Experimentation in Conceptual Modeling Research: A Systematic Review

Completed Research

Markus Fischer
University of Hagen, Hagen, Germany
markus.fischer@fernuni-hagen.de

Kristina Rosenthal
University of Hagen, Hagen, Germany
kristina.rosenthal@fernuni-hagen.de

Stefan Strecker
University of Hagen, Hagen, Germany
stefan.strecker@fernuni-hagen.de

Abstract

Experiments are increasingly applied as a data generation method in conceptual modeling research. Conceptual models, modeling languages and, in particular, their graphical notations are subject to experimental studies. The present comprehensive survey examines the current state of experimentation in conceptual modeling research. Based on an exhaustive sampling of publications, the study identifies and analyzes 98 publications reporting on experiments in conceptual modeling research published between 2005 and 2018 to develop an organizing overview along essential methodological dimensions. Findings reveal a remarkable variety in theoretical foundations, research designs and experiment procedures, and suggest to rethink hypotheses development and to further develop methodical standards for experimentation in conceptual modeling research.

Keywords

Experiment, Conceptual Modeling, Literature Review.

Introduction

Experimental studies in conceptual modeling research ask, for example, what factors impact viewer's comprehension of a business process model or if a person's prior modeling experience affects the time required to construct a data model (e.g., Recker and Dreiling 2007; Reggio et al. 2015). Such studies complement conceptual modeling research on designing new modeling languages or on devising conceptual reference models (Frank et al. 2014), and, hence, serve an important methodological purpose with regard to design science research (Kampling et al. 2016; Mettler et al. 2014), e.g., when evaluating modeling language designs. Scientific experimentation has been receiving appreciation in conceptual modeling research for the past 15 years with a significant share of experiments reported after 2004. As related literature reviews focus on aspects different from hypotheses development, experimental design and experiment procedure, the main objective of the present work is to provide a structuring overview of experiments in conceptual modeling research, particularly with regard to methodological concerns (e.g., Friedman and Sunder 1994), i.e., (i) *theoretical lenses* referred to and utilized for developing (ii) *hypotheses* underlying an experimental design, and (iii) *independent and dependent variables* constituting the experimental design. A preliminary, selective review of experiments in conceptual modeling research indicated a variety of theoretical lenses informing the development of hypotheses involving a multitude of independent and dependent variables and their conceptualizations (cf. Houy et al. 2014). However, a comprehensive inquiry into these dimensions in experimental modeling research is—to the best of our knowledge—missing at present. A secondary objective of the present review is to inform the discussion on experimentation in conceptual modeling research (following, e.g., Rowe, 2014). To assess the state of recent research, we perform a systematic and comprehensive review of literature reporting on experiments in conceptual modeling research published between 2005 and June 2018 based on a pluralistic search strategy (vom Brocke et al. 2009, p. 2214; Webster and Watson 2002, pp. 14–19).

This literature survey ties in with former research suggesting guidelines for experimental evaluations taking into account independent and dependent variables, participants and experimental procedures (Parsons and Cole 2005) as well as research investigating theories underlying the development of hypotheses on business process model understandability (Houy et al. 2014). Literature reviews related to the present work have a narrower focus on process models and restrict the literature retrieval to model understandability/comprehension as dependent variable in the experimental design (Dikici et al. 2018; Figl 2017). Similarly, the surveys by Houy et al. (2012, 2014) focus on business process model understandability and on theories used or cited in process model understandability research. The recent review by Figl (2017) investigates the influence of primary notation, i.e., the used modeling language (e.g., BPMN) and the influence of secondary notation (e.g., color coding) on one dependent variable (understandability/comprehension). The literature review by Dikici et al. (2018), again, focuses on process model comprehension. In contrast, the present work does neither restrict the literature retrieval to business process models nor to (process) model understandability/comprehension as dependent measure. Different from earlier reviews, the presented study searched for experiments involving static abstractions (e.g., data models), functional abstractions (e.g., dataflow diagrams), dynamic abstractions (e.g., business process models), and mixed abstractions such as object models (e.g., UML Class Diagrams).

Methodological and Theoretical Background

Experimentation in conceptual modeling research transfers the notion and methodological tenets of scientific experimentation in the natural sciences, in particular, in experimental physics (Kuhn 1976), to investigating conceptual models, modeling languages, and their graphical notation under controlled conditions. Scientific experimentation has a long and intricate history and is subject to controversial debate in the history and the philosophy of science (Kuhn 1996)—a broad debate we cannot do justice here (see, e.g., Radder 2009 for an overview). Generally, an experimental design includes hypotheses, independent and dependent variables, participants and procedures (e.g., Parsons and Cole 2005, p. 330). A hypothesis proposes a presumed influence of a specific independent measure (independent variable) on a specific dependent measure (dependent variable). Further, dependent variables serve as operationalization of specific phenomena (e.g., comprehension) which cannot be observed directly (latent constructs). Hypotheses are tested under controlled conditions to determine whether the influence of an independent on a dependent variable can be falsified, or can be regarded as non-falsified and hence, as preliminary confirmed for the time being. In this Popperian school of thought, experimentation aims to falsify hypotheses following the notion of falsification (Popper 1959). In conceptual modeling research, an experiment, generally speaking, means that an experimenter assigns a modeling task to cohorts of experimental subjects, often students, with deliberate variation of independent factors among cohorts, e.g., the graphical notations (concrete syntax of a modeling language) shown to subjects, while deliberately controlling for other factors, e.g., prior modeling experience, to study dependent measures such as the time required to "solve" the modeling task.

