CrossMark

An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models

Jan Mendling[1] · Jan Recker[2] · Hajo A. Reijers[3] · Henrik Leopold[3]

Abstract

Understanding conceptual models of business domains is a key skill for practitioners tasked with systems analysis and design. Research in this field predominantly uses experiments with specific user proxy cohorts to examine factors that explain how well different types of conceptual models can be comprehended by model viewers. However, the results from these studies are difficult to compare. One key difficulty rests in the unsystematic and fluctuating consideration of model viewer characteristics (MVCs) to date. In this paper, we review MVCs used in prominent prior studies on conceptual model comprehension. We then design an empirical review of the influence of MVCS through a global, cross-sectional experimental study in which over 500 student and practitioner users were asked to answer comprehension questions about a prominent type of conceptual model - BPMN process models. As an experimental treatment, we used good versus bad layout in order to increase the variance of performance. Our results show MVC to be a multi-dimensional construct. Moreover, process model comprehension is related in different ways to different traits of the MVC construct. Based on these findings, we offer guidance for experimental designs in this area of research and provide implications for the study of MVCs.

Keywords Process modeling · Model viewer characteristics · Experiment · Conceptual modeling · Literature review · BPMN

1 Introduction

The complexity of contemporary information systems draws much attention to how their analysis and design can be supported by appropriate methods and tools. Efforts are spent on new techniques that support the modeling of system requirements and, increasingly, on how these techniques actually aid the analysis and design

process (Xiao and Zheng 2012). Of special interest in this stream are studies that focus on conceptual models as an aid to facilitate the comprehension of certain domain facts that relate to an information system, which will contribute to better design decisions and eventually a better system. Therefore, investigating the factors that influence the way people make sense of conceptual models is instrumental in improving the analysis and design of information systems in terms of their effectiveness and efficiency. Not surprisingly, conceptual modeling remains an active field of study, with contributions regularly occurring in the field's main journals.

Studies that discuss the comprehension of various modeling artifacts acknowledge *model viewer characteristics (MVCs)* as a factor of influence. Various aspects of MVCs have been discussed in the literature, partially relating to theoretical knowledge (Khatri et al. 2006; Mendling et al. 2012; Reijers and Mendling 2011), duration of practice (Recker 2010a; Reijers et al. 2011b; Recker and Dreiling 2011), education (Recker 2010a), or familiarity (Burton-Jones and Meso 2008).

We observe, however, that MVCs are hardly considered prominently in research on the comprehension of conceptual models. First, several experiments in this area cover them as control variables, but not as independent variables

✉ Jan Mendling
jan.mendling@wu.ac.at

Jan Recker
jan.recker@wiso.uni-koeln.de

Hajo A. Reijers
h.a.reijers@vu.nl

Henrik Leopold
h.leopold@vu.nl

[1] Wirtschaftsuniversität Wien, Welthandelsplatz 1, 1020 Vienna, Austria

[2] University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany

[3] Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

in their own right. Second, experiments use different operationalizations of different aspects of viewer characteristics such as the years of modeling experience (Burton-Jones and Meso 2008) or the number of models created (Recker and Dreiling 2011), which makes the results difficult to compare. Third, experiments often involve specific cohorts of potentially limited variation in MVCs such as students, although it is not fully understood in how far they have model viewer characteristics that are similar or different to other cohorts.

These observations call for research into the role of MVCs in model comprehension and into respective profiles of different model readers, such as students and practitioners. We are not the first to make this observation. Burton-Jones et al. (2009), for example, stated in their review already that "given the importance of these concepts, more work needs to be done" (p. 514). We take this step and examine in more detail how users assign meaning to the elements represented in a conceptual model presented to them. This is important because up until now, it is unclear to which extent results reported for certain cohorts can be extrapolated to other populations, in particular from students to IS professionals or business experts. Do students possess MVCs that professionals exploit in working with conceptual models or do students lack the decisive type of MVCs altogether? Furthermore, there is a lack of understanding on how different operationalizations of MVCs correlate with one another. For instance, Burton-Jones and Meso (2008) find a negative correlation of self-reported UML modeling experience with comprehension and problem-solving tasks among post-graduate students, while Mendling et al. (2012) find process modeling knowledge to have a significant and positive impact on model comprehension by students. Do these findings relate to different MVCs or to different profiles of the involved participants?

