Factors of Process Model Comprehension - Findings from a Series of Experiments

J. Mendling[a,*] M. Strembeck[a] J. Recker[b]

[a] *Wirtschaftsuniversität Wien (WU), Augasse 2-6, 1090 Wien, Austria*

[b] *Queensland University of Technology, 126 Margaret Street, QLD 4000 Brisbane, Australia*

Abstract

In order to make good decisions about the design of information systems, an essential skill is to understand process models of the business domain the system is intended to support. Yet, little knowledge to date has been established about the factors that affect how model users comprehend the content of process models. In this study, we use theories of semiotics and cognitive load to theorize how model and personal factors influence how model viewers comprehend the syntactical information of process models. We then report on a four-part series of experiments, in which we examined these factors. Our results show that additional semantical information impedes syntax comprehension, and that theoretical knowledge eases syntax comprehension. Modeling experience further contributes positively to comprehension efficiency, measured as the ratio of correct answers to the time taken to provide answers. We discuss implications for practice and research.

Key words: Business Process Modeling, Model Comprehension, Experiment;

* Corresponding author

Email addresses: `jan.mendling@wu.ac.at` (J. Mendling), `mark.strembeck@wu.ac.at.` (M. Strembeck), `j.recker@qut.edu.au` (J. Recker).

1 Introduction

In recent years, the documentation of business processes and the analysis and design of process-aware information systems has gained attention as a primary focus of modeling in information systems practice [11]. The so-called practice of process modeling has emerged as a key instrument to enable decision making in the context of the analysis and design of process-aware enterprise systems [12], service-oriented architectures [14], workflow operation [27] and web services [15] alike.

Process models typically capture in some graphical notation the tasks, events, states, and control flow logic that constitute a business process. Process models may also contain information regarding the data that is processed by the execution of tasks, which organizational and IT resources are involved, and potentially capture other artifacts such as external stakeholders and performance metrics, see e.g. [50].

Many benefits are associated with business process modeling. For instance, practitioners have identified process improvement, communication and shared understanding as the most important process modeling benefits [18]. A prerequisite for realizing these benefits, however, is that the quality of process models are perceived as good by their audience, making the *understandability of process models* an important topic for research relevant to all potential uses of process models [3]. Several studies support this view. For instance, the perceived quality of a process model is a key factor contributing to organizational re-design project success [22]. Accordingly, our interest in this papers is to examine how analysts develop an understanding of process models.

More specifically, we study (a) factors characterizing the process model in terms of the activity labels used in the models, (b) factors characterizing the person interpreting the models in terms of relevant modeling expertise, and (c) how these factors affect process model comprehension. The relevance of this research stems from companies making significant investments in process modeling training, with the view of developing a body of process modeling expertise. Indeed, modeler expertise has been established by surveys as an important factor for process modeling success [4] and modeling grammar usage [41]. Furthermore, prior experiments demonstrate that model factors (e.g., an increase in model complexity) affect understanding [48,47]. Notably, these experiments use abstract activity labels (A, B, C etc.) in their process models, which, in turn, raises the question whether the usage of activity labels that carry real domain semantics leverages or impedes understanding.

The aim of the research reported here is to combine these preliminary insights in the definition of a series of experiments. Accordingly, the contributions of this paper are threefold. First, we build on the cognitive load theory to conjecture that real activity labels should decrease syntactical process model understanding. This hypothesis is confirmed in our experiments. Second, we argue in line with prior research that higher modeling expertise results in better understanding performance. This hypothesis is generally confirmed, too. Third, we define different measures of expertise including theoretical knowledge, prior modeling experience, and intensity of modeling. The experiments show that theoretical knowledge is most significant with its impact on performance. Our findings have implications for research on model understanding, in particular regarding cognitive load considerations, and for practice by demonstrating the relevance of theoretical knowledge of process modeling to understanding these models. This insight, in turn, is relevant to informing a staged teaching strategy that educates practitioners about how to read process models.

The rest of this paper is structured as follows. Section 2 introduces the theoretical foundations of process model comprehension. We identify matters of process model understanding and respective challenges. This leads us to factors of understanding. Section 3 describes the research design and Section 4 the results along with a discussion of threats to validity. Section 5 highlights implications for research and practice. Section 6 concludes the article.

2 Background

In this section, we discuss the background of our research. Section 2.1 summarizes which formal conclusions can be drawn from a process model and how understanding performance can be measured. Section 2.2 formalizes our hypotheses.

2.1 Process Model Comprehension

Process modeling has emerged as an important practice to guide decisions in systems analysis and design. In fact, process modeling is the number one reason to engage in conceptual modeling altogether [11], and also considered the number one skill demanded from IT graduates [1]. Analysts develop process models to capture relevant information about a business process they seek to re-design, analyze, or support with an appropriate information system. A business process that is in place to deal with a book order may, for example, contain a task to receive the order, which is followed by another one specifying that the book is to be sent to the customer who ordered it. A model of this process would, therefore, include sequences of graphical elements to describe these tasks and the order in which they have to be performed. Process models can be elicited through interviews with relevant stakeholders, or derived from organizational documents such as business policies [54]. Figures 1 and 2 show two variants of a typical process model, conveying information about important tasks and the control flow that specifies the execution of these tasks.

In reaching an understanding about how individuals comprehend the content of process models, we realize that there is a broad spectrum of matters that can be understood from a process model. The SEQUAL model by Lindland et al. [25], for instance, distinguishes syntactic, semantic, and pragmatic dimensions of model quality. Consider Figures 1 and 2, which show two structurally

[1] http://www.networkworld.com/news/2009/040609-10-tech-skills.html

equivalent process models. The model of Figure 1 contains activities that are labeled with capital letters. Therefore, this model can only be analyzed from a *syntactical* point of view. On the other hand, the model of Figure 2 includes German language activity labels. As these labels point to a specific real-world application domain (i.e., they describe which activities in the real-world domain *specifically* are to be executed), they enable the discussion of the model from a *semantic* point of view. If now this model is communicated in a particular context, e.g. it is communicated as a normative model, then we can also investigate its *pragmatics*. In this way, a process model can represent knowledge for action [23].