Review Design and Conduct

Literature retrieval: As a standalone, systematic literature review (vom Brocke et al. 2009, p. 2207), the present study aims at an exhaustive sampling of publications reporting on controlled experiments in the field of conceptual modeling research. The review builds upon a widely accepted scope of conceptual modeling research involving contributions discussing static, functional, dynamic and mixed abstractions (see Embley and Thalheim 2011; Frank et al. 2014). Complementary search strategies are employed to include publications in journals and conference proceedings as well as in other types of sources such as monographs and anthologies: Keyword searches in electronic databases are combined with selective searches in journals and conference proceedings following, e.g., Webster and Watson (2002). We complement these searches with backward searches based on three recent overview articles identified as "key articles" following, e.g., vom Brocke et al. (2015, pp. 215ff.). The time frame for the searches is set to 2005 up to and including June 2018. The choice of 2005 coincides with emerging interest in experimentation in conceptual modeling research (e.g., Parsons and Cole 2005). The present study is limited to publications in English and German while the searches in electronic databases are performed using search strings in English only. As we refer to a subset of the final sample of publications for the paper at hand, the coverage of the literature review in the present work can be characterized as "exhaustive with selective citation" following Cooper (1988) (a list of all publications in the final sample is available as supplementary material in a data repository, see Fischer et al. 2019).

The initial step of the literature retrieval was performed by keyword searches in electronic databases. Based on the focus and research objectives of the study, we purposefully constructed and tested a search string (vom Brocke et al. 2009, p. 2214): Since the focus is on experimental research, the term experiment/experimental constitutes the first part of the search string. As second part, the phrase conceptual/business process/data/object-oriented model/modeling is utilized to specify the application area of the experiments to the field of conceptual modeling. The conjunction of these two phrases (complemented with the terms in British English) was used as generic database search string:

> "(experiment OR experimental) AND ("conceptual model" OR "business process model" OR "data model" OR "object-oriented model" OR "conceptual modeling" OR "business process modeling" OR "data modeling" OR "object-oriented modeling")"

We performed searches in a cross-disciplinary database, i.e., *EBSCOhost (Business Source Ultimate)* complemented with searches in publication databases on computer science, i.e., the *ACM Digital Library (The ACM Guide to Computing Literature)* and the *IEEE Xplore Digital Library*—to include publications on experiments in the broader context of conceptual modeling research. Overall, the keyword searches resulted in 75 publications after removing duplicates and results not qualifying as research publications as, for example, a short article from a poster session and a summary of a panel discussion.

Complementing database searches, we performed manual searches in a selection of journals and conference proceedings addressing topics including conceptual modeling, yet not covered by the database searches. First, three journals were added to the selection covering sources relevant for the Information Systems and Business Informatics disciplines: Journal on Software and Systems Modeling (SoSyM) and Information Systems (ISSN 0306-4379) as well as Business & Information Systems Engineering (BISE)/WIRTSCHAFTSINFORMATIK as relevant outlet for the German-language Business Informatics community which directly addresses conceptual modeling research. Second, to also account for more recent publications, proceedings of seven conferences were added to the selection whose focus and scope include topics complying with the focus of this study: European Conference on Information Systems (ECIS), International Conference on Advanced Information Systems Engineering (CAiSE), International Conference on Conceptual Modeling (ER), International Conference on Information Systems (ICIS), International Conference on Model-Driven Engineering Languages and Systems (MODELS). The proceedings of the Internationale Tagung Wirtschaftsinformatik (WI) and the Multikonferenz Wirtschaftsinformatik (MKWI) were added to the sources as two important outlets for recent research in the German-language Business Informatics community. Scrutinizing the table of contents of the selected sources in June 2018 and viewing titles, abstracts and, in doubt, the full texts of publications resulted in adding 19 journal articles and 63 articles published in conference proceedings to the sample.

As further step, to also include relevant additional publications not published in journals and conference proceedings, we performed backward searches based on the following three recent overview articles ("key articles"): Dikici et al. (2018), Figl (2017) and Houy et al. (2014). The selection of these articles is based on the following two criteria: (1) A key article focuses on experimentation in conceptual modeling, and (2) a key article synthesizes prior work and, thus, refers to a wide variety of publications entailing different perspectives and research fields. We scrutinized the bibliographies of the three identified key articles for publications not covered so far by reason of publication date, type or source—leading to 25 additional results. After removing duplicates, the literature retrieval at this stage resulted in an *intermediate sample* of 182 publications.

