

ePub^{WU} Institutional Repository

Jan Mendling and Mark Strembeck and Jan Recker

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

Article (Accepted for Publication)
(Refereed)

Original Citation:

Mendling, Jan and Strembeck, Mark and Recker, Jan
(2012)

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

Decision Support Systems, 53 (1).

pp. 195-206. ISSN 0167-9236

This version is available at: <https://epub.wu.ac.at/3499/>

Available in ePub^{WU}: April 2012

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the version accepted for publication and — in case of peer review — incorporates referee comments. There are minor differences between this and the publisher version which could however affect a citation.

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

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

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

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

25 More specifically, we study (a) factors characterizing the process model in
26 terms of the activity labels used in the models, (b) factors characterizing the
27 person interpreting the models in terms of relevant modeling expertise, and
28 (c) how these factors affect process model comprehension. The relevance of

29 this research stems from companies making significant investments in process
30 modeling training, with the view of developing a body of process modeling
31 expertise. Indeed, modeler expertise has been established by surveys as an
32 important factor for process modeling success [4] and modeling grammar usage
33 [41]. Furthermore, prior experiments demonstrate that model factors (e.g.,
34 an increase in model complexity) affect understanding [48,47]. Notably, these
35 experiments use abstract activity labels (A, B, C etc.) in their process models,
36 which, in turn, raises the question whether the usage of activity labels that
37 carry real domain semantics leverages or impedes understanding.

38 The aim of the research reported here is to combine these preliminary insights
39 in the definition of a series of experiments. Accordingly, the contributions of
40 this paper are threefold. First, we build on the cognitive load theory to con-
41 jecture that real activity labels should decrease syntactical process model under-
42 standing. This hypothesis is confirmed in our experiments. Second, we argue
43 in line with prior research that higher modeling expertise results in better un-
44 derstanding performance. This hypothesis is generally confirmed, too. Third,
45 we define different measures of expertise including theoretical knowledge, prior
46 modeling experience, and intensity of modeling. The experiments show that
47 theoretical knowledge is most significant with its impact on performance. Our
48 findings have implications for research on model understanding, in particu-
49 lar regarding cognitive load considerations, and for practice by demonstrating
50 the relevance of theoretical knowledge of process modeling to understanding
51 these models. This insight, in turn, is relevant to informing a staged teaching
52 strategy that educates practitioners about how to read process models.

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

59 2 Background

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

64 2.1 Process Model Comprehension

65 Process modeling has emerged as an important practice to guide decisions
66 in systems analysis and design. In fact, process modeling is the number one
67 reason to engage in conceptual modeling altogether [11], and also considered
68 the number one skill demanded from IT graduates¹. Analysts develop pro-
69 cess models to capture relevant information about a business process they seek
70 to re-design, analyze, or support with an appropriate information system. A
71 business process that is in place to deal with a book order may, for example,
72 contain a task to receive the order, which is followed by another one specifying
73 that the book is to be sent to the customer who ordered it. A model of this
74 process would, therefore, include sequences of graphical elements to describe
75 these tasks and the order in which they have to be performed. Process mod-
76 els can be elicited through interviews with relevant stakeholders, or derived
77 from organizational documents such as business policies [54]. Figures 1 and
78 2 show two variants of a typical process model, conveying information about
79 important tasks and the control flow that specifies the execution of these tasks.