We chose to study, first, how different MVCs as used in prior experiments relate to one another; second, how important populations, namely students and practitioners, differ in these MVCs; and third, and how these differences have an impact on model comprehension performance. To this end, we designed and conducted an experimental study that compares and contrasts MVCs discussed in the literature in terms of their impact on model comprehension tasks. We use a popular type of conceptual model - BPMN process models, and utilize an experimental treatment of good versus bad layout in order to increase the variance of performance and in this way to better study the connections between MVCs and performance aspects.

The findings advance the literature in two directions. First, we systematically describe connections between independent measures, MVCs and their impact on comprehension. It is a unique feature of our study that these connections are grounded in empirical data, and in this way pave the way for theory building in future studies (Miller 2007). Second, we derive recommendations for covering MVCs in future experiments on model comprehension. These contributions are important also for system analysis and design as they extend our knowledge about core subject matters of the field (Sidorova et al. 2008) and the student-practitioner dichotomy (Compeau et al. 2012). They also complement recent research on the process of creating models (Claes et al. 2015) and previous surveys of conceptual modeling research (Houy et al. 2012; Figl 2017; Cognini et al. 2016). In this way, we aim to contribute to improving the external validity of model comprehension experiments toward the population of practitioners (Venable 2007; Kock et al. 2002).

We proceed as follows. Section 2 describes the background of our study. We recap conceptual modeling research and then review in some detail how the notion of MVCs has been operationalized, used and tested in prominent experimental research on model comprehension. Section 3 then presents the design of our study, and Section 4 provides the results. Section 5 discusses implications of this research. Section 6 concludes the paper with a review of its contributions.

2 Background

2.1 Conceptual Modeling

Conceptual modeling is a key task during system analysis and design, where professionals attempt to develop a representation of elements of a real-world domain that they believe to be important to consider when analyzing or designing information systems (Wand and Weber 2002). Thereby, conceptual models are used to facilitate a communicative process among relevant stakeholders, they document relevant process and data requirements pertinent to the implementation of a system, and they guide end users in operating and maintaining the system. For all these purposes, it is of importance that professionals are able to understand the content of these models to be able to reason about them. This makes model comprehension an important and active stream of research (Burton-Jones et al. 2009). Conceptual models are developed by using modeling grammars that provide various, often graphical, constructs to model different types of phenomena (Wand and Weber 2002). Depending on the type of grammar chosen, the focus may be on important things in the real-world and their properties, which are important to know to understand the data structure of an information system (Weber 1997). Other grammars focus on behaviors and dynamics of events and resulting actions in the real-world; these are important to

understand how processes can be modeled or enacted in an information system (Dumas et al. 2013).

Research in this area has been three-fold. One stream of research has examined how conceptual modeling grammars might be improved such that the ability is enhanced to develop good conceptual models with them. This line of research has established and examined design principles as well as guidelines for the use of conceptual modeling grammars, e.g., Evermann and Wand (2005), Mendling et al. (2010), and Figl et al. (2013a). A second stream of research is examining conceptual modeling in practice and establishes findings about their usefulness (Recker et al. 2011; Becker et al. 2016) or purposes and challenges of its use, e.g., Fettke (2009) and Indulska et al. (2009). A third stream of research, which is most relevant to this paper, is examining the conditions which determine how well a conceptual model is understood by those using it. This research, to date, has largely examined semantics (i.e., the meaning of constructs in a model, e.g., Weber (1997)), syntax (i.e., the rules about how a model can be constructed with a grammar, e.g., Reijers et al. (2011a) and Mendling et al. (2010)) and to a much lesser extent pragmatics (i.e., how existing user knowledge may influence how a model is understood, e.g., Khatri et al. (2006) and Khatri and Vessey (2016)). Some studies, finally, have attempted to review the relevant works in these areas, e.g., Burton-Jones et al. (2009), with the aim to provide guidelines for future research. Researchers from these and other studies, e.g., Recker et al. (2014), have repeatedly lamented that not enough emphasis has been on non-model related factors, in particular on *MVCs* (Gemino and Wand 2003), which is the reason we undertook the work reported in this paper.