As last step of the literature retrieval, publications were excluded from the intermediate sample if considered outside of the scope of the present study. A publication had to fulfill the following inclusion criteria to be included in the final sample: (1) Original research contribution; (2) reporting on an experiment involving a conceptual model, modeling language or corresponding graphical notation; and (3) involving humans as subjects in the experiment(s)—as the term *experiment* is occasionally used for (agent-based) simulations or other non-experimental research designs. Further, in line with the inclusion criteria, the following exclusion criteria were developed which classified a publication to be outside of the scope of this study. A publication was excluded from this study if it reports on: (1) Exploratory experiments which are rather aimed at generating hypotheses than testing them and, thus, usually do not introduce (in)dependent variables and hypotheses; (2) experiments which serve the comparison of two different (query) languages like DBS Oracle and MySQL; (3) experiments performing automated model testing and not involving human subjects (e.g., Němec 2015); (4) outlines of experiments without reporting on the experiment execution and its results; (5) experiments which are performed in order to validate or evaluate an artifact different from a conceptual model or modeling languages, e.g., a software tool. We further excluded textbooks, editorials, book reviews, tutorials, summaries of, for example, panels or posters and education-related publications like course books since the focus of the present study is on original research contributions. For the exclusion step, all 182 publications in the intermediate sample were reviewed by two researchers and discarded based on the inclusion and exclusion criteria by considering titles, abstracts, and, if in doubt, a review of the full-text was performed. Exclusion of a publication required a consensus among the two researchers. This step of the literature retrieval resulted in a *final sample* of 98 publications (with 96 in English language and 2 in German language) in the focus of the present review.

Literature analysis: To develop a structuring overview of prior work, the literature analysis is first directed at characterizing the final sample via its publication profile. For that, we analyze the publications with regard to publication year and modeling language (cf. Sect. Findings). The literature analysis is then guided by the following dimensions developed and refined while selectively reviewing the intermediate and final sample: (1) Research questions addressed and the development of hypotheses including the theory or theoretical lens informing hypotheses development, for example, the Cognitive Load Theory (Chandler and Sweller 1991) underlying the hypotheses development in Petrusel et al. (2017). Thus, the first dimension of analysis is aimed at identifying the *theoretical lens informing hypotheses development*. (2) To achieve in-depth insights into experiment designs reported in the final sample, we analyze *independent and dependent variables* and corresponding operationalizations in the final sample. (3) Last, the developed and *tested hypotheses* of the reviewed experiments are addressed, i.e., how the investigated independent and dependent variables are related in the experiment design.

Findings

Publication Profile

With 98 publications, the final sample illustrates the considerable extent of experiments in conceptual modeling research. Figure 1 displays the distribution of publications in the final sample by year. It is safe to conclude from the publication numbers per year that experimentation has achieved acceptance in conceptual modeling research and has been accepted into the methodical "toolbox" of modeling researchers.

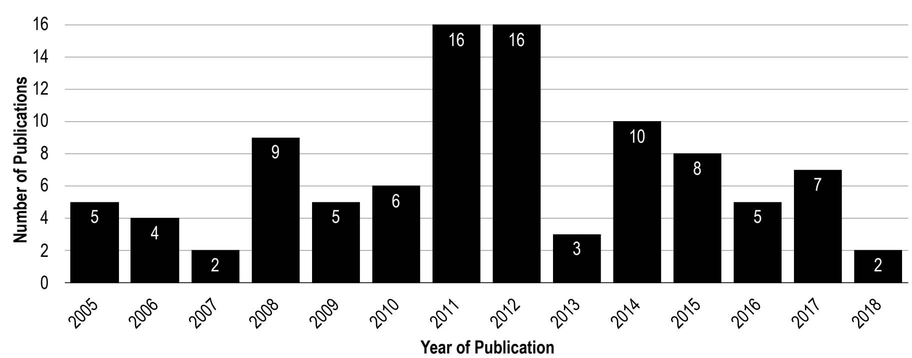

Figure 1. Numbers of Publications in the Final Sample from 2005 to 2018

Regarding modeling languages applied in the reviewed experiments (cf. Figure 2), a prevalence of modeling languages with a primarily focus on (business) process modeling is indicated with 29 assignments to the Business Process Model and Notation (BPMN) and 13 assignments to the Event-driven Process Chain (EPC) while data modeling with the "Entity-Relationship Model" (ER Model) is less often investigated with 14 assignments. The experiments addressing the Unified Modeling Language (UML, 55 assignments in total) split into UML activity diagrams (UML AD, 12 assignments), UML class diagrams (UML CD, 20 assignments), UML object diagrams (UML OD, 1 assignment), UML statecharts (UML SC, 5 assignments), UML sequence diagrams (UML SD, 7 assignments), UML use case diagrams (UML UCD, 6 assignments) and 4 assignments to UML without specifying the used diagram type. We identify 32 different modeling languages (e.g., C3 in Nielen et al. 2011 or the Common Variability Language (CVL) in Reinhartz-Berger et al. 2013, or Basic Message Sequence Chart (bMSC) in Daun et al. 2017) which are only investigated in isolated cases in the review sample and therefore summarized in the category *Others* in Figure 2.

