business process models from a gateway complexity perspective. *Inf Softw Technol* 54(11):1159–1174. doi:[10.1016/j.infsof.2012.05.001](https://doi.org/10.1016/j.infsof.2012.05.001)
- Sandkuhl K, Wiebring J (2015) Experiences from selecting a BPM notation for an enterprise. *Business information systems workshops*. Springer, Heidelberg, pp 126–138
- Sarshar K, Loos P (2005) Comparing the control-flow of epc and petri net from the end-user perspective. *International conference on business process management*. Springer, Heidelberg, pp 434–439
- Schrepfer M, Wolf J, Mendling J, Reijers HA (2009) The impact of secondary notation on process model understanding. In: *The practice of enterprise modeling, 2nd IFIP WG8.1 working conference (PoEM 2009)*, Stockholm, Sweden
- Siegmund J, Kästner C, Apel S, Parnin C, Bethmann A, Leich T, Saake G, Brechmann A (2014) Understanding understanding source code with functional magnetic resonance imaging. In: *36th International conference on software engineering*, ACM, pp 378–389
- Simon HA (1955) A behavioral model of rational choice. *Q J Econ* 69(1):99–118
- Škrinjar R, Bosilj-Vukšić V, Indihar-Štemberger M (2008) The impact of business process orientation on financial and non-financial performance. *Bus Process Manag J* 14(5):738–754
- Soffer P, Wand Y (2007) Goal-driven multi-process analysis. *J Assoc Inf Syst* 8(3):175–203
- Soffer P, Wand Y, Kaner M (2015) Conceptualizing routing decisions in business processes: theoretical analysis and empirical testing. *J Assoc Inf Syst* 16(5):2
- Stitzlein C, Sanderson P, Indulska M (2013) Understanding healthcare processes. An evaluation of two process model notations. In: *Human factors and ergonomics society annual meeting*, vol 1. Sage, New York, pp 240–244
- Storch A, Laue R, Gruhn V (2013) Measuring and visualising the quality of models. In: *IEEE 1st International workshop on communicating business process and software models quality, understandability, and maintainability (CPSM)*, pp 1–8. doi:[10.1109/CPSM.2013.6703084](https://doi.org/10.1109/CPSM.2013.6703084)
- Sweller J (1988) Cognitive load during problem solving: effects on learning. *Cognit Sci Multidiscip J* 12(2):257–285
- Sweller J (2010) Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ Psychol Rev* 22(2):123–138
- Trkman M, Mendling J, Krisper M (2016) Using business process models to better understand the dependencies among user stories. *Inf Softw Technol* 71:58–76. doi:[10.1016/j.infsof.2015.10.006](https://doi.org/10.1016/j.infsof.2015.10.006)
- Turetken O, Rompen T, Vanderfeesten I, Dikici A, van Moll J (2016) The effect of modularity representation and presentation medium on the understandability of business process models in BPMN. In: *International conference on business process management*. Springer, Heidelberg, pp 289–307
- Vom Brocke J, Riedl R, Léger P-M (2013) Application strategies for neuroscience in information systems design science research. *J Comput Inf Syst* 53(3):1–13
- Vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Clevén A (2015) Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. *Commun Assoc Inf Syst* 37(9):205–224
- Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol Bull* 138(6):1172
- Wand Y, Weber R (1995) On the deep structure of information systems. *Inf Syst J* 5(3):203–223

- Weber B, Reichert M, Mendling J, Reijers HA (2011) Refactoring large process model repositories. *Comput Ind* 62(5):467
- Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. *MIS Q* 26(2):13–23
- Weitlaner D, Guettinger A, Kohlbacher M (2013) Intuitive comprehensibility of process models. In: Fischer H, Schneeberger J (eds) *S-BPM ONE—running processes*, vol 360. Communications in computer and information science. Springer, Heidelberg, pp 52–71. doi:[10.1007/978-3-642-36754-0_4](https://doi.org/10.1007/978-3-642-36754-0_4)
- Wilson TD (2000) Human information behavior. *Inf Sci* 3(2):49–56
- Zugal S, Pinggera J, Weber B, Mendling J, Reijers HA (2012) Assessing the impact of hierarchy on model understandability—a cognitive perspective. In: Kienzle J (ed) *Models in software engineering: workshops and symposia at MODELS 2011*, Wellington, New Zealand, Reports and revised selected papers. Springer, Heidelberg, pp 123–133. doi:[10.1007/978-3-642-29645-1_14](https://doi.org/10.1007/978-3-642-29645-1_14)
- Zugal S, Soffer P, Haisjackl C, Pinggera J, Reichert M, Weber B (2015) Investigating expressiveness and understandability of hierarchy in declarative business process models. *Softw Syst Model* 14(3):1081–1103. doi:[10.1007/s10270-013-0356-2](https://doi.org/10.1007/s10270-013-0356-2)