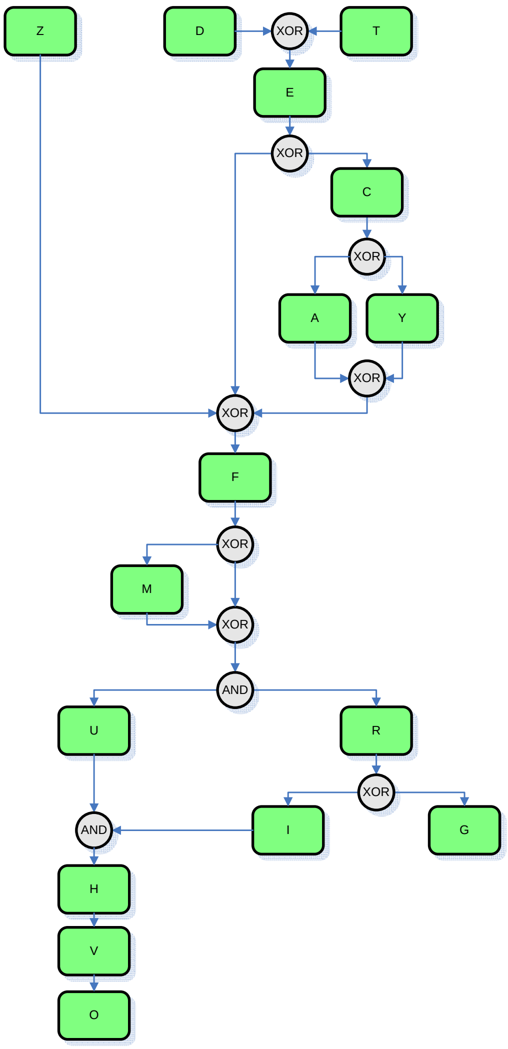

Figure 1. Model 4 with Letters

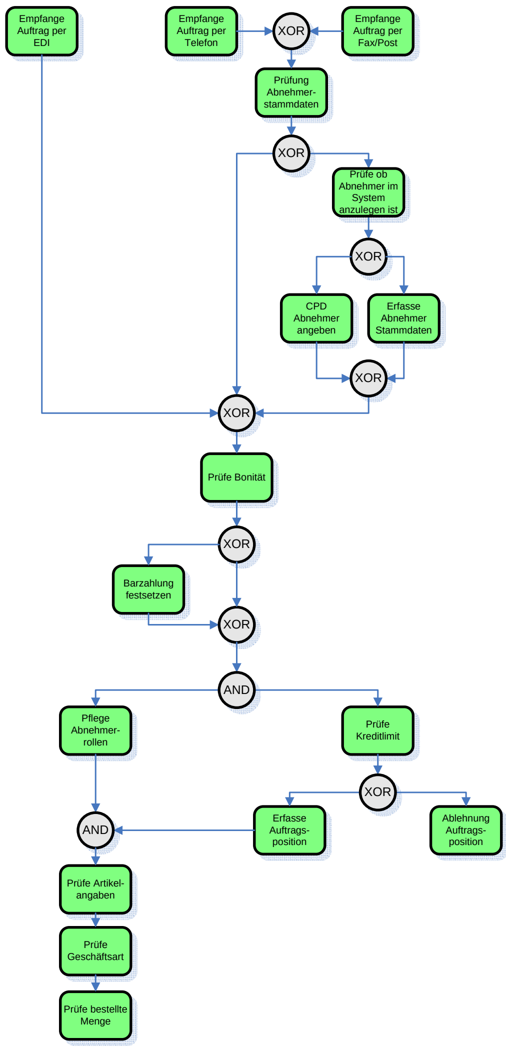

Figure 2. Model 4 with German Text

Semiotic theory postulates that comprehension, and consequently, communication, can be understood as a ladder: syntax (how do I faithfully combine grammatical elements in a process model? [8]) must be clear before semantics can be discussed, and semantics (what do the grammatical elements in a process model mean? [8]) must be clear before pragmatics can be considered. In this regard, it is a primary interest to analyze in how far stakeholders are able to understand process models on a syntactical level. Other interpretations are flawed if syntax is not correctly understood. This is also acknowledged by prior studies that focus on formal and syntactical aspects of process models [46,47].

Looking at which factors influence the comprehension of the syntactical content of process models, prior research has discussed several factors of process model understanding including model purpose [47], problem domain [24], modeling notation [49,16,2], visual presentation [35,40,45], and process model complexity [9,28]. Personal factors, on the other hand, have been less intensively researched to date. This is not to say that no research as been conducted. The experiment by Recker and Dreiling, for instance, operationalized the notion of process modeling expertise through a measure of familiarity with a particular modeling notation [42]. In an experiment by Mendling, Reijers, and Cardoso, participants were characterized based on the number of process models they created and the years of modeling experience they had achieved [31]. This study, furthermore, also indicated the specific importance of theoretical process modeling knowledge. In the latter experiment the participants from TU Eindhoven with strong Petri net education scored better than other participants with less theoretical education in process modeling.

These studies emphasize the value of looking into more details for the impact of expertise, in a sense of *previous experience with modeling*, and in a sense of *knowledge of fundamental process modeling concepts*, which is the intent of our study.

Aside from these important personal factors, we also aim to examine model factors that have not received much attention in prior studies. Specifically, we aim to investigate the effect of semantical information on formal syntactical process model understanding. Therefore, we consider model semantics as expressed in the textual labels, which are used to annotate the graphical activity constructs in a process model (see Fig. 2), and which are important to the usefulness of the models [32]. While one may expect that people might be able to better recall a model with textual information due to a broader activation of different concepts [26], there is an opposite effect to be expected when only questions about syntax are asked. The theoretical rationale for this expectation stems from the cognitive load theory [52]. The main assumptions of the cognitive load theory are limited working memory and its interaction with a practically unlimited long-term memory [52]. When individuals study new material (e.g., information about a business process from a process model) they increase their cognitive load, i.e., the burden on their working memory. This is important because working memory has the capacity to process approximately seven items of information at any given time [34]. Clearly, a long text label in comparison to a single letter implies a higher cognitive load. Textual labels might accordingly distract persons from drawing correct conclusions about formal and syntactical aspects of a process model because a larger share of the working memory is required to process the textual information and the domain information they represent. In this way, a variation of activity labels is an interesting treatment as it should be more detrimental to inexperienced model readers due to the implied cognitive load [53].

On the basis of these theoretical arguments, we define the following research objective: analyze business process models for the purpose of understanding with respect to their syntactical and semantic content from the point of view of model readers in the context of varying prior experience with modeling. Now formalize our expectations in a set of testable hypotheses.

2.2 Hypotheses

In theorizing anticipated effects of the factors discussed above on process model understanding, we first define our operationalization of process model understanding. Similar to [39], we investigate syntactic understanding from two angles, these being *comprehension task performance* (how faithfully does the interpretation of the process model allow the reader to comprehend the formal content of the model?) and *comprehension task efficiency* (what resources are used by the reader to comprehend the process model?). Both factors are important elements in Norman's theory of action [37], and relate to what Norman calls "the gulf of interpretation" (a difference between what the model tries to convey and what is interpreted by the model reader). The gulf of interpretation is an important measure of the performance of modeling efforts, because model comprehension by relevant stakeholders is a necessary prerequisite for various model application tasks, such as systems analysis, communication, design, organizational re-engineering, project management, end user querying and others [44]. In other words, for a model to be useful for any modeling-related task, it is imperative that the stakeholders doing these tasks are able to comprehend the model well (performance) and timely (efficiency).

We now draw hypotheses regarding the effects of personal and model factors on model readers' comprehension task performance and efficiency. Figure 3 shows our research model. The model proposes that process model understanding (in terms of comprehension accuracy and comprehension efficiency) is a function of the characteristics of the model of the process, and of the characteristics of the user interpreting the model.

Our first hypothesis addresses model factors. While prior studies have examined model characteristics such as model structure and complexity [30], our interest is in the textual labels that are used in process models to annotate the graphical constructs. Graphical constructs, and their relationships, are used to convey information about the structure of a process and its formal behavior.

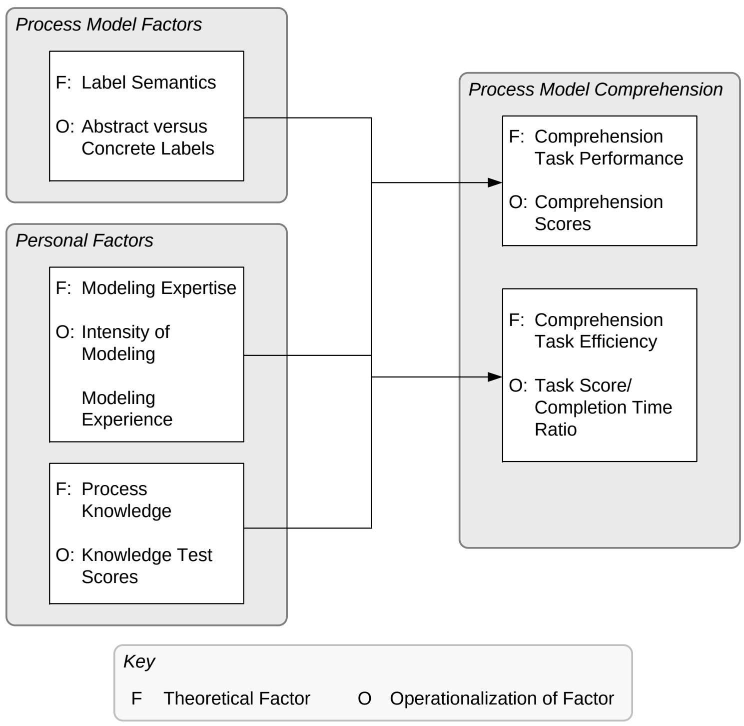

Figure 3. Research model

Textual labels used to annotate the graphical constructs, on the other hand, convey important information about the domain (e.g., what activity has to be performed, what is an important document, who within an organization is responsible for execution, and so forth). Based on this distinction, we expect that model readers will be able to more easily understand the formal, syntactical aspects of a process model, as expressed in the grammatical constructs and their relationships, when they are not presented with additional, semantic information about the application domain (in the textual labels). This is because the textual labels increase the cognitive burden on the model viewer in that the textual labels are an additional set of information material that needs to be processed by the working memory [53], but which is largely irrelevant to the comprehension of the formal content of a process model, which is the interest in our study.