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

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

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

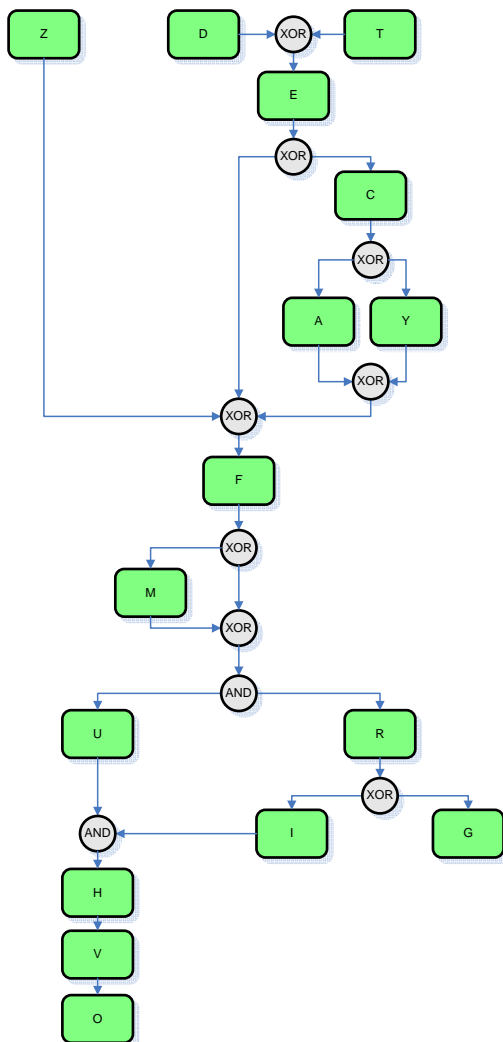


Figure 1. Model 4 with Letters

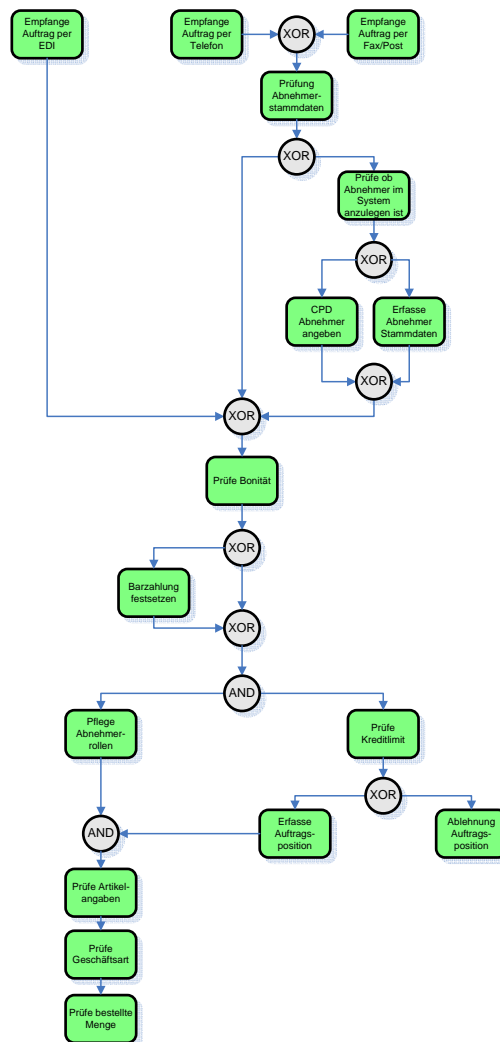


Figure 2. Model 4 with German Text

95 Semiotic theory postulates that comprehension, and consequently, communi-
96 cation, can be understood as a ladder: syntax (how do I faithfully combine
97 grammatical elements in a process model? [8]) must be clear before seman-
98 tics can be discussed, and semantics (what do the grammatical elements in a
99 process model mean? [8]) must be clear before pragmatics can be considered.
100 In this regard, it is a primary interest to analyze in how far stakeholders are
101 able to understand process models on a syntactical level. Other interpretations
102 are flawed if syntax is not correctly understood. This is also acknowledged by
103 prior studies that focus on formal and syntactical aspects of process models
104 [46,47].

105 Looking at which factors influence the comprehension of the syntactical con-
106 tent of process models, prior research has discussed several factors of pro-
107 cess model understanding including model purpose [47], problem domain [24],
108 modeling notation [49,16,2], visual presentation [35,40,45], and process model
109 complexity [9,28]. Personal factors, on the other hand, have been less inten-
110 sively researched to date. This is not to say that no research as been con-
111 ducted. The experiment by Recker and Dreiling, for instance, operationalized
112 the notion of process modeling expertise through a measure of familiarity with
113 a particular modeling notation [42]. In an experiment by Mendling, Reijers,
114 and Cardoso, participants were characterized based on the number of process
115 models they created and the years of modeling experience they had achieved
116 [31]. This study, furthermore, also indicated the specific importance of theo-
117 retical process modeling knowledge. In the latter experiment the participants
118 from TU Eindhoven with strong Petri net education scored better than other
119 participants with less theoretical education in process modeling.

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

124 Aside from these important personal factors, we also aim to examine model

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

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

154 In theorizing anticipated effects of the factors discussed above on process
155 model understanding, we first define our operationalization of process model
156 understanding. Similar to [39], we investigate syntactic understanding from
157 two angles, these being *comprehension task performance* (how faithfully does
158 the interpretation of the process model allow the reader to comprehend the for-
159 mal content of the model?) and *comprehension task efficiency* (what resources
160 are used by the reader to comprehend the process model?). Both factors are
161 important elements in Norman’s theory of action [37], and relate to what Nor-
162 man calls “the gulf of interpretation” (a difference between what the model
163 tries to convey and what is interpreted by the model reader). The gulf of in-
164 terpretation is an important measure of the performance of modeling efforts,
165 because model comprehension by relevant stakeholders is a necessary prereq-
166 uisite for various model application tasks, such as systems analysis, commu-
167 nication, design, organizational re-engineering, project management, end user
168 querying and others [44]. In other words, for a model to be useful for any
169 modeling-related task, it is imperative that the stakeholders doing these tasks
170 are able to comprehend the model well (performance) and timely (efficiency).

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

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

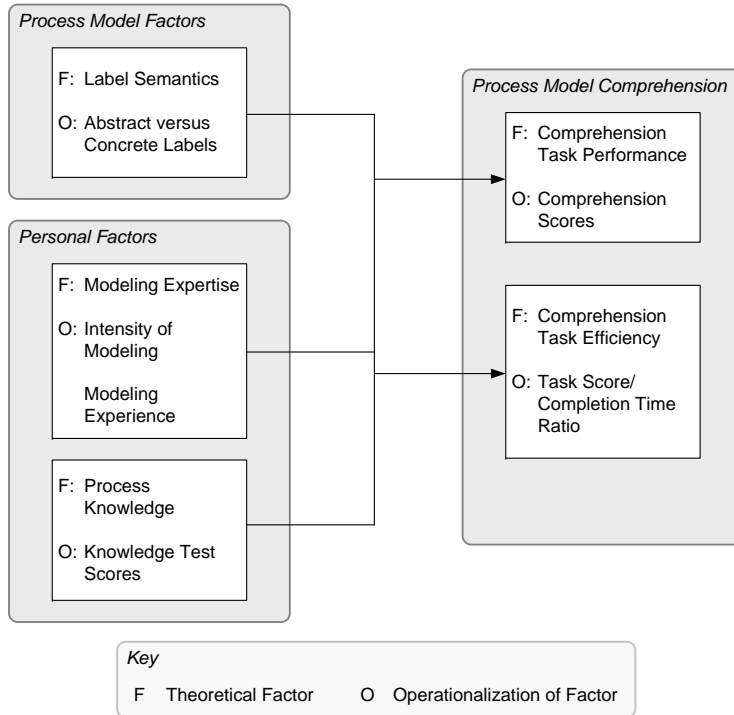


Figure 3. Research model

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

195 We further expect that comprehension occurs quicker for people working with
 196 process models featuring abstract textual labels, because they require less