2.2 Prior Research on Conceptual Model Comprehension

Several authors suggested that the *comprehension* of a conceptual model can be considered as the outcome of a learning process that requires model viewers to actively organize and integrate the model information with their own knowledge and previous experience (Gemino and Wand 2003; Burton-Jones and Meso 2008; Mayer 2009). This conceptualization explicitly emphasizes the role of MVCs as one important factor for model comprehension. Still, in the wider information systems field of research, the emphasis on MVCs in conceptual model comprehension studies has been cursory at best. This may be because the emphasis of this discipline has been foremost on the model as a representation artifact of an information system - the identity core of the discipline (Weber 2006). Or, it may be that no strong theory base has been available yet to conceptualize MVCs and their influence.

To substantiate our argument, we reviewed the literature on model comprehension studies to examine whether and how relevant operationalizations of MVCs have been previously used. Table 1 summarizes this review. It describes specifically whether and how available studies included variables to capture and consider MVCs in their research models, analyzes, and results.

Table 1 is a comprehensive (not exhaustive) classification of prior studies in this area. It highlights several points relevant to this paper. First, it shows that comprehension as a dependent (*affected*) variable can be examined in terms of effectiveness (accuracy of comprehension) and efficiency (resources required to attain comprehension) (Burton-Jones et al. 2009). Traditionally, the question of effectiveness or accuracy of comprehension has been of predominant interest (Bodart et al. 2001; Burton-Jones et al. 2009). The reason for this is that the extent of comprehension is a key quality criterion for all model-based problem solving tasks, over and above the question how much time is available to the analysts in developing this understanding. Furthermore, task completion time is a dimension that has to be considered (Gemino and Wand 2004). Second, Table 1 allows us to develop three key arguments that characterize our current understanding of MVCs and their impact, which are important to our study:

1. The studies differ vastly in their consideration of MVCs in their experimental settings. Most studies to date include MVCs as a control variable, if at all. Only recently have MVCs been considered in some attempts as an independent factor.

2. The studies to date have used inconsistent measures to operationalize MVCs. Some rely on self-report scales, others use counts of experience in years or number of models.

3. The studies to date rely on different cohorts as participants, with under-graduate students being used in the majority of studies. Comparison of results across the participant cohorts is difficult because of the differences between the participant groups used in the studies. There are notably few studies that involve practitioners in their sample.

The deeper exploration of these three arguments is the aim of our work. First of all, we wish to establish the significance of MVCs for predicting model comprehension performance. Second, it seems important to examine the different operationalizations of MVCs in more detail. This paves the way to offer a better conceptual definition of MVCs in modeling plus an appropriate empirical examination of the effects of various dimensions of MVCs on comprehension performance. Third, we wish to reflect on different cohorts of participants in experiments, notably whether we can rely on students as adequate proxies for

Table 1 Prior model comprehension experiments and the consideration of MVCs

Ref.	Measures for MVCs	Variable	Study context	Dependent variable	Sample population	Findings
Allen and Parsons (2010)	Years of experience in database query formulation; Self-reported level of domain familiarity	Covariate	Database query formulation	Query formulation accuracy; Task completion time; Self-reported confidence in query accuracy.	Under-graduate students	Domain familiarity leads to increased query formulation accuracy. Domain familiarity does not lead to decreased task completion time. Domain familiarity leads to increased confidence.
Bera (2012)	Experience with BPMN modeling; Self-reported level of domain	Control	BPMN models with and without swimlanes	Comprehension accuracy; Problem solving score	Under-graduate students	Novices suffer from cognitive overload but swimlanes improve on this
Bera et al. (2014)	Self-reported level of domain knowledge	Moderator	Importance of clear semantics in data models	Comprehension performance	Post-graduate students	The benefit of clear semantics is greatest when readers have moderate knowledge of the represented domain.
Bowen et al. (2009)	University course grade point average	Covariate	Database query formulation	Semantic correctness of query formulation; Task completion time; Self-reported confidence in query accuracy	Post- and under-graduate students	University course performance is a significant predictor of semantic query correctness. University course performance is a significant predictor of task completion time. University course performance is a significant predictor of confidence in query accuracy.
Burton-Jones and Meso (2008)	Experience with UML modeling; Problem-solving; Ease of understanding	Control	UML models	Comprehension accuracy	Post-graduate students	Negative correlations with comprehension and problem-solving
Figl et al. (2013a)	Work experience; Hours of training; Self-reported modeling confidence; Self-reported domain knowledge	Control	Process models with different symbols as derived from a set of different languages	Comprehension accuracy; Task completion time	Under-graduate students	Notational deficiencies in symbol sets lower comprehension
Figl et al. (2013b)	Knowledge test score	Covariate	Effect of routing symbol design in process models	Comprehension accuracy; Time; Perceived cognitive load; Perceived control flow comprehension difficulty	Post- and under-graduate students	Perceptual discriminability and pop out affect perceived cognitive load and model comprehension accuracy but not comprehension efficiency
Gemino and Wand (2005)	Familiarity with modeling; Problem-solving	Control	UML models	Information retention	Under-graduate students	Not significant as covariate
Khatri et al. (2006)	Test in data modeling course; Problem-solving	Independent	Data models	Comprehension accuracy	Under-graduate students	Significant