We further expect that comprehension occurs quicker for people working with process models featuring abstract textual labels, because they require less effort to retrieve and assemble pieces of information in their working memory, when only having to consider graphical constructs but not additional textual information. We formalize these observations in the first two hypotheses:

H_0^1 The use of abstract labels will have no impact on comprehension task performance.

H_a^1 The use of abstract labels will have a significant positive impact on comprehension task performance.

H_0^2 The use of abstract labels will have no impact on comprehension task efficiency.

H_a^2 The use of abstract labels will have a significant positive impact on comprehension task efficiency.

Next, we consider personal factors. First, we theorize that individuals with higher levels of knowledge about formal process model concepts such as deadlocks, soundness, concurrency and so forth will achieve better comprehension task performance and efficiency. This is because, when interpreting a process model, these individuals can make use of prior knowledge, i.e., relevant knowledge material stored in long term memory can be applied to reduce the cognitive load on their working memory, which will ease, and improve their understanding of the material (the process model) presented to them. Accordingly, we have:

H_0^3 Users with higher levels of process knowledge will not have higher comprehension task performance.

H_a^3 Users with higher levels of process knowledge will have significantly higher comprehension task performance.

H_0^4 Users with higher levels of process knowledge will not have higher comprehension task efficiency.

H_a^4 Users with higher levels of process knowledge will have significantly better comprehension task efficiency.

Second, we realize that modeling expertise is an important factor in process modeling [4,41]. Experienced modelers often possess a repertoire of workarounds for challenging modeling situations, and can often refer to their previous experiences and knowledge about modeling when attempting to interpret complex models. Less experienced modelers, on the other hand, often lack such knowledge, which, in turn, can be expected to affect their comprehension accuracy and efficiency.

The resource allocation theory [20] suggests that when users build up experience in modeling, their demand for cognitive attentional effort required to perform the model-related tasks is reduced, thereby freeing cognitive resources that can be allocated to improving task performance and outcome production (i.e., better and faster understanding). This situation would suggest that experienced modelers can read process models better and with less effort. We distinguish between modelers that have modeled for a long time (i.e., that have *modeling experience*) and those that model often (i.e., that have *modeling intensity*), to be able to examine modeling experience in a more detailed manner. We state the following hypotheses:

H_0^5 Users with higher levels of modeling experience will have equal comprehension task performance.

H_a^5 Users with higher levels of modeling experience will have significantly higher comprehension task performance.

H_0^6 Users with higher levels of modeling experience will have equal comprehension task efficiency.

H_a^6 Users with higher levels of modeling experience will have significantly better comprehension task efficiency.

H_0^7 Users with higher levels of modeling intensity will have equal comprehension task performance.

H_a^7 Users with higher levels of modeling intensity will have significantly higher comprehension task performance.

H_0^8 Users with higher levels of modeling intensity will have significantly better comprehension task efficiency.

H_a^8 Users with higher levels of modeling intensity will have significantly better comprehension task efficiency.

In the following, we describe design and results of a series of experiments we conducted to test these hypotheses.

3 Experiment Description

For investigating the hypotheses, we define an experiment following established guidelines for experimental software engineering [5,19,55]. Because there is only limited research on cognitive load effects in the process modeling domain, we chose an experimental method as it affords a higher internal validity than other methods [10]. With this experiment definition, we aim to analyze process models for the purpose of understanding with respect to comprehension task performance and comprehension task efficiency. In particular, the analyses are conducted from the perspective of a reader of the model, and the experiment's context is given through persons with process modeling skills answering questions about the meaning of a process model.

3.1 Experiment Design

To test our hypotheses, we selected a 2 x (4 x 4 x 4) mixed balanced experimental design that allowed us to focus on personal factors and model characteristics whilst eliminating potentially confounding other variables (e.g., domain knowledge). Our experimental design featured one between-subjects factor and three within-subjects factors.

3.1.1 Experimental Condition and Tasks

The between-subjects factor, *Label Type*, had two levels. We provided participants with process models that contained either abstract or concrete labels. To operationalize this factor, we gathered a set of six process models from practice that capture business processes in two different domains, order processing and price calculation. The models were provided by a partner organization, which has these models in real use for process documentation purposes. The models were randomly selected from their collection of process models. The models all could be displayed on an A4 page and ranged from nine to twenty activities, and contained between six and fifteen connectors. These characteristics are similar to those found in process model collections in practice [38]. Therefore, we deemed these models to be adequate experimental treatments given that the cases reflect modeling scenarios typically encountered in real-life process modeling practice. Based on the observation in [49] that EPCs appear to be easier to understand than Petri nets, we chose an EPC-like notation without events. The participants received a short informal description of the semantics similar to [29, p. 25]. Finally, we drew all models in the same top-to-bottom style with the start element at the top and end element at the bottom. Altogether, each participant was challenged with four tasks (see Appendix):

(1) self-assess process modeling intensity,
(2) self-assess process modeling experience,
(3) answer theoretical knowledge test, and
(4) answer process model comprehension questions.

3.1.2 Independent Variables

To operationalize the between-subjects factor *Label Type* as an independent variable, for each of the process models used we constructed a variant where the activity labels were replaced by abstract capital letters as identifiers. Figures 1 and 2 depict model number 4 of the models we used in our experiment. For the 6 models we identified 6 yes/no questions related to the structure and the process flow specified by the model. These questions together with questions on personal experience and knowledge of process modeling were packed into two variants of the questionnaire, one for models with original activity labels (*textual labels*), one for models with letters (*abstract labels*).

Aside from the between-subjects factor *Label Type*, we also defined three within-subject factors. The first within-subjects factor *Knowledge* had four levels. The participants had to answer twelve theoretical yes/no questions before seeing the models about selected topics related to process modeling such as choices, concurrency, loops, and deadlocks (see Appendix). These questions concern grammatical rules of process model logic, derived from fundamental work in this area [21] and as previously used in [33]. We transformed the *knowledge score* into an ordinal *knowledge* scale with four levels: very low (0-3 correct answers), somewhat low (4-6 correct answers), somewhat high (7-9 correct answers) and very high (10-12 correct answers). This ordinal measure served as a second independent variable. The second within-subjects factor *Experience* had four levels. The participants were asked for how long they have been involved with business process modeling. The variable was measured on an ordinal scale with four levels: less than one month, less than a year, less than three years, and longer than three years. This measure served as a third independent variable. Finally, the third within-subjects factor *Intensity* also had four levels. The participants had to indicate how often they work with process models. We used an ordinal scale with four options to answer: daily, monthly, less frequent than monthly, never. This measure served as a fourth independent variable.

3.1.3 Dependent Variables

We use two dependent variables, comprehension task performance and comprehension task efficiency. *Comprehension Task Performance* is calculated based on the answers given by the participant to the model comprehension questions. It captures the number of correct answers by the person. The maximum value is 36 for six questions on six models. This measure serves as an operationalization of formal process model understanding of a person.

Comprehension Task Efficiency is based on the task completion time that the participants invested in answering the different questions in the questionnaire. The measure is calculated by dividing the number of correct answers (Comprehension Task Performance) by the time take to complete the respective questions, and served as a second dependent variable in our study.

3.2 Experiment Execution

We implemented the experiment in two ways. First, we defined an online experiment in order to make access to practitioners with modeling experience more easy. The automated system further allowed us to record the answer times, randomly assign the subject to a label type, and randomly define the presentation order of the six models in the corresponding label type, thereby ensuring a balanced treatment. Participation was voluntary. As an incentive the participants received feedback about their test performance.