197 effort to retrieve and assemble pieces of information in their working memory,
198 when only having to consider graphical constructs but not additional textual
199 information. We formalize these observations in the first two hypotheses:

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

202 H_a^1 The use of abstract labels will have a significant positive impact on com-
203 prehension task performance.

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

206 H_a^2 The use of abstract labels will have a significant positive impact on com-
207 prehension task efficiency.

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

217 H_0^3 Users with higher levels of process knowledge will not have higher com-
218 prehension task performance.

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

221 H_0^4 Users with higher levels of process knowledge will not have higher com-
222 prehension task efficiency.

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

225 Second, we realize that modeling expertise is an important factor in process

226 modeling [4,41]. Experienced modelers often possess a repertoire of workarounds
227 for challenging modeling situations, and can often refer to their previous expe-
228 riences and knowledge about modeling when attempting to interpret complex
229 models. Less experienced modelers, on the other hand, often lack such knowl-
230 edge, which, in turn, can be expected to affect their comprehension accuracy
231 and efficiency.

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

242 H_0^5 Users with higher levels of modeling experience will have equal compre-
243 hension task performance.

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

246 H_0^6 Users with higher levels of modeling experience will have equal compre-
247 hension task efficiency.

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

250 H_0^7 Users with higher levels of modeling intensity will have equal comprehen-
251 sion task performance.

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

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

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

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

260 **3 Experiment Description**

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

271 *3.1 Experiment Design*

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

277 *3.1.1 Experimental Condition and Tasks*

278 The between-subjects factor, *Label Type*, had two levels. We provided partici-
279 pants with process models that contained either abstract or concrete labels. To

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

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

300 3.1.2 *Independent Variables*

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

309 labels (*textual labels*), one for models with letters (*abstract labels*).

310 Aside from the between-subjects factor *Label Type*, we also defined three
311 within-subject factors. The first within-subjects factor *Knowledge* had four
312 levels. The participants had to answer twelve theoretical yes/no questions be-
313 fore seeing the models about selected topics related to process modeling such
314 as choices, concurrency, loops, and deadlocks (see Appendix). These questions
315 concern grammatical rules of process model logic, derived from fundamental
316 work in this area [21] and as previously used in [33]. We transformed the
317 *knowledge score* into an ordinal *knowledge* scale with four levels: very low (0-
318 3 correct answers), somewhat low (4-6 correct answers), somewhat high (7-9
319 correct answers) and very high (10-12 correct answers). This ordinal measure
320 served as a second independent variable. The second within-subjects factor *Ex-*
321 *perience* had four levels. The participants were asked for how long they have
322 been involved with business process modeling. The variable was measured on
323 an ordinal scale with four levels: less than one month, less than a year, less
324 than three years, and longer than three years. This measure served as a third
325 independent variable. Finally, the third within-subjects factor *Intensity* also
326 had four levels. The participants had to indicate how often they work with
327 process models. We used an ordinal scale with four options to answer: daily,
328 monthly, less frequent than monthly, never. This measure served as a fourth
329 independent variable.

330 3.1.3 *Dependent Variables*

331 We use two dependent variables, comprehension task performance and compre-
332 hension task efficiency. *Comprehension Task Performance* is calculated based
333 on the answers given by the participant to the model comprehension ques-
334 tions. It captures the number of correct answers by the person. The maximum
335 value is 36 for six questions on six models. This measure serves as an opera-
336 tionalization of formal process model understanding of a person.

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

342 3.2 *Experiment Execution*

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

350 In 2007, we distributed the link to the experiment via the German mailing lists
351 EMISA and WI as well as among students that followed courses on process
352 modeling at the Vienna University of Economics and Business. Typically, both
353 academics and practitioners with an interest in conceptual modeling and infor-
354 mation systems development are registered with these lists. The questionnaire
355 was started by 200 persons and completed by 46. From these 46 we excluded
356 4 people who spent less than 10 minutes time on the questionnaire since we
357 assumed that to be the minimum time to provide meaningful answers. The
358 remaining 42 persons and their answers to the 36 questions establish the first
359 part of the sample for our statistical analysis below. Altogether, 1512 answers
360 are recorded in the sample. 65% of the participants had more than three years
361 experience in process modeling.

362 To increase confidence in the conclusion validity of our study, we collected
363 further data with paper-based replications of the experiment. The first repli-
364 cation in April 2009 involved 23 graduate students from Vienna University

365 of Economics and Business who followed a course on modeling. The second
366 sample includes 22 graduate students who followed the same course in June
367 2009.² The third replication was conducted with 32 graduate students who
368 followed the system analysis and design course at Humboldt-Universität zu
369 Berlin. From all four experiments we collected data from altogether 119 per-
370 sons. With each answering 36 questions, we get 4284 answers to model under-
371 standing questions.