Table 1 (continued)

Study						
Khatri and Vessey (2016)	Model comprehension task score; Self-rated domain familiarity	Independent	Data models	Schema-based and inferential problem solving	Under-graduate students	The two types of domain knowledge (model comprehension and domain familiarity) have similar roles in explaining problem-solving
Kummer et al. (2016)	National culture; Self reported BPMN familiarity; Knowledge test score	Independent; Control; Control	Syntax highlighting in process models	Comprehension accuracy; Perceived difficulty; Task completion time	Post-graduate students	Colors in Asian cultures aid comprehension, in Germanic cultures different effects.
Lukyanenko et al. (2014)	None	Not available	information completeness in citizen science	Number of observations	Citizen scientists; Number of new observations	Not available
Mendling et al. (2012)	Knowledge test score; Years of Experience; Intensity of process model use	Independent	Abstract and domain process models	Comprehension performance; Task completion time	Post-graduate students	Knowledge correlated with comprehension accuracy. Knowledge and intensity correlated to completion time.
Parsons (2011)	None	Not available	Property precedence	Comprehension performance; Task confidence	Under-graduate students	Not available
Recker (2013)	Knowledge test score	Covariate	Usefulness of gateway constructs	Comprehension performance; Time	Post-graduate students	Use of gateways positively affects comprehension
Recker and Dreiling (2011)	Number of models created	Covariate	EPC vs BPMN models	Comprehension performance; Problem-solving; Task completion times	Post-graduate students	Not significant for comprehension Interaction effects for problem-solving
Recker et al. (2014)	Prior Domain Knowledge; Prior Method Knowledge; Self - efficacy	Control	Process model comprehension as a learning process	Comprehension scores	Modeling practitioners	Cognitive abilities, different learning styles and strategies affect level of comprehension
Reijers and Mendling (2011)	Knowledge test score; Intensity of process model use; Education	Independent	Process model complexity	Comprehension accuracy	Post- and under-graduate students	Education yields difference in comprehension.
Reijers et al. (2011a)	Modeling expertise (novice or expert)	Independent	Syntax highlighting in process models	Comprehension accuracy; Task completion time	Post-graduate students and industry experts	Syntax highlighting aids comprehension accuracy for novices but not for experts. No effects on task completion time.
Shanks et al. (2010)	None	Not available	Classes of things	Comprehension performance; Problem-solving; Discrepancy checking; Comprehension time	Post-graduate students with working experience	Not available
Shanks et al. (2008)	None	Not available	Part-whole relationships	Problem solving; Task completion time; Ease of use	Post-graduate students	Not available

modeling practitioners. To achieve these aims, we now report on the design and execution of an experiment designed as an empirical review.

3 Research Method

To examine the role of MVCs and the measurement thereof in explaining how well users understand conceptual models, several options exist. Our specific objective was to evaluate measures of MVCs and their impact on individuals' understanding of conceptual models. To that end, an experiment appears to be the best choice, also because it is congruent to past research in this area. In what follows, we describe relevant design choices about this experiment.

3.1 Design

We implemented our study as an online quasi-experiment (Wohlin et al. 2000) that featured MVCs as within-subject variations and layout (good versus bad) as between-subject variations. Our study classifies as a quasi-experiment because a random assignment of participants to groups was not feasible. Instead, we collected and examined several key demographic variables to evidence an appropriate variety in the responses. The main dependent variable was performance in model comprehension tasks, that is, tasks designed to measure how well participants understand a conceptual model presented to them.