In 2007, we distributed the link to the experiment via the German mailing lists EMISA and WI as well as among students that followed courses on process modeling at the Vienna University of Economics and Business. Typically, both academics and practitioners with an interest in conceptual modeling and information systems development are registered with these lists. The questionnaire was started by 200 persons and completed by 46. From these 46 we excluded 4 people who spent less than 10 minutes time on the questionnaire since we assumed that to be the minimum time to provide meaningful answers. The remaining 42 persons and their answers to the 36 questions establish the first part of the sample for our statistical analysis below. Altogether, 1512 answers are recorded in the sample. 65% of the participants had more than three years experience in process modeling.

To increase confidence in the conclusion validity of our study, we collected further data with paper-based replications of the experiment. The first replication in April 2009 involved 23 graduate students from Vienna University of Economics and Business who followed a course on modeling. The second sample includes 22 graduate students who followed the same course in June 2009.[2] The third replication was conducted with 32 graduate students who followed the system analysis and design course at Humboldt-Universität zu Berlin. From all four experiments we collected data from altogether 119 persons. With each answering 36 questions, we get 4284 answers to model understanding questions.

These four experiments correspond to a strict replication according to [5], with the variation between the experiments being only in the institution of the participants and the mode of presentation (web versus paper). Because neither institutional affiliation nor mode of presentation are relevant factors in our study, our replication can be considered strict and therefore allows not only combination of experimental results but also pooling of data. To be able to examine any potential threats to validity stemming from the replication, we created two dummy variables, *affiliation*, and *experimentMode*, to examine whether experimental results differed significantly across the replications. Table 1 gives the results. All test results were insignificant, with p values ranging from 0.17 to 0.41, suggesting that none of the relevant data differed significantly for the dummy variables, thereby justifying to our pooling of the data.

Each of the experiments used feedback about the performance as an inducement. While this feedback was meant to be informative to practitioners, it served the students for the preparation towards their exams.

[2] Vienna University of Economics and Business runs the modeling course on a half-semester turn.

Dependent Variable	Dummy variable	Levels	N	Mean	Std. Dev.	Sig.
Comprehension	affiliation	Original study	42	26.26	4.94	0.17
Task		Replication 1	23	25.44	4.02	
Performance		Replication 2	22	26.36	4.28	
		Replication 3	32	25.78	4.90	
	ExperimentMode	Online	42	26.60	4.49	0.23
		Paper	77	25.58	4.25	
Comprehension	affiliation	Original study	42	1.31	0.66	0.27
Task		Replication 1	23	1.22	0.29	
Efficiency		Replication 2	22	1.14	0.25	
	ExperimentMode	Online	42	1.31	0.66	0.41
		Paper	45	1.18	0.28	

Table 1

Test Results Regarding Experiment Replication

4 Data Analysis and Interpretation

In this section, we first discuss distribution and correlation before we turn to hypothesis testing. Last, we discuss threats to validity.

4.1 Distribution and Correlation Analysis

Table 2 shows descriptive statistics for our measures. All results are in line with expectations. Table 3 gives the correlation matrix. First, we check for potential interactions between our between-subject factor (label type) and our within-subject factors (experience, intensity, knowledge). The data in Table 3 clearly shows that no significant interaction terms are present between these

Type of variable	Variable	N	Mean	Std. Dev.	Scale
Independent variables	Knowledge	119	2.66	0.84	1-4
	Label Type	119	1.47	0.50	1/2
	Experience	119	2.75	1.21	1-4
	Intensity	119	2.30	0.95	1-4
Dependent variables	Comprehension Task Performance	119	25.94	4.34	0-36
	Comprehension Task Efficiency	87	1.22	0.52	0-inf.

Table 2

Descriptive Statistics

factors, thereby suggesting independence of the experimental conditions used in our study. The insignificant correlations of the between-subjects factor and the within-subject factors allows to run the hypothesis tests independently. Further inspection of Table 3 suggests that *Label type* and formal process knowledge (*knowledge*) are meaningful independent factors as they correlate significantly with the dependent measures. By contrast, *experience* and *intensity* do not correlate largely with the dependent measures but with each other. This correlation between intensity and experience, however, behaves in accordance with general expectations (in the sense that people that model longer often model more frequently, too). Next, the correlation between intensity and experience to knowledge is expected, as people with more intensive or overall longer process modeling experiences build up higher levels of knowledge about process modeling. The correlations between *comprehension score* and *efficiency*, likewise, were expected. Overall, we do not find counter-intuitive correlations in Table 3. Note that in Table 2 we see that the sample size for the efficiency measure is 87, which is because we failed to accurately record task completion times in our experiment replication with the students in Berlin.

	Label type	Knowledge	Intensity	Experience	Comprehension Task Performance
Knowledge	-0.01				
Intensity	0.08	0.31**			
Experience	0.04	0.28**	0.24*		
Comprehension Task Performance	-0.08	0.42**	0.15	0.15	
Comprehension Task Efficiency	-0.35**	0.16	0.13	-0.11	-0.31**

Table 3

Correlation Matrix. * Correlation is significant at the 0.05 level (2-tailed), ** Correlation is significant at the 0.01 level (2-tailed).

4.2 Testing Hypotheses on Comprehension Task Performance

After screening the data, we now discuss the test of our predictions. We argued in our Hypothesis H_a^1, H_a^3, H_a^5 and H_a^7 that process model comprehension task performance would be positively impacted by

- the use of abstract labels,
- higher levels of formal process knowledge,
- higher levels of process modeling experience, and
- higher levels of process modeling intensity.

As a dependent measure, we used the process model comprehension task performance scores (0-36). We first checked whether the data met the assumption of equal variances in the dependent measures across the levels of each independent variable. Levene's test was insignificant ($F = 1.45, p = 0.19$), indicating that the data met this assumption. Hypothesis testing was completed individually for each of the four independent factors above, using SPSS Version 16.0. First, we performed an Analysis of Variance (ANOVA) for our between-subjects factor *Label Type*. Then, for each of the three factors formal process knowledge, process modeling experience, and process modeling intensity, we used a non-parametric Kruskal-Wallis test to examine our hypotheses, because a Kolmogorov-Smirnov test confirmed that the normality assumption did not hold for these measures, i.e. $Z = 2.51$ (*knowledge*), 2.68 (*experience*), 2.52 (*intensity*), all $p < 0.01$. Therefore, we used the Kruskal-Wallis test, which is accepted as an alternative to ANOVA in case the considered variables are not normally distributed [51]. We examined the hypotheses individually because our correlation analysis suggested independence of the between-subjects and within-subjects factors. Also, our experimental design features three ordinal variables, for which we required non-parametric tests, and the Kruskal-Wallis test we selected considers one independent variable at a time. We chose this test over others (e.g., ANOVA, Mann-Whitney) because, first, the Kruskal-Wallis test is the generalization of the Mann-Whitney test when there are more than two independent groups, like in our study (four levels) [17]. Second, even though we replicated the experiment to gather more data, the number of respondents overall is rather small, and the subgroups for each ordinal scale level are smaller. The distribution-free nature of non-parametric tests places few restrictions on the sample size in contrast with parametric tests, which rely on asymptotic properties or normality of the sample distribution [51]. Third, the ordinal measures used in our study called for the use of non-parametric methods, which yield higher power than corresponding parametric tests (e.g., ANOVA) [36]. Finally, rank-based non-parametric tests are not affected by outliers [17], which allows us to also consider those data where respondents took unusually long (or short) for answering the experimental questions. Table 4 gives the descriptive results and Table 5 gives the results from the statistical tests.

Perusal of the data in Table 4 and Table 5 leads to the following observations.