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

385 Each of the experiments used feedback about the performance as an induce-
386 ment. While this feedback was meant to be informative to practitioners, it
387 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 Task Performance	affiliation	Original study	42	26.26	4.94	0.17
		Replication 1	23	25.44	4.02	
		Replication 2	22	26.36	4.28	
	ExperimentMode	Replication 3	32	25.78	4.90	0.23
		Online	42	26.60	4.49	
		Paper	77	25.58	4.25	
Comprehension Task Efficiency	affiliation	Original study	42	1.31	0.66	0.27
		Replication 1	23	1.22	0.29	
		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

388 4 Data Analysis and Interpretation

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

391 4.1 Distribution and Correlation Analysis

392 Table 2 shows descriptive statistics for our measures. All results are in line
393 with expectations. Table 3 gives the correlation matrix. First, we check for po-
394 tential interactions between our between-subject factor (label type) and our
395 within-subject factors (experience, intensity, knowledge). The data in Table 3
396 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

397 factors, thereby suggesting independence of the experimental conditions used
398 in our study. The insignificant correlations of the between-subjects factor and
399 the within-subject factors allows to run the hypothesis tests independently.
400 Further inspection of Table 3 suggests that *Label type* and formal process
401 knowledge (*knowledge*) are meaningful independent factors as they correlate
402 significantly with the dependent measures. By contrast, *experience* and *inten-*
403 *sity* do not correlate largely with the dependent measures but with each other.
404 This correlation between intensity and experience, however, behaves in accor-
405 dance with general expectations (in the sense that people that model longer
406 often model more frequently, too). Next, the correlation between intensity and
407 experience to knowledge is expected, as people with more intensive or over-
408 all longer process modeling experiences build up higher levels of knowledge
409 about process modeling. The correlations between *comprehension score* and
410 *efficiency*, likewise, were expected. Overall, we do not find counter-intuitive
411 correlations in Table 3. Note that in Table 2 we see that the sample size for the
412 efficiency measure is 87, which is because we failed to accurately record task
413 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 indi-

427 vidually for each of the four independent factors above, using SPSS Version
428 16.0. First, we performed an Analysis of Variance (ANOVA) for our between-
429 subjects factor *Label Type*. Then, for each of the three factors formal process
430 knowledge, process modeling experience, and process modeling intensity, we
431 used a non-parametric Kruskal-Wallis test to examine our hypotheses, because
432 a Kolmogorov-Smirnov test confirmed that the normality assumption did not
433 hold for these measures, i.e. $Z = 2.51$ (*knowledge*), 2.68 (*experience*), 2.52
434 (*intensity*), all $p < 0.01$. Therefore, we used the Kruskal-Wallis test, which is
435 accepted as an alternative to ANOVA in case the considered variables are not
436 normally distributed [51]. We examined the hypotheses individually because
437 our correlation analysis suggested independence of the between-subjects and
438 within-subjects factors. Also, our experimental design features three ordinal
439 variables, for which we required non-parametric tests, and the Kruskal-Wallis
440 test we selected considers one independent variable at a time. We chose this
441 test over others (e.g., ANOVA, Mann-Whitney) because, first, the Kruskal-
442 Wallis test is the generalization of the Mann-Whitney test when there are
443 more than two independent groups, like in our study (four levels) [17]. Sec-
444 ond, even though we replicated the experiment to gather more data, the num-
445 ber of respondents overall is rather small, and the subgroups for each ordinal
446 scale level are smaller. The distribution-free nature of non-parametric tests
447 places few restrictions on the sample size in contrast with parametric tests,
448 which rely on asymptotic properties or normality of the sample distribution
449 [51]. Third, the ordinal measures used in our study called for the use of non-
450 parametric methods, which yield higher power than corresponding parametric
451 tests (e.g., ANOVA) [36]. Finally, rank-based non-parametric tests are not
452 affected by outliers [17], which allows us to also consider those data where
453 respondents took unusually long (or short) for answering the experimental
454 questions. Table 4 gives the descriptive results and Table 5 gives the results
455 from the statistical tests.

456 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

⁴⁵⁷ H_a^1 hypothesized higher comprehension task performance scores for the group
⁴⁵⁸ of users working with models with abstract labels. Table 4 shows that the av-
⁴⁵⁹ erage comprehension task performance scores indeed were higher (mean score
⁴⁶⁰ = 26.45 vs. 25.48), and Table 5 confirms that the differences are significant
⁴⁶¹ ($F = 5.05, p = 0.03$). These results lead to the rejection of null hypothesis
⁴⁶² H_0^1 and suggest people viewing models with no textual labels achieve a higher
⁴⁶³ 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

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

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

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

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

491 4.3 Testing Hypotheses on Comprehension Task Efficiency

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

- 496 • the use of abstract labels,
- 497 • higher levels of formal process knowledge,
- 498 • higher levels of process modeling experience, and
- 499 • higher levels of process modeling intensity.

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

510 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

511 H_a^2 hypothesized better comprehension task efficiency scores for the group
512 of users working with models with abstract labels. Table 6 shows that the
513 average comprehension task efficiency score, i.e., the ratio between correct
514 answers and time taken to complete the answers, indeed were lower for this
515 group (mean score = 1.39 vs. 1.03). Table 7 shows that the group differences
516 are significant ($F = 3.90, p = 0.05$). Therefore, the results suggest rejecting
517 null hypothesis H_0^2 , which means that textual semantics, being a significant
518 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

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

520 H_a^4 hypothesized better comprehension task efficiency scores for the group
521 of users working with higher levels of formal process knowledge. We note
522 from Table 7 that the differences in comprehension task efficiency across the
523 groups of users with different levels of knowledge are significant ($Chi - 2 =$
524 $8.38, p = 0.04$), and from Table 6 that the efficiency scores are better for
525 users with higher levels of knowledge. We note, however, that Table 6 also
526 shows a somewhat unexpected exception. The group of users with low levels
527 of knowledge completed their tasks the with the second-best efficiency score
528 (mean = 1.34), superseded only by those with high levels of knowledge (mean
529 = 1.51). We note that these results may have been over-compensated through
530 quick task completion, independent from correct results (as shown in Table 4).
531 Indeed, it seems plausible that users with low knowledge levels just quickly
532 selected answers without engaging in a thorough consideration of the content
533 presented to them. Overall, the results are in line with our expectations, the
534 null hypothesis H_0^4 is rejected.