The first main design decision concerned the type of conceptual model to use in a model comprehension task. We chose process models as the type of conceptual model. This decision was based on the fact that process models, unlike most other forms of conceptual models, are not the sole domain of IT experts. Instead, they are meant to be used by a large variety of business users with little or no training in IT, analysis and design methods, let alone process modeling. For example, the BPMN specification notes that its primary goal is to provide a notation that is readily understandable by all business users, from business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes (OMG 2010). In turn, these models are in use by a largely heterogenous group with various levels of knowledge and experience - which makes the consideration of MVCs in studies an important precondition for ecological validity of the findings. Also, we complement existing studies relating to object or data models, e.g. Aguirre-Urreta and Marakas (2008), with our work.

To tease out performance differences among the participants, we implemented the treatment of good and bad layout as between-subject variations. This choice was informed by prior research on the relationship between secondary notation and model comprehension. The term *secondary notation* refers to visual cues that are not part of the actual modeling grammar (Petre 1995). Several studies have shown that particularly *layout* represents a secondary notation aspect that affects model comprehension (Purchase 1997; Purchase et al. 2001, 2002; Petre 2006; Turetken and Schuff 2007). Among others, these studies suggest that models should avoid line crossings, that constructs should be arranged symmetrically, and that semantically related constructs should be placed close to each other. Besides the general importance of secondary notation for model comprehension, Schrepfer et al. (2009) also discuss the relationship between secondary notation and MVCs. They argue that in particular inexperienced model viewers may benefit from good layout.

3.2 Participants

Suitable subjects for our study were individuals with previous experience in using process models. To examine the MVC differences between two subject groups that are typically encountered in experimental studies, we sought to recruit subjects from two populations: students and practitioners, both of which we could reach via our interactive website. In this way, we could also implement direct feedback on which answers were correct, which allowed for a gamification of the experiment (by assigning scores and ranks to participants in comparison to others) to incentivize participation as well as performance.

To recruit student subjects, graduate students were recruited from our ongoing and prior courses on business process modeling at TU Eindhoven, HU Berlin, QUT Brisbane, and WU Vienna. We sent out invitations via course coordinators to those students that previously received at least basic training on process modeling concepts and theory. They were invited via email and motivated with the hint that answering the comprehension questions and learning from the online feedback would be a good preparation for unit exams. We were not ourselves involved in the teaching of these courses and had no control over course marks or exam composition.

Practitioners were recruited through an international advertisement campaign using news forums, practitioner communities, and special interest groups on the internet. The invitation was posted, among others, in relevant XING and LinkedIn groups, Bruce Silver's BPMN blog, the ARIS process modeling community, and the BPMN Forum. Practitioners were attracted by the challenge to test their understanding of the BPMN standard and the prospect of receiving feedback on the results.
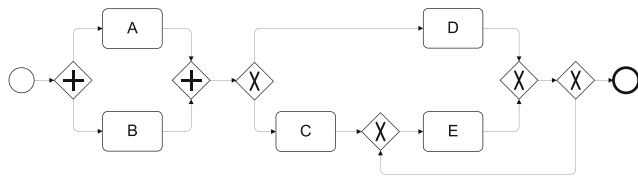
Fig. 1 Example of a simple BPMN process model

3.3 Materials

In the experiment, we designed five BPMN process model comprehension tasks.[1] The comprehension challenge with BPMN, as with other types of process models, is to correctly understand the *control flow* between different activities (e.g. A, B, C, D, E in Fig. 1). The control flow of a process model defines temporal and logical constraints between those activities, for instance whether the execution of certain activities depends upon decisions, whether paths are concurrent, or whether activities can be repeated. Control flow is the key mechanism to describe behavioral dynamics in processes (Dijkman et al. 2008) and is the key differentiator to other forms of modeling, such as object structures or data relationships. Control flow aspects are modeled in BPMN using so-called gateways. In essence, there are gateways with three different routing logics (XOR, AND, OR), which can be used as splits (multiple outgoing arcs) or joins (multiple ingoing arcs). XOR-splits represent exclusive choices and XOR-joins capture respective merges without synchronization. AND-splits introduce concurrency of all outgoing branches while AND-joins synchronize all incoming arcs. OR-splits define inclusive choices of a 1-to-all fashion. OR-joins synchronize such multiple choices, which requires a sophisticated implementation (Kindler 2006; Mendling 2008). Furthermore, there are specific nodes to indicate the start and end of a process.