Differences among groups	Treatment Group	N	Mean	Std. Dev.	Mean Rank
Label Type	Abstract Labels	62	26.35	4.06	N/A
	Textual Labels	56	25.48	4.67	N/A
Knowledge	Very low	9	24.78	2.44	43.78
	Somewhat low	41	23.80	4.66	45.42
	Somewhat high	49	26.57	3.77	63.93
	Very high	19	29.47	3.10	89.79
Experience	Less than one month	28	24.39	4.65	48.58
	Less than a year	20	26.25	4.27	58.54
	Less than three years	23	26.78	3.87	71.22
	Longer than three years	47	26.32	4.36	60.33
Intensity	Never	26	24.81	3.38	46.09
	Less than monthly	45	25.56	4.47	62.85
	Monthly	32	27.56	4.23	63.67
	Daily	15	25.60	5.24	64.02

Table 4

Descriptive Results of Model Comprehension Task Performance Scores

H_a^1 hypothesized higher comprehension task performance scores for the group of users working with models with abstract labels. Table 4 shows that the average comprehension task performance scores indeed were higher (mean score = 26.45 vs. 25.48), and Table 5 confirms that the differences are significant ($F = 5.05, p = 0.03$). These results lead to the rejection of null hypothesis H_0^1 and suggest people viewing models with no textual labels achieve a higher level of comprehension of formal syntactic aspects of process models.

Independent factor	df	Statistic	Sig.
Label Type	1	5.05	0.03
Theory	3	24.48	0.00
Experience	3	6.37	0.10
Intensity	3	5.70	0.13

Table 5

Test Results of Model Comprehension Task Performance Scores

H_a^3 hypothesized higher comprehension task performance scores for users with higher levels of formal process knowledge. And indeed, we observe that comprehension task performance scores were higher, relatively, for users with very high knowledge levels, over those with somewhat high, and somewhat low knowledge (means = 29.47, 26.57 and 23.80). [3] Table 5 suggests that the comprehension task performance across the four groups is significantly different ($Chi{-}2 = 24.48, p = 0.00$). We note, interestingly, that the group of users with very low knowledge performed somewhat better than the group with somewhat low knowledge (mean = 24.78). A follow-up ANOVA analysis of these two groups, however, showed these differences to be insignificant. A second-follow up ANOVA analysis of comprehension task performance based on the actual comprehension task performance scores (0-12) also yielded significant results ($df = 11, F = 2.05, p = 0.03$). Therefore, we suggest to reject the null hypothesis and tentatively accept hypothesis H_a^3.

H_a^5 and H_a^7 hypothesized higher comprehension task performance scores for users with higher levels of modeling expertise (in the sense of modeling experience and intensity). Table 4 shows that the comprehension task performance scores for the four groups of users (for both experience and intensity) follow an inverse U-shaped curve in that task scores increase for the users with very low, somewhat low, and somewhat high expertise (both for experience and intensity) but drop for the groups of users classified as very experienced/very

[3] Note that higher rank scores indicate higher comprehension task performance.

intensive. The results from the Kruskal-Wallis test in Table 5 show, furthermore, that group differences for both factors experience and intensity are insignificant ($Chi-2 = 6.37, p = 0.10$ and $Chi-2 = 5.70, p = 0.13$). In light of these results, we cannot reject the null hypotheses H_0^5 and H_0^7, suggesting that modeling expertise is not an important factor in explaining process model comprehension task performance.

4.3 Testing Hypotheses on Comprehension Task Efficiency

Next, we argued in our Hypothesis H_a^2, H_a^4, H_a^6 and H_a^8 that process model comprehension task efficiency (measured by the normalized ratio between comprehension task performance and comprehension task completion times) would be positively impacted by

- the use of abstract labels,
- higher levels of formal process knowledge,
- higher levels of process modeling experience, and
- higher levels of process modeling intensity.

Because during our conduct of the experiment at Humboldt-Universität zu Berlin we were unable to accurately record time measures for comprehension tasks, for this second analysis we had to exclude 32 entries from our data set, resulting in an effective sample size of 87. Again, we first checked whether the data met the assumption of equal variances in the dependent measures across groups. Levene's test was insignificant ($F = 1.30p = 0.08$), indicating that the data met this assumption. Hypothesis testing was completed in the same manner as above, using the same four measures as independent factors. As a dependent measure, we used the process model comprehension task efficiency scores. The descriptive analysis results are displayed in Table 6 and Table 7.

Perusal of the data in Table 6 and Table 7 leads to the following observations.

Differences among groups	Treatment Group	N	Mean	Std. Dev.	Mean Rank
Label type	Abstract Labels	44	1.39	0.60	N/A
	Textual Labels	42	1.03	0.32	N/A
Formal knowledge	Very low	9	1.34	0.39	54.50
	Somewhat low	33	1.08	0.40	48.92
	Somewhat high	33	1.24	0.42	65.98
	Very high	11	1.51	0.85	71.68
Modeling experience	Less than one month	16	1.36	0.49	69.81
	Less than a year	13	1.29	0.64	53.10
	Less than three years	16	1.01	0.60	62.83
	Longer than three years	41	1.21	0.44	58.13
Modeling intensity	Never	14	1.09	0.30	74.41
	Less than monthly	37	1.19	0.58	64.22
	Monthly	23	1.28	0.49	52.74
	Daily	12	1.30	0.58	51.91

Table 6

Descriptive Results of Model Comprehension Task Efficiency Scores

H_a^2 hypothesized better comprehension task efficiency scores for the group of users working with models with abstract labels. Table 6 shows that the average comprehension task efficiency score, i.e., the ratio between correct answers and time taken to complete the answers, indeed were lower for this group (mean score = 1.39 vs. 1.03). Table 7 shows that the group differences are significant ($F = 3.90, p = 0.05$). Therefore, the results suggest rejecting null hypothesis H_0^2, which means that textual semantics, being a significant factor for how well people understand the formal content of process models,

Independent Factor	df	Statistic	Sig.
Type	1	3.90	0.05
Theory	3	8.38	0.04
Experience	3	4.29	0.23
Intensity	3	9.09	0.03

Table 7

Test Results of Model Comprehension Task Efficiency Scores

also significantly affects the effort that is required to reach this understanding.

H_a^4 hypothesized better comprehension task efficiency scores for the group of users working with higher levels of formal process knowledge. We note from Table 7 that the differences in comprehension task efficiency across the groups of users with different levels of knowledge are significant ($Chi - 2 = 8.38, p = 0.04$), and from Table 6 that the efficiency scores are better for users with higher levels of knowledge. We note, however, that Table 6 also shows a somewhat unexpected exception. The group of users with low levels of knowledge completed their tasks the with the second-best efficiency score (mean = 1.34), superseded only by those with high levels of knowledge (mean = 1.51). We note that these results may have been over-compensated through quick task completion, independent from correct results (as shown in Table 4). Indeed, it seems plausible that users with low knowledge levels just quickly selected answers without engaging in a thorough consideration of the content presented to them. Overall, the results are in line with our expectations, the null hypothesis H_0^4 is rejected.

H_a^6 and H_a^8 hypothesized better comprehension task efficiency scores for users with higher levels of modeling expertise (in the sense of modeling experience and intensity). We note from Table 7 that the differences in task completion efficiency across the user groups with different levels of modeling intensity are significant ($Chi - 2 = 9.09, p = 0.03$), and provide the correct directionality (means = 1.09, 1.19, 1.28 and 1.30). The results support hypothesis H8a. For modeling experience, however, the results are not in line with hypothesis H6a. There are fluctuations in comprehension task efficiency scores noted in Table 6 (means = 1.36, 1.29, 1.01 and 1.21), and the Kruskal-Wallis tests suggests that the differences across the groups are insignificant ($Chi - 2 = 4.29, p = 0.23$). Therefore, we cannot reject null hypothesis H_0^6.