535 H_a^6 and H_a^8 hypothesized better comprehension task efficiency scores for users
536 with higher levels of modeling expertise (in the sense of modeling experience
537 and intensity). We note from Table 7 that the differences in task completion
538 efficiency across the user groups with different levels of modeling intensity are
539 significant ($Chi - 2 = 9.09, p = 0.03$), and provide the correct directionality

540 (means = 1.09, 1.19, 1.28 and 1.30). The results support hypothesis H8a. For
541 modeling experience, however, the results are not in line with hypothesis H6a.
542 There are fluctuations in comprehension task efficiency scores noted in Table 6
543 (means = 1.36, 1.29, 1.01 and 1.21), and the Kruskal-Wallis tests suggests that
544 the differences across the groups are insignificant ($Chi - 2 = 4.29, p = 0.23$).
545 Therefore, we cannot reject null hypothesis H_0^6 .

546 4.4 Discussion of Results

547 Our experimental study provides support for five out of eight hypothesized
548 factors of process model comprehension task performance and efficiency (see
549 Table 8). The results for hypotheses H_a^1 and H_a^2 suggest that a plus in seman-
550 tical information in terms of text labels seems to be a burden when analyzing
551 the syntactical content of a process. These findings are in line with argu-
552 ments that are founded on the grounds of cognitive load theory as well as the
553 premise of the semiotic ladder. Hypotheses H_a^3 to H_a^8 are interesting to be dis-
554 cussed relative to each other. Theoretical knowledge turned out to be a strong
555 indicator for both comprehension task performance and efficiency on syntax-
556 related comprehension of process models (H_a^3 and H_a^4). In contrast, modeling
557 experience and intensity were found not to contribute significantly to either
558 comprehension task performance or efficiency, set aside the result obtained in
559 relation to hypothesis H_a^8 . We interpret this result as an indication that the-
560 oretical knowledge is of paramount importance to understanding syntactical
561 aspects of a process model, over and above any practical experience with the
562 exercise of process modeling. Indeed, the non-significance of experience and
563 intensity here might suggest that these factors are more important for the se-
564 mantical interpretation of process models and that theory is the prerequisite
565 for understanding syntax.

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

Table 8

Summary of Hypotheses Tests

566 *4.5 Threats to Validity*

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

570 *Conclusion validity* is concerned with the relationship between treatment and
571 outcome, and the conclusions drawn from it. Two aspects have to be consid-
572 ered: The first aspect concerns the appropriateness of the statistical tests. As
573 reported above, we have screened our data for conformance with the assump-
574 tions of the statistical tests we used (ANOVA, Kruskal-Wallis test). We used
575 Levene’s test to show that the dependent variables across the treatment groups
576 shared approximately equal variance. We used the non-parametric Kruskal-
577 Wallis test for our ordinal measures because the independent data was not
578 normally distributed. A Kolmogorov-Smirnov test confirmed that the normal-
579 ity assumption did not hold for the measures *knowledge*, *experience*, or *inten-*
580 *sity* ($Z = 2.51, 2.68, 2.52$, all $p = 0.00$). Therefore, we used the Kruskal-Wallis

581 test, which is accepted as an alternative to ANOVA in case the considered vari-
582 ables are not normally distributed [51]. The second aspect concerns the effect
583 sizes of the results. In order to reach a sample size sufficient to solve potential
584 issues regarding the statistical significance, we conducted strict replications
585 [5] of our experiment. In order to show that our replications did not induce
586 bias into our analysis, we created two dummy variables, *affiliation* and *ex-*
587 *perimentMode*, to examine whether experimental results differed significantly
588 across the replications. Affiliation with one of the universities partaking in
589 our study did not affect results for comprehension task performance or task
590 completion time - the Kruskal-Wallis test was insignificant ($p = 0.16$ and p
591 $= 0.09$). The mode of experiment (paper versus online), likewise, was an in-
592 significant factor, as shown in an independent samples t-test ($p = 0.20$ and p
593 $= 0.80$ for comprehension task performance and task completion time).

594 *Internal validity* demands that the treatment causes the effect. In order to
595 avoid maturation and learning effects, we used a random sampling of the
596 questions. Other threats relate to resentful demoralization and mortality. In
597 general, we can assume that those who perform better would be less likely
598 to interrupt or stop answering the questionnaire. This is presumably not a
599 problem when this dropout is equally relevant for both treatments. As we
600 observe in the results, it appears to require a higher cognitive load to inspect
601 the models with text labels. Participants receiving this treatment might be
602 more likely to give up due to higher mental effort. While we did not have drop
603 outs in the student replications, we noticed some instances in which online
604 participants failed to answer all questions. For the online participants ($N =$
605 42), cases for the comprehension questions ranged from 0 missing answers to
606 a maximum of 8 missing answers (out of 36 questions), with the mean being
607 1.69. We then performed a linear regression analysis to examine whether the
608 number of missing answers has a significant effect on the number of correct
609 answers. The regression model showed that number of missing answers was
610 an insignificant predictor ($t = -1.64$, $p = 0.11$), thereby alleviating concerns
611 about internal validity of our results.

612 *Construct validity* can be related to potential interactions between the mea-
613 sures. To that end, first, we inspected the measure correlations as reported
614 above. We did not find any unexpected correlations, but only those that es-
615 tablish confidence in the convergent validity of our comprehension measures
616 (task performance and task efficiency: $r = -0.31$, $p < 0.01$) and expertise mea-
617 sures (experience and intensity: $r = 0.24$, $p < 0.05$), and the discriminant
618 validity of our model and personal factors (e.g., label type and knowledge: r
619 $= -0.01$, $p > 0.05$).