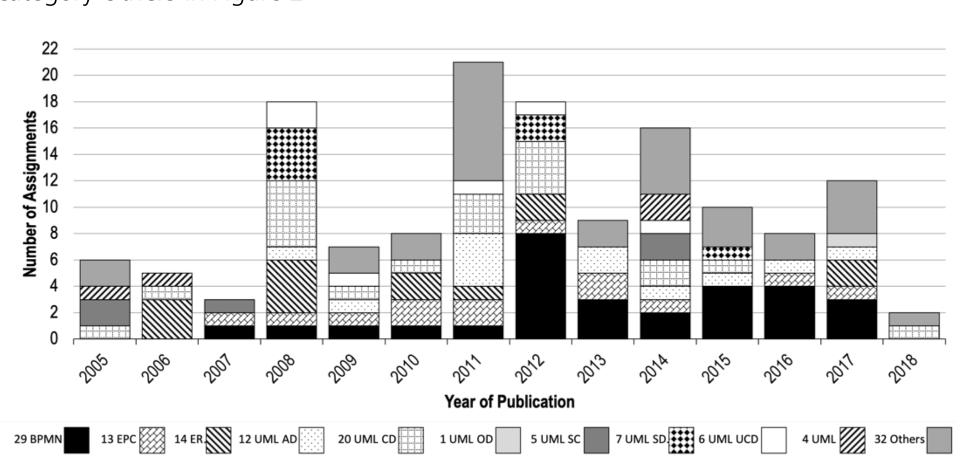

Figure 2. Applied Modeling Languages in the Final Sample According to Year of Publication (Multiple Assignments Possible)

Theoretical Lenses

Less than half of the publications in our review sample, precisely 44 publications, explicitly refer to one or more theory informing hypotheses development (see Table 1)—in several cases, more than one theory is mentioned in the publications (e.g., Cognitive Load Theory, Ornstein's left-brain/right-brain theory, cognitive style index in Turetken et al. 2017). The remaining 54 publications do not explicitly discuss theoretical lenses underlying hypotheses development. It becomes apparent that theoretical lenses in experimentation in conceptual modeling research originate from different scientific fields including Information Systems, Cognitive Psychology and Philosophy (e.g., Bunge's Ontology). With 38 assignments, theoretical lenses referring to theories originating from the field of Cognitive Psychology are prevalent in the review sample which we subsume under the term *Cognitive Dimension*, including the Cognitive Load Theory (e.g., Chandler and Sweller 1991) with 13 assignments, the Cognitive Fit Theory (e.g., Vessey 1991) with 9 assignments, Cognitive Theory of Multimedia Learning (e.g., Mayer 1989) with 7 assignments and publications referring to other cognitive aspects (e.g., cognitive dimensions framework for notational systems in Figl et al. 2013; Poels et al. 2011) with 8 assignments. Epistemological or ontological considerations are much less represented in the review sample (e.g., Evermann and Wand 2006) with 12 assignments referred to as *Ontological Dimension*. We subsume several other theories which are not directly related to cognitive, epistemological or ontological considerations under the term *Others* (e.g., Ornstein's left-brain/right-brain theory) with 9 assignments. We identify a small subset of 11 publications which mention but do not further specify the theoretical lens(es) informing hypotheses development (e.g., Beimel and Peleg 2010).

(In)Dependent Variables and Hypotheses

Independent variables are made explicit in 72 publications in the review sample while 26 publications do not explicitly discuss the investigated independent variables. In order to provide a structuring overview, collecting and clustering independent variables used in the reported experiments led us to establish the following three categories of investigated independent variables: (1) *Model- or language-related* comprises all factors directly related to the conceptual model (e.g., a model's size or complexity, e.g., Sánchez-González et al. 2012) or the applied modeling language constituting the largest group with 75 assignments; (2) *Individual* comprises all factors directly related to the individual experience or knowledge of the modeler, e.g., (self-rated) familiarity with a notation (e.g., Petrusel et al. 2017) with 26 assignments to this category; (3) *Situational* comprises all situational or social aspects related to the activity of modeling (e.g., social distance between modeler and model recipient, e.g., in Kolb et al. 2014) which constitutes the smallest group with 11 assignments. The used categorization ties in with prior publications on experiments in conceptual modeling research. In Dikici et al. (2018, p. 119) process model and personal factors are suggested to influence process model understandability indicators. The category *model- or language-related* corresponds with process model factors while the category *individual* corresponds with personal factors in Dikici et al. (2018, p. 119). The study in Figl (2017, pp. 47–48) builds on the categories personal, model and content factors introduced in Mendling and Strembeck (2008) and investigates additional dimensions (presentation medium, notation, secondary notation, characteristics of the process models, labels, the users, and the types of comprehension tasks). The category *individual* corresponds with personal while the category *model- or language-related* corresponds with the category content. We are aware that the segmentation of independent variables into three categories is a broad distinction accompanied by the limitation that clearly assigning variables may not succeed in all cases. However, we consider the categories tying in with prior work as helpful analytical element reducing complexity and enabling a structuring overview of investigated independent variables.