4.4 Discussion of Results

Our experimental study provides support for five out of eight hypothesized factors of process model comprehension task performance and efficiency (see Table 8). The results for hypotheses H_a^1 and H_a^2 suggest that a plus in semantical information in terms of text labels seems to be a burden when analyzing the syntactical content of a process. These findings are in line with arguments that are founded on the grounds of cognitive load theory as well as the premise of the semiotic ladder. Hypotheses H_a^3 to H_a^8 are interesting to be discussed relative to each other. Theoretical knowledge turned out to be a strong indicator for both comprehension task performance and efficiency on syntax-related comprehension of process models (H_a^3 and H_a^4). In contrast, modeling experience and intensity were found not to contribute significantly to either comprehension task performance or efficiency, set aside the result obtained in relation to hypothesis H_a^8. We interpret this result as an indication that theoretical knowledge is of paramount importance to understanding syntactical aspects of a process model, over and above any practical experience with the exercise of process modeling. Indeed, the non-significance of experience and intensity here might suggest that these factors are more important for the semantical interpretation of process models and that theory is the prerequisite for understanding syntax.

Hypothesis	Result
H_a^1: Label Type → Comprehension Task Performance	Supported
H_a^2: Label Type → Comprehension Task Efficiency	Supported
H_a^3: Knowledge → Comprehension Task Performance	Supported
H_a^4: Knowledge → Comprehension Task Efficiency	Supported
H_a^5: Experience → Comprehension Task Performance	Not Supported
H_a^6: Experience → Comprehension Task Efficiency	Not Supported
H_a^7: Intensity → Comprehension Task Performance	Not Supported
H_a^8: Intensity → Comprehension Task Efficiency	Supported

Table 8

Summary of Hypotheses Tests

4.5 Threats to Validity

The results of this experiment have to be discussed against different threats to validity. We focus on those threats of [55, p. 67] that are most relevant for our experiment.

Conclusion validity is concerned with the relationship between treatment and outcome, and the conclusions drawn from it. Two aspects have to be considered: The first aspect concerns the appropriateness of the statistical tests. As reported above, we have screened our data for conformance with the assumptions of the statistical tests we used (ANOVA, Kruskal-Wallis test). We used Levene's test to show that the dependent variables across the treatment groups shared approximately equal variance. We used the non-parametric Kruskal-Wallis test for our ordinal measures because the independent data was not normally distributed. A Kolmogorov-Smirnov test confirmed that the normality assumption did not hold for the measures *knowledge*, *experience*, or *intensity* (Z = 2.51, 2.68, 2.52, all p = 0.00). Therefore, we used the Kruskal-Wallis test, which is accepted as an alternative to ANOVA in case the considered variables are not normally distributed [51]. The second aspect concerns the effect sizes of the results. In order to reach a sample size sufficient to solve potential issues regarding the statistical significance, we conducted strict replications [5] of our experiment. In order to show that our replications did not induce bias into our analysis, we created two dummy variables, *affiliation* and *experimentMode*, to examine whether experimental results differed significantly across the replications. Affiliation with one of the universities partaking in our study did not affect results for comprehension task performance or task completion time - the Kruskal-Wallis test was insignificant ($p = 0.16$ and $p = 0.09$). The mode of experiment (paper versus online), likewise, was an insignificant factor, as shown in an independent samples t-test ($p = 0.20$ and $p = 0.80$ for comprehension task performance and task completion time).

Internal validity demands that the treatment causes the effect. In order to avoid maturation and learning effects, we used a random sampling of the questions. Other threats relate to resentful demoralization and mortality. In general, we can assume that those who perform better would be less likely to interrupt or stop answering the questionnaire. This is presumably not a problem when this dropout is equally relevant for both treatments. As we observe in the results, it appears to require a higher cognitive load to inspect the models with text labels. Participants receiving this treatment might be more likely to give up due to higher mental effort. While we did not have drop outs in the student replications, we noticed some instances in which online participants failed to answer all questions. For the online participants ($N = 42$), cases for the comprehension questions ranged from 0 missing answers to a maximum of 8 missing answers (out of 36 questions), with the mean being 1.69. We then performed a linear regression analysis to examine whether the number of missing answers has a significant effect on the number of correct answers. The regression model showed that number of missing answers was an insignificant predictor ($t = -1.64$, $p = 0.11$), thereby alleviating concerns about internal validity of our results.

Construct validity can be related to potential interactions between the measures. To that end, first, we inspected the measure correlations as reported above. We did not find any unexpected correlations, but only those that establish confidence in the convergent validity of our comprehension measures (task performance and task efficiency: r = -0.31, $p < 0.01$) and expertise measures (experience and intensity: r = 0.24, $p < 0.05$), and the discriminant validity of our model and personal factors (e.g., label type and knowledge: r = -0.01, $p > 0.05$).

As reported above, we also cared to eliminate potential bias stemming from non-equivalency between the treatment groups, by conducting manipulation checks to assess differences between the groups of participants across treatments. We noted above that there were no significant differences in the independent and dependent variables used, based on independent samples t-tests using the experimental medium used (paper versus online), student cohort (two from Vienna University of Economics and Business versus one from Humboldt-Universität zu Berlin), or time of experiment (2007, April 2009, June 2009). These results indicate that the participants were effectively randomized across treatments. We can also assume that there was no hypothesis guessing by the participants as we did not even reveal that two different treatments were used. The students participated as a preparation for the exam while the practitioners expected to receive feedback on their performance.

External validity is concerned with how generalizable the results are to the wider population of process modelers. Our set of replications was particularly motivated by external validity considerations, since we aim to generalize to the population of professionals involved in process modeling initiatives. Our manipulation checks confirmed that our replications can be considered strict, thereby increasing the external validity of our findings. One particular aspect of the external validity of the presented research relates to the extent to which the used models are representative for real-world models. As explained, we countered this threat by our choice of real process models from an partnering organization. A third important aspect that refers to a potentially limited external validity, relates to the involvement of students. We note that some of the students possessed prior practical experience with process modeling. Also, prior research found that students tend to have higher theoretical knowledge [47]. While we explicitly built both these factors into our research model, this could be seen as a limitation of this research, as the population in our study is potentially more knowledgeable of formal aspects of process modeling theory than the wider population. And indeed, our results confirm that theoretical knowledge is a key factor in explaining process model comprehension. One may argue, however, that process modeling students will form the next generation of junior analysts, and therefore our results may be predictive of the future generations of process analysts.

Last, we consider the effect of setting as a potential threat to external (as well as internal) validity: We used an online and a paper-based system. Therefore, participants either viewed process models on screen or as a printout. Both these practices are widespread in industry practice, where models are either provided through an intranet web page linked to a modeling tool (e.g., ARIS Web Publisher), or provided in print out format as part of process handbooks or manuals of procedures. Our study used both options, thereby increasing the external validity of the study. As noted above, we observed no statistical differences in relation to the *experimentMode*, thereby alleviating concerns about the internal validity of this treatment.

5 Implications

In this section, we discuss implications for Research (Section 5.1) and for practice (Section 5.2).

5.1 *Implications For Research*

The findings presented in this paper have three major implications for research. First, we have shown that textual labels hamper syntax comprehension of process models. This finding emphasizes the relevance of cognitive load theory for interpreting comprehension phenomena in this context. This is in line with prior research that identified size and complexity as factors having a negative impact on process model comprehension [28], although a direct reference to cognitive load theory is missing in these works. Cognitive load theory might offer a useful perspective to study the impact of process model complexity on comprehension in a more detailed way in future research. We further identify research on textual labels, e.g., [32] to be an important extension of our work, given that we identified textual labels to be a potential barrier to syntactical process model comprehension. Indeed, future work may examine how textual labels could be specified in order to decrease the additional cognitive burden on the model viewer.

Second, research on expert performance has established a close link between expertise and the duration and extent of training [26,13]. Our findings point to the fact that expertise is a task-specific phenomenon, as emphasized in [6]. Knowledge in theoretical aspects of process model syntax have been found as a significant factor of comprehension while general modeling intensity and general modeling experience were not significant. We speculated that semantic comprehension might be much more dependent on these factors than syntactical comprehension appeared to be. This speculation suggests that experience might have a different impact on comprehension of syntax, semantics, and pragmatics of a process model. These levels of comprehension might even be in conflict with each other. This aspect requires a deeper investigation in future research, both from a theoretical and from a behavioral perspective.