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

633 *External validity* is concerned with how generalizable the results are to the
634 wider population of process modelers. Our set of replications was particularly
635 motivated by external validity considerations, since we aim to generalize to
636 the population of professionals involved in process modeling initiatives. Our
637 manipulation checks confirmed that our replications can be considered strict,
638 thereby increasing the external validity of our findings. One particular aspect
639 of the external validity of the presented research relates to the extent to which
640 the used models are representative for real-world models. As explained, we
641 countered this threat by our choice of real process models from an partnering

642 organization. A third important aspect that refers to a potentially limited
643 external validity, relates to the involvement of students. We note that some of
644 the students possessed prior practical experience with process modeling. Also,
645 prior research found that students tend to have higher theoretical knowledge
646 [47]. While we explicitly built both these factors into our research model, this
647 could be seen as a limitation of this research, as the population in our study is
648 potentially more knowledgeable of formal aspects of process modeling theory
649 than the wider population. And indeed, our results confirm that theoretical
650 knowledge is a key factor in explaining process model comprehension. One may
651 argue, however, that process modeling students will form the next generation
652 of junior analysts, and therefore our results may be predictive of the future
653 generations of process analysts.

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

664 **5 Implications**

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

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

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

694 Third, our research showed that there is a trade-off in understanding the for-
695 mal, syntactical structure of a model and its semantical content (as conveyed

696 through textual labels). In this paper, therefore, we chose to examine process
697 model understanding in terms of comprehension of syntactical content. Other
698 research, by contrast, has examined semantic understanding, e.g., [42] whilst
699 neglecting the syntactical comprehension. Future research should now com-
700 bine these streams of study to be able to assert the relevant factors important
701 to syntactic and semantic understanding, as well as the interactions between
702 understanding of syntax and semantics. Ultimately, this vein of research can
703 then arrive at a body of knowledge informing pragmatic understanding of
704 process models as representations of knowledge for action [23], and study the
705 factors the influence how individuals use process models to solve tasks such
706 as organizational re-design, software specification, certification and others.

707 5.2 *Implications For Practice*

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

720 Second, we note that there are several situations in practice when syntactical
721 aspects have to be investigated for a process model. This is, for instance, the
722 case when a process model needs to be verified for soundness [1] before it is
723 deployed in a workflow system. Our findings suggest that a tool option to

724 hide, or to abbreviate the activity labels, could help analysts when correcting
725 a syntactically unsound model. The abbreviation would reduce the cognitive
726 load of the modeler, which would permit her to focus her attention on control
727 flow. Corresponding features are not yet part of nowadays modeling tools.

728 **6 Conclusions**

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

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

748 Our work extends the body of knowledge in the field of process modeling, and
749 thereby paves the way to more effective and efficient process modeling - which
750 will significantly increase the benefits of process modeling in organizations [18],

751 and also reduce associated direct and indirect costs. In moving forward, we
752 discussed a number of speculations and possible directions for future research
753 in our implications section. Most notably, it will be an important objective for
754 future research to study the joint impact of various factors on different levels
755 of comprehension, from syntactical to semantical to pragmatic.

756 **References**

- 757 [1] W. M. P. van der Aalst, Business Process Management, Vol. LNCS 1806,
758 Springer Verlag, 2000, Ch. Workflow Verification: Finding Control-Flow Errors
759 Using Petri-Net-Based Techniques, pp. 161–183.
- 760 [2] R. Agarwal, P. De, A. Sinha, Comprehending object and process models: An
761 empirical study, IEEE Transactions on Software Engineering 25 (4) (1999) 541–
762 556.
- 763 [3] M. I. Aguirre-Urreta, G. M. Marakas, Comparing conceptual modeling
764 techniques: A critical review of the eer vs. oo empirical literature, The DATA
765 BASE for Advances in Information Systems 39 (2) (2008) 9–32.
- 766 [4] W. Bandara, Factors and measures of business process modelling: model
767 building through a multiple case study, European Journal of Information
768 Systems 14 (2005) 347–360.
- 769 [5] V. Basili, F. Shull, F. Lanubile, Building knowledge through families of
770 experiments, Software Engineering, IEEE Transactions on 25 (4) (1999) 456–
771 473.
- 772 [6] S. Bonner, N. Pennington, Cognitive processes and knowledge as determinants
773 of auditor expertise, Journal of Accounting Literature 10 (1) (1991) 1–50.
- 774 [7] A. Burton-Jones, P. Meso, The effects of decomposition quality and multiple
775 forms of information on novices understanding of a domain from a conceptual
776 model, Journal of the Association for Information Systems 9 (12) (2008) 784–
777 802.