Table 1. Identified Theoretical Lenses

Category	*Theoretical Lens*	*Sum of Assignments*
Cognitive Dimension	Cognitive Load Theory	13
	Cognitive Fit Theory	9
	Cognitive Theory of Multimedia Learning	7
	Cognitive Style Index	1
	Other cognitive foundations	8
Ontological Dimension	Bunge-Wand-Weber-Ontology	4
	Bunge's Theory of Ontology	3
	Theory of Ontological Clarity	2
	Bunge's ontological notation of property precedence	1
	Theory of Ontological Expressiveness	1
	"Theory of Ontology"	1
Others	Comprehensibility	1
	Construal Level Theory	1
	Dual Coding Theory	1
	Feature Integration Theory	1
	Ornstein's left-brain/right-brain theory	1
	Pattern Recognition	1
	Usability Theory	1
	Short references to cognitive load and semiotic theory	1
	Theory of effective visual notation	1

Dependent variables in hypotheses are made explicit in 72 publications in the review sample while 26 publications do not explicitly discuss the investigated dependent variables. As with the independent variables, collecting and clustering dependent variables used in the reported experiments led us to group the variables in categories. We established three main categories and one subcategory of dependent variables of which the three main categories correspond with the categories of independent variables: (1) *Model- or language-related* comprises all dependent variables directly connected to the used model or modeling language (e.g., number of syntactical errors as operationalization of a model's syntactic quality, e.g., Kolb et al. 2014) constituting the majority with 55 assignments; (2) *Individual* comprises variables connected with personal capabilities (e.g., a person's cognitive style, e.g., Turetken et al. 2017) with 50 assignments. To further investigate the important individual factor comprehension, we decided to build a subcategory *Comprehension* in this category comprising dependent variables directly related to participants' comprehension (e.g., understanding accuracy, e.g., Reijers et al. 2011) with 52 assignments. In our interpretation, the individual factor "comprehension" marks an important category of dependent variables, since it is one goal of experimentation in conceptual modeling research to understand how modelers comprehend given models. (3) *Situational* comprises situational or social aspects related to the activity of modeling (e.g., a laboratory environment or aural silence, e.g., Luebbe and Weske 2011) marking the smallest group with 3 assignments. Further, we identify 14 publications for which assigning dependent variables to one of the categories does not succeed by reason of missing information (e.g., Kayama et al. 2015). Although this categorization is a broad distinction and assigning dependent variables may not succeed in each case, it serves as a useful analytical distinction to reduce complexity.

We identify a variety of operationalizations of dependent variables: For example, model- or language-related dependent variables are measured via references to the number of elements and via syntactical considerations according to the investigated modeling language. Further, it is assumed that modelers with long-time experience in modeling are faster in fulfilling a modeling task so that the experience as an individual factor is operationalized via time taken for the process of modeling. Situational dependent variables are, e.g., measured via the experiments' circumstances (e.g., experiment in a laboratory or in the field). Comprehension is, for example, measured as comprehension level which means a subject's comprehension of a business process and comprehension effort which is operationalized via, e.g., the time taken for a task.

Hypotheses are presented in 81 publications in the review sample while 17 publications do not explicitly state investigated hypotheses (e.g., Allen and March 2012). Table 2 displays categories of independent variables and theoretical lenses underlying hypotheses development and the combinations with categories of dependent variables. Apparently, theories related to the cognitive dimension are often investigated in combination with model- or language-related dependent and independent variables. Further, the influence of individual independent variables on individual dependent variables is investigated in several cases. Remarkably, the influence of model- or language-related independent on individual dependent variables seems to be of special interest which may relate to the fact that model- or language-related variables (e.g., diagram size) can be changed easier than individual variables (e.g., a person's experience with a modeling notation). It becomes apparent that the majority of experiments in the review sample focuses on model- or language-related and individual factors and refers to theories from cognitive sciences like cognitive psychology (e.g., Cognitive Load Theory) or does not mention a theoretical lens. Please note that there may be combinations of variables and theories in Table 2 which do not seem reasonable, e.g., an experiment testing a correlation between an individual independent and a situational dependent variable informed by a cognitive theory. However, pursuing an open coding of experiments we did not exclude combinations a priori.