Third, our research showed that there is a trade-off in understanding the formal, syntactical structure of a model and its semantical content (as conveyed through textual labels). In this paper, therefore, we chose to examine process model understanding in terms of comprehension of syntactical content. Other research, by contrast, has examined semantic understanding, e.g., [42] whilst neglecting the syntactical comprehension. Future research should now combine these streams of study to be able to assert the relevant factors important to syntactic and semantic understanding, as well as the interactions between understanding of syntax and semantics. Ultimately, this vein of research can then arrive at a body of knowledge informing pragmatic understanding of process models as representations of knowledge for action [23], and study the factors the influence how individuals use process models to solve tasks such as organizational re-design, software specification, certification and others.

5.2 Implications For Practice

Our research has at least two relevant implications for practice. First, we note that the importance of theoretical knowledge for syntactical process model comprehension was supported by our tests. In contrast, practical experience does not seem to have a significant impact. These facts suggest that it is essential to provide formal process modeling education to staff members before letting them take part in a project. Such a training program should proceed in two stages. Initially, it should develop sufficient expertise in the syntactical rules of process modeling to ensure that practitioners appropriately understand the syntax of process models. Subsequently, the training program could proceed to more realistic process models that carry domain semantics, to teach practitioners how to reason about the processes being modeled. The recommendations in [43] could guide the development of a staged training program.

Second, we note that there are several situations in practice when syntactical aspects have to be investigated for a process model. This is, for instance, the case when a process model needs to be verified for soundness [1] before it is deployed in a workflow system. Our findings suggest that a tool option to hide, or to abbreviate the activity labels, could help analysts when correcting a syntactically unsound model. The abbreviation would reduce the cognitive load of the modeler, which would permit her to focus her attention on control flow. Corresponding features are not yet part of nowadays modeling tools.

6 Conclusions

Using process modeling for the analysis and design of process-aware information systems is an emerging, highly relevant domain of Information Systems practice. In this paper, we have described the formulation and execution of an experimental study to examine factors of process model comprehension.

We identify two key limitations to the work carried out. First, congruent to other studies, e.g. [7,32], we used post-graduate students as proxies for novice business analysts. Second, our operationalization of model comprehension was focused on the syntactical structure of a process model. Future work could investigate other aspects of understanding, for instance, through problem-solving tasks, e.g. [42]. In spite of the boundaries set by these limitations, we believe our work offers two central contributions. First, we provided a theoretical framework to define levels of process model comprehension task performance and efficiency, and the set of factors relevant to reaching comprehension on basis of cognitive load theory and semiotic considerations. Second, our series of experiments examined two sets of relevant factors - model factors and personal factors. We found that theoretical knowledge and, to a small extent, process modeling expertise, are important personal factors, and also found a negative effect of textual domain semantics - a model factor - on the comprehension of the formal content of process models.

Our work extends the body of knowledge in the field of process modeling, and thereby paves the way to more effective and efficient process modeling - which will significantly increase the benefits of process modeling in organizations [18], and also reduce associated direct and indirect costs. In moving forward, we discussed a number of speculations and possible directions for future research in our implications section. Most notably, it will be an important objective for future research to study the joint impact of various factors on different levels of comprehension, from syntactical to semantical to pragmatic.

References

[1] W. M. P. van der Aalst, Business Process Management, Vol. LNCS 1806, Springer Verlag, 2000, Ch. Workflow Verification: Finding Control-Flow Errors Using Petri-Net-Based Techniques, pp. 161–183.

[2] R. Agarwal, P. De, A. Sinha, Comprehending object and process models: An empirical study, IEEE Transactions on Software Engineering 25 (4) (1999) 541–556.

[3] M. I. Aguirre-Urreta, G. M. Marakas, Comparing conceptual modeling techniques: A critical review of the eer vs. oo empirical literature, The DATA BASE for Advances in Information Systems 39 (2) (2008) 9–32.

[4] W. Bandara, Factors and measures of business process modelling: model building through a multiple case study, European Journal of Information Systems 14 (2005) 347–360.

[5] V. Basili, F. Shull, F. Lanubile, Building knowledge through families of experiments, Software Engineering, IEEE Transactions on 25 (4) (1999) 456–473.

[6] S. Bonner, N. Pennington, Cognitive processes and knowledge as determinants of auditor expertise, Journal of Accounting Literature 10 (1) (1991) 1–50.

[7] A. Burton-Jones, P. Meso, The effects of decomposition quality and multiple forms of information on novices understanding of a domain from a conceptual model, Journal of the Association for Information Systems 9 (12) (2008) 784–802.

[8] A. Burton-Jones, Y. Wand, R. Weber, Guidelines for empirical evaluations of conceptual modeling grammars, Journal of the Association for Information Systems 10 (6) (2009) 495–532.

[9] G. Canfora, F. García, M. Piattini, F. Ruiz, C. Visaggio, A family of experiments to validate metrics for software process models., Journal of Systems and Software 77 (2) (2005) 113–129.

[10] T. D. Cook, D. T. Campbell, Quasi-Experimentation: Design and Analysis Issues, Houghton Mifflin, Boston, Massachusetts, 1979.

[11] I. Davies, P. Green, M. Rosemann, M. Indulska, S. Gallo, How do practitioners use conceptual modeling in practice?, Data & Knowledge Engineering 58 (3) (2006) 358–380.

[12] A. Dreiling, M. Rosemann, W. M. P. van der Aalst, W. Sadiq, From conceptual process models to running systems: A holistic approach for the configuration of enterprise system processes, Decision Support Systems 45 (2) (2008) 189–207.

[13] K. Ericsson, A. Lehmann, Expert and exceptional performance: Evidence of maximal adaptation to task constraints, Annual Review of Psychology 47 (1) (1996) 273–305.

[14] T. Erl, Service-oriented Architecture: Concepts, Technology, and Design, Prentice Hall, Upple Saddle Revier, New Jersey, 2005.

[15] C. Ferris, What are web services?, Communications of the ACM 46 (6) (2003) 31–32.

[16] J. Hahn, J. Kim, Why are some diagrams easier to work with? effects of diagrammatic representation on the cognitive integration process of systems analysis and design, ACM Transactions on Computer-Human Interaction 6 (3) (1999) 181–213.

[17] M. Hollander, D. A. Wolfe, Nonparametric Statistical Methods, 2nd Edition, John Wiley and Sons, New York, New York, 1999.

[18] M. Indulska, P. Green, J. Recker, M. Rosemann, Business process modeling: Perceived benefits, in: S. Castano, U. Dayal, A. H. F. Laender (Eds.),

Conceptual Modeling - ER 2009, Lecture Notes in Computer Science, Springer, Gramado, Brazil, 2009, pp. 458–471.

[19] A. Jedlitschka, M. Ciolkowski, D. Pfahl, Reporting experiments in software engineering, in: Guide to advanced empirical software engineering, Springer, 2008, pp. 201–228.

[20] R. Kanfer, P. L. Ackerman, T. C. Murtha, B. Dugdale, L. Nelson, Goal setting, conditions of practice, and task performance: A resource allocation perspective, Journal of Applied Psychology 79 (6) (1994) 826–835.

[21] B. Kiepuszewski, A. H. M. ter Hofstede, W. M. P. van der Aalst, Fundamentals of control flow in workflows, Acta Informatica 39 (3) (2003) 143–209.

[22] N. Kock, J. Verville, A. Danesh-Pajou, D. DeLuca, Communication flow orientation in business process modeling and its effect on redesign success: results from a field study, Decision Support Systems 46 (2) (2009) 562–575.

[23] J. Krogstie, G. Sindre, H. D. Jorgensen, Process models representing knowledge for action: a revised quality framework, European Journal of Information Systems 15 (1) (2006) 91–102.

[24] A. Lakhotia, Understanding someone elses code: Analysis of experiences, Journal of Systems and Software 23 (3) (1993) 269–275.

[25] O. Lindland, G. Sindre, A. Sølvberg, Understanding quality in conceptual modeling., IEEE Software 11 (2) (1994) 42–49.