- 778 [8] A. Burton-Jones, Y. Wand, R. Weber, Guidelines for empirical evaluations
779 of conceptual modeling grammars, *Journal of the Association for Information*
780 *Systems* 10 (6) (2009) 495–532.
- 781 [9] G. Canfora, F. García, M. Piattini, F. Ruiz, C. Visaggio, A family of
782 experiments to validate metrics for software process models., *Journal of Systems*
783 *and Software* 77 (2) (2005) 113–129.
- 784 [10] T. D. Cook, D. T. Campbell, *Quasi-Experimentation: Design and Analysis*
785 *Issues*, Houghton Mifflin, Boston, Massachusetts, 1979.
- 786 [11] I. Davies, P. Green, M. Rosemann, M. Indulska, S. Gallo, How do practitioners
787 use conceptual modeling in practice?, *Data & Knowledge Engineering* 58 (3)
788 (2006) 358–380.
- 789 [12] A. Dreiling, M. Rosemann, W. M. P. van der Aalst, W. Sadiq, From conceptual
790 process models to running systems: A holistic approach for the configuration of
791 enterprise system processes, *Decision Support Systems* 45 (2) (2008) 189–207.
- 792 [13] K. Ericsson, A. Lehmann, Expert and exceptional performance: Evidence of
793 maximal adaptation to task constraints, *Annual Review of Psychology* 47 (1)
794 (1996) 273–305.
- 795 [14] T. Erl, *Service-oriented Architecture: Concepts, Technology, and Design*,
796 Prentice Hall, Upple Saddle Revier, New Jersey, 2005.
- 797 [15] C. Ferris, What are web services?, *Communications of the ACM* 46 (6) (2003)
798 31–32.
- 799 [16] J. Hahn, J. Kim, Why are some diagrams easier to work with? effects of
800 diagrammatic representation on the cognitive integration process of systems
801 analysis and design, *ACM Transactions on Computer-Human Interaction* 6 (3)
802 (1999) 181–213.
- 803 [17] M. Hollander, D. A. Wolfe, *Nonparametric Statistical Methods*, 2nd Edition,
804 John Wiley and Sons, New York, New York, 1999.
- 805 [18] M. Indulska, P. Green, J. Recker, M. Rosemann, Business process modeling:
806 Perceived benefits, in: S. Castano, U. Dayal, A. H. F. Laender (Eds.),

- 807 Conceptual Modeling - ER 2009, Lecture Notes in Computer Science, Springer,
808 Gramado, Brazil, 2009, pp. 458–471.
- 809 [19] A. Jedlitschka, M. Ciolkowski, D. Pfahl, Reporting experiments in software
810 engineering, in: Guide to advanced empirical software engineering, Springer,
811 2008, pp. 201–228.
- 812 [20] R. Kanfer, P. L. Ackerman, T. C. Murtha, B. Dugdale, L. Nelson, Goal setting,
813 conditions of practice, and task performance: A resource allocation perspective,
814 *Journal of Applied Psychology* 79 (6) (1994) 826–835.
- 815 [21] B. Kiepuszewski, A. H. M. ter Hofstede, W. M. P. van der Aalst, Fundamentals
816 of control flow in workflows, *Acta Informatica* 39 (3) (2003) 143–209.
- 817 [22] N. Kock, J. Verville, A. Danesh-Pajou, D. DeLuca, Communication flow
818 orientation in business process modeling and its effect on redesign success:
819 results from a field study, *Decision Support Systems* 46 (2) (2009) 562–575.
- 820 [23] J. Krogstie, G. Sindre, H. D. Jorgensen, Process models representing knowledge
821 for action: a revised quality framework, *European Journal of Information*
822 *Systems* 15 (1) (2006) 91–102.
- 823 [24] A. Lakhotia, Understanding someone else's code: Analysis of experiences,
824 *Journal of Systems and Software* 23 (3) (1993) 269–275.
- 825 [25] O. Lindland, G. Sindre, A. Sølvsberg, Understanding quality in conceptual
826 modeling., *IEEE Software* 11 (2) (1994) 42–49.
- 827 [26] P. Lindsay, D. Norman, *Human Information Processing: An introduction to*
828 *psychology*, second edition Edition, Academic Press, 1977.
- 829 [27] D. Liu, M. Shen, Business-to-business workflow interoperation based on process-
830 views, *Decision Support Systems* 38 (3) (2004) 399–419.
- 831 [28] J. Mendling, *Metrics for Process Models: Empirical Foundations of Verification,*
832 *Error Prediction and Guidelines for Correctness*, Vol. 6 of *Lecture Notes in*
833 *Business Information Processing*, Springer, Berlin, Germany, 2008.
- 834 [29] J. Mendling, W. M. P. van der Aalst, Towards EPC Semantics based on
835 State and Context, in: M. Nüttgens and F.J. Rump and J. Mendling (Ed.),

- 836 Proceedings of the 5th GI Workshop on Business Process Management with
837 Event-Driven Process Chains (EPK 2006), German Informatics Society, Vienna,
838 Austria, 2006, pp. 25–48.
- 839 [30] J. Mendling, H. A. Reijers, W. M. P. van der Aalst, Seven process modeling
840 guidelines (7PMG), *Information and Software Technology* 52 (2) (2010) 127–
841 136.
- 842 [31] J. Mendling, H. A. Reijers, J. Cardoso, What makes process models
843 understandable?, in: G. Alonso, P. Dadam, M. Rosemann (Eds.), *Business
844 Process Management - BPM 2007*, Vol. 4714 of *Lecture Notes in Computer
845 Science*, Springer, Brisbane, Australia, 2007, pp. 48–63.
- 846 [32] J. Mendling, H. A. Reijers, J. Recker, Activity labeling in process modeling:
847 Empirical insights and recommendations, *Information Systems* 35 (4) (2010)
848 467–482.
- 849 [33] J. Mendling, M. Strembeck, Influence factors of understanding business process
850 models, in: W. Abramowicz, D. Fensel (Eds.), *Business Information Systems -
851 BIS 2008*, Vol. 7 of *Lecture Notes in Business Information Processing*, Springer,
852 Innsbruck, Austria, 2008, pp. 142–153.
- 853 [34] G. Miller, The magical number seven, plus or minus two: Some limits on our
854 capacity for processing information, *The Psychological Review* 63 (2) (1956)
855 343-355.
- 856 [35] T. Moher, D. Mak, B. Blumenthal, L. Leventhal, Comparing the
857 Comprehensibility of Textual and Graphical Programs: The Case of Petri Nets,
858 in: C. Cook, J. Scholtz, J. Spohrer (Eds.), *Empirical Studies of Programmers:
859 Fifth Workshop: Papers Presented at the Fifth Workshop on Empirical Studies
860 of Programmers*, December 3-5, 1993, Ablex Pub, 1993, pp. 137–161.
- 861 [36] M. J. Nanna, S. S. Sawilowsky, Analysis of likert scale data in disability and
862 medical rehabilitation research, *Psychological Methods* 3 (1) (1998) 55–67.
- 863 [37] D. A. Norman, Cognitive engineering, in: D. A. Norman, S. W. Draper
864 (Eds.), *User Centered System Design: New Perspectives on Human-computer
865 Interaction*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986, pp. 31–
866 61.