Table 2. Categories of Independent and Dependent Variables and Theoretical Lenses in the Review Sample (Multiple Assignments Possible)

Categories of Independent Variables / Categories of Dependent Variables	Model- or language-related				Individual				Situational				No assignment possible			
Theoretical lenses	CD	OD	OT	NM	CD	OD	OT	NM	CD	OD	OT	NM	CD	OD	OT	NM
Model- or language-related	15	5	2	27	5	2	0	7	3	0	0	4	1	0	0	0
Individual	14	2	3	23	6	1	1	8	3	0	0	3	1	0	1	0
Comprehension	17	4	2	25	7	1	1	6	1	0	0	3	0	0	0	0
Situational	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
Not mentioned	3	0	1	3	3	0	1	1	0	0	0	1	4	0	0	3

CD: Cognitive Dimension; OD: Ontological Dimension; OT: Others; NM: Not Mentioned

Discussion and Conclusion

Experimentation in conceptual modeling research—as exemplified by the analyzed sample—has an emphasis on business process models (e.g., BPMN, EPC, UML Activity Diagrams) with data models (e.g., Entity-Relationship diagrams) and object models (e.g., UML Class Diagrams) much less investigated (cf. Figure 2). In the following, we discuss our findings along the introduced three dimensions of analysis (1) *theoretical lenses*, (2) *independent and dependent variables*, and (3) *investigated and tested hypotheses.* Suggestions for paths for future research are developed to inform the discussion on experimentation in conceptual modeling research.

Reviewing *theoretical lenses* underlying hypotheses development indicates that experimenters in the review sample refer to a variety of theories or theoretical lenses with a focus on *cognitive theories* (e.g., Cognitive

Load Theory, Cognitive Fit Theory) originating from Cognitive Psychology (cf. Table 2)—which is not surprising in the light of demands on a modeler's cognitive capabilities. Strikingly, a number of experiments in the review sample do not refer to any theoretical foundation for hypotheses development (e.g., Sánchez-González et al. 2012). Equally noticeable is that we do not find any theories or "theoretical lenses" specifically tailored to conceptual modeling. There is also no debate on whether experimentation studies aim to contribute to such theories, and if so, how they could lead to them. Hence, the present state of research is presumed to benefit strongly from a methodological and epistemological debate on experimentation in conceptual modeling research.

Reviewing *independent and dependent variables* and *investigated and tested hypotheses* suggests an exceptional diversity of investigated variables and corresponding operationalizations in the review sample, as exemplified by the prevalent dependent variable "comprehension". Contrary to methodological demand, we encounter publications in the review sample which do not explicitly document hypotheses at all (e.g., Gassen et al. 2015)—another noticeable and somewhat unexpected finding to be further discussed in the methodological debate. The findings reveal that especially model- and language-related variables (e.g., diagram size or gateway complexity) are a particular focus in modeling experiments (cf. Table 2) which is in line with identified theoretical lenses focusing on cognitive capabilities (cf. Table 1). However, the diversity in documenting and investigating, e.g., divergent operationalizations of constructs such as model comprehension, challenges comparability of results and replicability of studies (e.g., Davis and Holt 1993, p. 14; Dennis and Valacich 2014). As one consequence, the study at hand provides a structuring overview instead of a meta-analysis of experiments in conceptual modeling research. Meta-analytical procedures presuppose similar conceptualizations and methodical standards (e.g., King and He 2005)—requirements not met in the identified review sample.

Reviewing 98 publications on experiments in conceptual modeling research published between 2005 and 2018 reveals a remarkable variety in theoretical foundations, research designs and experiment procedures. Rather than facilitating study replicability, experiment documentation lacks commonly accepted standards as presented findings indicate. Observed diversity in experiment execution and documentation impedes cumulative research and falsification attempts difficult. Methodologically, however, replicability and falsification are essential to the experimental method (e.g., Davis and Holt 1993, p. 14; Dennis and Valacich 2014).

Complementing prior work (e.g., Figl 2017, p. 62), present findings lead us to suggest future work to rethink how hypotheses are developed and justified, and to further develop and establish methodical standards for experimentation in conceptual modeling research. As part of the latter, it appears advisable to develop a shared understanding of key concepts (e.g., model understanding/comprehension) and to establish corresponding operationalizations in experimental studies—to improve comparability and traceability of results which benefits researchers designing and conducting experiments as well as practitioners interpreting study findings.

The present literature review is subject to limitations: (1) As a principle limitation, the literature retrieval does not necessarily lead to a complete census of relevant literature—despite a systematic and purposeful sampling of publications (vom Brocke et al. 2009, p. 2207). There is always the risk of misleading decisions, for example, in selecting databases and sources or that we have overlooked publications on experiments which do not mention conducting an experiment in the title or abstract. Furthermore, the filtering process of excluding publications outside of the focus of the study entails the risk of misleading decisions, i.e., to have erroneously misjudged an excluded publication. (2) The set time frame for the literature retrieval neglects reports on experiments published before 2005 (e.g., Bodart et al. 2001; Wand et al. 1999). (3) The identified key articles used for conducting backward searches focus on (business) process modeling, (business) process model understanding (dynamic abstraction) as well as on theories used or cited in process model understandability research and do *not* cover functional, static or mixed abstractions. To the best of our knowledge, there are no further publications available focusing on, e.g., static abstractions which would qualify as key articles. (4) Publications in the review sample show heterogenous research questions as well as varying designs and documentations of experiments. However, the presented findings relate the divergent studies to each other. As a consequence, we suggest future research should investigate the documentation of experiments in conceptual modeling research in more detail to further examine the challenges of comparison and replication. (5) Further, the investigated dimensions of analysis (theoretical lenses, (in)dependent variables, and tested hypotheses) cover experimental research designs only partially (e.g., recruitment of experimental subjects, or incentives are not investigated in the present review). The developed dimensions of analysis and corresponding categorizations in the review at hand assist in providing the structuring overview. Research following-up this literature study may investigate other specific aspects of experiments in conceptual modeling research and elaborate on different categorizations to deepen insights into those specific aspects.