[26] P. Lindsay, D. Norman, Human Information Processing: An introduction to psychology, second edition Edition, Academic Press, 1977.

[27] D. Liu, M. Shen, Business-to-business workflow interoperation based on process-views, Decision Support Systems 38 (3) (2004) 399–419.

[28] J. Mendling, Metrics for Process Models: Empirical Foundations of Verification, Error Prediction and Guidelines for Correctness, Vol. 6 of Lecture Notes in Business Information Processing, Springer, Berlin, Germany, 2008.

[29] J. Mendling, W. M. P. van der Aalst, Towards EPC Semantics based on State and Context, in: M. Nüttgens and F.J. Rump and J. Mendling (Ed.),

Proceedings of the 5th GI Workshop on Business Process Management with Event-Driven Process Chains (EPK 2006), German Informatics Society, Vienna, Austria, 2006, pp. 25–48.

[30] J. Mendling, H. A. Reijers, W. M. P. van der Aalst, Seven process modeling guidelines (7PMG), Information and Software Technology 52 (2) (2010) 127–136.

[31] J. Mendling, H. A. Reijers, J. Cardoso, What makes process models understandable?, in: G. Alonso, P. Dadam, M. Rosemann (Eds.), Business Process Management - BPM 2007, Vol. 4714 of Lecture Notes in Computer Science, Springer, Brisbane, Australia, 2007, pp. 48–63.

[32] J. Mendling, H. A. Reijers, J. Recker, Activity labeling in process modeling: Empirical insights and recommendations, Information Systems 35 (4) (2010) 467–482.

[33] J. Mendling, M. Strembeck, Influence factors of understanding business process models, in: W. Abramowicz, D. Fensel (Eds.), Business Information Systems - BIS 2008, Vol. 7 of Lecture Notes in Business Information Processing, Springer, Innsbruck, Austria, 2008, pp. 142–153.

[34] G. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information, The Psychological Review 63 (2) (1956) 343-355.

[35] T. Moher, D. Mak, B. Blumenthal, L. Leventhal, Comparing the Comprehensibility of Textual and Graphical Programs: The Case of Petri Nets, in: C. Cook, J. Scholtz, J. Spohrer (Eds.), Empirical Studies of Programmers: Fifth Workshop: Papers Presented at the Fifth Workshop on Empirical Studies of Programmers, December 3-5, 1993, Ablex Pub, 1993, pp. 137–161.

[36] M. J. Nanna, S. S. Sawilowsky, Analysis of likert scale data in disability and medical rehabilitation research, Psychological Methods 3 (1) (1998) 55–67.

[37] D. A. Norman, Cognitive engineering, in: D. A. Norman, S. W. Draper (Eds.), User Centered System Design: New Perspectives on Human-computer Interaction, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986, pp. 31–61.

[38] S. Patig, V. Casanova-Brito, B. Vgeli, It requirements of business process management in practice an empirical study, in: R. Hull, J. Mendling, S. Tai (Eds.), Business Process Management - BPM 2010, Vol. 6336 of Lecture Notes in Computer Science, Springer, Hoboken, New Jersey, 2010, pp. 13–28.

[39] G. Poels, A. Maes, F. Gailly, R. Paemeleire, The pragmatic quality of resources-events-agents diagrams: An experimental evaluation, Information Systems Journal 21 (1) (2011) 63–89.

[40] H. Purchase, Which Aesthetic has the Greatest Effect on Human Understanding?, in: Graph Drawing: 5th International Symposium, Gd'97, Rome, Italy, September 18-20, 1997: Proceedings, Springer, 1997, pp. 248–261.

[41] J. Recker, Continued use of process modeling grammars: The impact of individual difference factors, European Journal of Information Systems 19 (1) (2010) 76–92.

[42] J. Recker, A. Dreiling, The effects of content presentation format and user characteristics on novice developers understanding of process models, Communications of the Association for Information Systems 28 (6) (2011) 65–84.

[43] J. Recker, M. Rosemann, Teaching business process modeling experiences and recommendations, Communications of the Association for Information Systems 25 (32) (2009) 379–394.

[44] J. Recker, M. Rosemann, M. Indulska, P. Green, Business process modeling: A comparative analysis, Journal of the Association for Information Systems 10 (4) (2009) 333–363.

[45] H. A. Reijers, T. Freytag, J. Mendling, A. Eckleder, Syntax highlighting in business process models, Decision Support Systems 51 (3) (2011) 339–349.

[46] H. A. Reijers, J. Mendling, Modularity in process models: Review and effects, in: Proceedings of BPM 2008, Springer, 2008, pp. 20–35.

[47] H. A. Reijers, J. Mendling, A study into the factors that influence the understandability of business process models, IEEE Transactions on Systems Man and Cybernetics, Part A 41 (3) (2011) 449–462.

[48] E. Rolón Aguilar, F. García, F. Ruiz, M. Piattini, An exploratory experiment to validate measures for business process models, in: First International Conference on Research Challenges in Information Science (RCIS), 2007.

[49] K. Sarshar, P. Loos, Comparing the control-flow of epc and petri net from the end-user perspective., in: W. Aalst, B. Benatallah, F. Casati, F. Curbera (Eds.), Business Process Management, 3rd International Conference, BPM 2005, Nancy, France, September 5-8, 2005, Proceedings, LNCS 3649, 2005, pp. 434–439.

[50] A.-W. Scheer, ARIS - Business Process Modeling, 3rd Edition, Springer, Berlin, Germany, 2000.

[51] C. Soh, M. L. Markus, K. H. Goh, Electronic marketplaces and price transparency: Strategy, information technology, and success, MIS Quarterly 30 (3) (2006) 705–723.

[52] J. Sweller, P. Chandler, Why some material is difficult to learn, Cognition and Instruction 12 (3) (1994) 185–223.

[53] J. Sweller, J. Van Merrienboer, F. Paas, Cognitive architecture and instructional design, Educational psychology review 10 (3) (1998) 251–296.

[54] H. J. Wang, J. L. Zhao, L.-J. Zhang, Policy-driven process mapping (pdpm): Discovering process models from business policies, Decision Support Systems 48 (1) (2009) 267–281.

[55] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, A. Wesslen, Experimentation in software engineering: an introduction, Kluwer Academic Publishers, 2000.

Appendix: Experimental Material

A complete sample workbook of the questionnaire used in the printout experiment is available with abstract models (http://www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-abstrakt.pdf) and with textual models (http://www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-konkret.pdf).

Task 1: Process Modeling Intensity

- How often do you encounter process models in practice? (never, less than once a month, more than once a month, daily)

Task 2: Process Modeling Experience

- When did you first work with process models in practice? (less than a month ago, less than a year ago, less than three years ago, more than three years ago)

Task 3: Theoretical Knowledge

- After exclusive choices, at most one alternative path is executed (yes/no).
- Exclusive choices can be used to model repetition (yes/no).
- Synchronization is modeled in a Petri net by a place with two transitions in its preset (yes/no).
- Synchronization means that two activities are executed at the same time (yes/no).
- An inclusive OR can activate concurrent paths (yes/no).
- If two activities are concurrent, they have to be executed at the same time (yes/no).
- If an activity is modeled to be part of a loop, it has to be executed at least once (yes/no).
- Having an AND-split at the exit of a loop can lead to non-termination (yes/no).
- A deadlock is the result of an inappropriate combination of splits and joins

(yes/no).

- Processes without loops cannot deadlock (yes/no).
- Both an AND-join or an XOR-join can be used as a correct counterpart of an OR-split (yes/no).
- A multiple choice activates either one or all subsequent paths (yes/no).

Task 4: Comprehension Questions for Model 4 of Figure 1

(1) Is U always executed, when T has been executed? (yes/no)

(2) If F is executed, has Z or E been executed? (yes/no)

(3) Is it possible to execute U as well as I after F? (yes/no)

(4) Can this process be completed by executing less than five activities? (yes/no)

(5) When R is executed, is it possible that M has been executed before? (yes/no)

(6) Is it guaranteed that the process has neither deadlocks nor lack of synchronization? (yes/no)