- 867 [38] S. Patig, V. Casanova-Brito, B. Vgeli, It requirements of business process
868 management in practice an empirical study, in: R. Hull, J. Mendling, S. Tai
869 (Eds.), *Business Process Management - BPM 2010*, Vol. 6336 of *Lecture Notes*
870 *in Computer Science*, Springer, Hoboken, New Jersey, 2010, pp. 13–28.
- 871 [39] G. Poels, A. Maes, F. Gailly, R. Paemeleire, The pragmatic quality of resources-
872 events-agents diagrams: An experimental evaluation, *Information Systems*
873 *Journal* 21 (1) (2011) 63–89.
- 874 [40] H. Purchase, Which Aesthetic has the Greatest Effect on Human
875 Understanding?, in: *Graph Drawing: 5th International Symposium, Gd'97*,
876 Rome, Italy, September 18-20, 1997: *Proceedings*, Springer, 1997, pp. 248–261.
- 877 [41] J. Recker, Continued use of process modeling grammars: The impact of
878 individual difference factors, *European Journal of Information Systems* 19 (1)
879 (2010) 76–92.
- 880 [42] J. Recker, A. Dreiling, The effects of content presentation format and
881 user characteristics on novice developers understanding of process models,
882 *Communications of the Association for Information Systems* 28 (6) (2011) 65–
883 84.
- 884 [43] J. Recker, M. Rosemann, Teaching business process modeling experiences and
885 recommendations, *Communications of the Association for Information Systems*
886 25 (32) (2009) 379–394.
- 887 [44] J. Recker, M. Rosemann, M. Indulska, P. Green, Business process modeling: A
888 comparative analysis, *Journal of the Association for Information Systems* 10 (4)
889 (2009) 333–363.
- 890 [45] H. A. Reijers, T. Freytag, J. Mendling, A. Eckleder, Syntax highlighting in
891 business process models, *Decision Support Systems* 51 (3) (2011) 339–349.
- 892 [46] H. A. Reijers, J. Mendling, Modularity in process models: Review and effects,
893 in: *Proceedings of BPM 2008*, Springer, 2008, pp. 20–35.
- 894 [47] H. A. Reijers, J. Mendling, A study into the factors that influence the
895 understandability of business process models, *IEEE Transactions on Systems*
896 *Man and Cybernetics, Part A* 41 (3) (2011) 449–462.

- 897 [48] E. Rolón Aguilar, F. García, F. Ruiz, M. Piattini, An exploratory experiment to
898 validate measures for business process models, in: First International Conference
899 on Research Challenges in Information Science (RCIS), 2007.
- 900 [49] K. Sarshar, P. Loos, Comparing the control-flow of epc and petri net from
901 the end-user perspective., in: W. Aalst, B. Benatallah, F. Casati, F. Curbera
902 (Eds.), Business Process Management, 3rd International Conference, BPM
903 2005, Nancy, France, September 5-8, 2005, Proceedings, LNCS 3649, 2005, pp.
904 434–439.
- 905 [50] A.-W. Scheer, ARIS - Business Process Modeling, 3rd Edition, Springer, Berlin,
906 Germany, 2000.
- 907 [51] C. Soh, M. L. Markus, K. H. Goh, Electronic marketplaces and price
908 transparency: Strategy, information technology, and success, MIS Quarterly
909 30 (3) (2006) 705–723.
- 910 [52] J. Sweller, P. Chandler, Why some material is difficult to learn, Cognition and
911 Instruction 12 (3) (1994) 185–223.
- 912 [53] J. Sweller, J. Van Merriënboer, F. Paas, Cognitive architecture and instructional
913 design, Educational psychology review 10 (3) (1998) 251–296.
- 914 [54] H. J. Wang, J. L. Zhao, L.-J. Zhang, Policy-driven process mapping (pdpm):
915 Discovering process models from business policies, Decision Support Systems
916 48 (1) (2009) 267–281.
- 917 [55] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, A. Wesslen,
918 Experimentation in software engineering: an introduction, Kluwer Academic
919 Publishers, 2000.

920 Appendix: Experimental Material

921 A complete sample workbook of the questionnaire used in the printout ex-
922 periment is available with abstract models ([http://www.mendling.com/2009-](http://www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-abstrakt.pdf)
923 [Fragebogen-Rahmen-ABCDEF-abstrakt.pdf](http://www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-abstrakt.pdf)) and with textual models ([http://](http://www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-konkret.pdf)
924 www.mendling.com/2009-Fragebogen-Rahmen-ABCDEF-konkret.pdf).

925 **Task 1: Process Modeling Intensity**

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

928 **Task 2: Process Modeling Experience**

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

932 **Task 3: Theoretical Knowledge**

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

947 (yes/no).

948 • Processes without loops cannot deadlock (yes/no).

949 • Both an AND-join or an XOR-join can be used as a correct counterpart of
950 an OR-split (yes/no).

951 • A multiple choice activates either one or all subsequent paths (yes/no).

952 **Task 4: Comprehension Questions for Model 4 of Figure 1**

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

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

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

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

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

960 (6) Is it guaranteed that the process has neither deadlocks nor lack of syn-
961 chronization? (yes/no)