Acknowledgements

We would like to express our gratitude to Elisabeth Böhmer for her support during literature retrieval. We would also like to thank the Minitrack Co-Chair and reviewers for their constructive comments and suggestions which helped us to improve the present article.

REFERENCES

Allen, G. N., and March, S. T. 2012. "A Research Note on Representing Part-Whole Relations in Conceptual Modeling," MIS Quarterly (36:3), pp. 945–964.

Beimel, D., and Peleg, M. 2010. "The Context and the SitBAC Models for Privacy Preservation–An Experimental Comparison of Model Comprehension and Synthesis," IEEE Transactions on Knowledge and Data Engineering (22:10), pp. 1475–1488.

Bodart, F., Patel, A., Sim, M., and Weber, R. 2001. "Should the Optional Property Construct Be Used in Conceptual Modeling? A Theory and Three Empirical Tests," Information Systems Research (12:4), pp. 384–405.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in ECIS, S. Newell, E. A. Whitley, N. Pouloudi, J. Wareham, and L. Mathiassen (eds.), pp. 2206–2217.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," Communications of the Association for Information Systems (37:9), pp. 205–224.

Chandler, P., and Sweller, J. 1991. "Cognitive Load Theory and the Format of Instruction," Cognition and Instruction (8:4), pp. 293–332.

Cooper, H. M. 1988. "Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews," Knowledge in Society (1:1), pp. 104–126.

Daun, M., Brings, J., and Weyer, T. 2017. "On the Impact of the Model-Based Representation of Inconsistencies to Manual Reviews," in Conceptual Modeling (Vol. 10650), H. C. Mayr, G. Guizzardi, H. Ma, and O. Pastor (eds.), Cham: Springer International Publishing, pp. 466–473.

Davis, D. D., and Holt, C. A. 1993. Experimental Economics, Princeton (NJ), USA: Princeton University Press.

Dennis, A. R., and Valacich, J. S. 2014. "A Replication Manifesto," AIS Transactions on Replication Research (1:1), pp. 1–4.

Dikici, A., Turetken, O., and Demirors, O. 2018. "Factors Influencing the Understandability of Process Models: A Systematic Literature Review," Information and Software Technology (93), pp. 112–129.

Embley, D. W., and Thalheim, B. (eds.). 2011. Handbook of Conceptual Modeling: Theory, Practice, and Research Challenges, Berlin ; Heidelberg: Springer.

Evermann, J., and Wand, Y. 2006. "Ontological Modeling Rules for UML: An Empirical Assessment," Journal of Computer Information Systems (46:5), pp. 14–29.

Figl, K. 2017. "Comprehension of Procedural Visual Business Process Models," Business & Information Systems Engineering (59:1), pp. 41–67.

Figl, K., Recker, J., and Mendling, J. 2013. "A Study on the Effects of Routing Symbol Design on Process Model Comprehension," Decision Support Systems (54:2), pp. 1104–1118.

Fischer, M., Rosenthal, K., and Strecker, S. 2019. "Supplementary material for 'Experimentation in Conceptual Modeling Research: A Systematic Review'," URL: https://doi.org/10.5281/zenodo.2649028.

Frank, U., Strecker, S., Fettke, P., vom Brocke, J., Becker, J., and Sinz, E. 2014. "The Research Field 'Modeling Business Information Systems': Current Challenges and Elements of a Future Research Agenda," Business & Information Systems Engineering (6:1), pp. 39–43.

Friedman, D., and Sunder, S. 1994. Experimental Methods: A Primer for Economists, Cambridge (MA), USA: Cambridge University Press.

Gassen, J. B., Mendling, J., Thom, L. H., and De Oliveira, J. P. M. 2015. "Towards Guiding Process Modelers Depending upon Their Expertise Levels," Lecture Notes in Business Information Processing (215), pp. 69–80.

Houy, C., Fettke, P., and Loos, P. 2012. "Understanding Understandability of Conceptual Models–What Are We Actually Talking About?," in International Conference on Conceptual Modeling, Springer, pp. 64–77.

Houy, C., Fettke, P., and Loos, P. 2014. "On the Theoretical Foundations of Research into the Understandability of Business Process Models," in ECIS 2014 Proceedings.

Kampling, H., Klesel, M., and Niehaves, B. 2016. "On Experiments in Design Science Research and Theory Development: A Literature Review," in 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa (HI), USA: IEEE, pp. 4454–4463.

Kayama, M., Ogata, S., Nagai, T., Yokoka, H., Masumoto, K., and Hashimoto, M. 2015. "Effectiveness of Model-Driven Development in Conceptual Modeling Education for University Freshmen," in EDUCON (Vol. 2015-April), pp. 274–282.

King, W. R., and He, J. 2005. "Understanding the Role and Methods of Meta-Analysis in IS Research," Communications of the Association for Information Systems (16), pp. 665–686.

Kolb, J., Zimoch, M., Weber, B., and Reichert, M. 2014. "How Social Distance of Process Designers Affects the Process of Process Modeling: Insights from a Controlled Experiment," in Proceedings of the ACM Symposium on Applied Computing, pp. 1364–1370.

Kuhn, T. S. 1976. "Mathematical vs. Experimental Traditions in the Development of Physical Science," The Journal of Interdisciplinary History (7:1), pp. 1–31.

Kuhn, T. S. 1996. The Structure of Scientific Revolutions, (3rd ed.), Chicago (IL), USA: University of Chicago Press.

Luebbe, A., and Weske, M. 2011. "Tangible Media in Process Modeling—A Controlled Experiment," in International Conference on Advanced Information Systems Engineering (Vol. 6741), H. Mouratidis and E. Rolland (eds.), Springer, pp. 283–298.

Mayer, R. E. 1989. "Models for Understanding," Review of Educational Research (59:1), pp. 43–64.

Mendling, J., and Strembeck, M. 2008. "Influence Factors of Understanding Business Process Models," in Business Information Systems: 11th International Conference, Berlin, Heidelberg: Springer, pp. 142–153.

Mettler, T., Eurich, M., and Winter, R. 2014. "On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework," Communications of the Association for Information Systems (34:10).

Němec, R. 2015. "Assessment of Query Execution Performance Using Selected Business Intelligence Tools and Experimental Agile Oriented Data Modeling Approach," in Proceedings of the Federated Conference on Computer Science and Information Systems (Vol. 5), pp. 1327–1333.

Nielen, A., Költer, D., Mütze-Niewöhner, S., Karla, J., and Schlick, C. M. 2011. "An Empirical Analysis of Human Performance and Error in Process Model Development," in Conceptual Modeling – ER 2011 (Vol. 6998), Berlin, Heidelberg: Springer, pp. 514–523.

Parsons, J., and Cole, L. 2005. "What Do the Pictures Mean? Guidelines for Experimental Evaluation of Representation Fidelity in Diagrammatical Conceptual Modeling Techniques," Data & Knowledge Engineering (55:3), pp. 327–342.

Petrusel, R., Mendling, J., and Reijers, H. A. 2017. "How Visual Cognition Influences Process Model Comprehension," Decision Support Systems (96), pp. 1–16.

Poels, G., Maes, A., Gailly, F., and Paemeleire, R. 2011. "The Pragmatic Quality of Resources-Events-Agents Diagrams: An Experimental Evaluation," Information Systems Journal (21:1), pp. 63–89.

Popper, K. 1959. The Logic of Scientific Discovery. Vienna: Hutchinson.

Radder, H. 2009. "The Philosophy of Scientific Experimentation: A Review," Automated Experimentation (1:1), pp. 1–8.

Recker, J., and 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 ACIS 2007 Proceedings, pp. 356–366.

Reggio, G., Ricca, F., Scanniello, G., Cerbo, F. D., and Dodero, G. 2015. "On the Comprehension of Workflows Modeled with a Precise Style: Results from a Family of Controlled Experiments," Software & Systems Modeling (14:4), pp. 1481–1504.

Reijers, H. A., Freytag, T., Mendling, J., and Eckleder, A. 2011. "Syntax Highlighting in Business Process Models," Decision Support Systems (51:3), pp. 339–349.

Reinhartz-Berger, I., Sturm, A., Clark, T., Cohen, S., and Bettin, J. (eds.). 2013. Domain Engineering: Product Lines, Languages, and Conceptual Models, Berlin, Heidelberg: Springer Vieweg.

Rowe, F. 2014. "What Literature Review Is Not: Diversity, Boundaries and Recommendations," European Journal of Information Systems (23), pp. 241–255.

Sánchez-González, L., García, F., Ruiz, F., and Mendling, J. 2012. "Quality Indicators for Business Process Models from a Gateway Complexity Perspective," Information and Software Technology (54:11), pp. 1159–1174.

Turetken, O., Vanderfeesten, I., and Claes, J. 2017. "Cognitive Style and Business Process Model Understanding," in Advanced Information Systems Engineering Workshops. CAiSE 2017. (Vol. 286), Cham: Springer, pp. 72–84.

Vessey, I. 1991. "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature," Decision Sciences (22:2), pp. 219–240.

Wand, Y., Storey, V. C., and Weber, R. 1999. "An Ontological Analysis of the Relationship Construct in Conceptual Modeling," ACM Transactions on Database Systems (24:4), pp. 494–528.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," MIS Quarterly (26:2), pp. 13–23.