Figure 1 shows the example of a simple BPMN process model. The process starts at the left with a start event (a circle with a thin line). Then, an AND-split introduces two branches of concurrent execution. Accordingly, both A and B are activated and can be executed without any order constraints. The subsequent AND-join synchronizes the two branches. Once both have been completed, a decision can be taken at the following XOR-split: the process has to continue either with C or D. If C is taken, the process continues via the XOR-join for executing activity E. Otherwise, only D is executed. In either case, the XOR-join after D and E leads to an XOR-split. There is the option to jump back to execute E, potentially multiple times.

In our study, we focused on behavioral constraints that can be derived from the process model. Domain content,

which is typically included in textual descriptions of activity labels, is ignored. The advantage is that there is an objective basis for judging process model comprehension and effects of domain knowledge are eliminated (Reijers and Mendling 2011). Also, a recent study showed that control flow comprehension is hindered by the presence of domain information (Mendling et al. 2012), which would have masked some of the effects and results that we are interested in in this work.

We utilize binary relationships between two activities in terms of execution order, exclusiveness, concurrency, and repetition. These relationships play an important role for reading, modifying, and validating the model.

- *Execution Order* relates to whether the execution of one activity a_i always happens before the execution of another activity a_j. In Fig. 1, the execution of A is always executed before D.
- *Exclusiveness* means that two activities a_i and a_j can never be executed in the same process instance. In Fig. 1, C and D are mutually exclusive. Note that D and E are not exclusive, since there is the option to jump back to E after having executed D.
- The *concurrency* relation covers two activities a_i and a_j if they can potentially be executed in an arbitrary order. In Fig. 1, A and B are concurrent. This means that the execution of A can precede that of B or the other way around.
- A single activity a is called *repeatable* if it is possible to execute it more than once for a process instance. In Fig. 1, E can be repeated.

Statements such as "Activity a_i can never be executed before a_j" can be formalized and verified using behavioral profiles, which capture the four different relationships described above (Weidlich et al. 2011). Most of the studies to date have perused these questions as comprehension measurement instruments, notably because they allow for an objective measurement of control flow comprehension, e.g., Reijers and Mendling (2011), Reijers et al. (2011a), and Mendling et al. (2012). A key question that emerges now is how individuals with different levels of abilities and skills can identify, comprehend, and reason about these control flow aspects in a process model. This is important because this understanding is essential for any subsequent deeper problem-solving task. In designing the BPMN models used for the comprehension tasks, we neutralized potentially confounding impacts of the notation (Sarshar and Loos 2005), model complexity (Mendling 2008), domain knowledge (Mendling et al. 2012), and the modeling purpose (Dehnert and van der Aalst. W.M.P. 2004) by choosing letters as activity names and models of comparable size.

---

[1]The original number of eight models was reduced after piloting in order to avoid potential fatigue.

To implement the treatment of layout, we created a well laid out version and a badly laid out version of each model. The variation was guided by well-known esthetic metrics from graph drawing research by Purchase et al. (1997), Ware et al. (2002), Purchase et al. (2002), Petre (2006), and Purchase (2014):

– *Line crossings:* An increase in the number of crossings has been found to decrease the readability of that layout. The models with good layout have no crossings while the bad layout ranges from 4 to 23 crossings.
– *Edge Bends:* An increase in the number of edge bends has been found to negatively affect the understanding of a model. The models with good layout have 29 to 61 bends while the bad layout ranges from 89 to 119 bends.
– *Symmetry:* Graphical layouts where elements are placed more symmetrically have been found to be easier to read. While the good layout models have only three violations of symmetry altogether, the badly laid out models have between 6 and 14 repositionings of elements that break symmetry.
– *Use of Locality:* When graphical elements that are related are placed close to each other, it is apparently easier to understand their connection. The models with good layout break locality between 3 and 7 times, the bad ones between 10 and 32 times.
– *Reading direction:* The direction of arcs should be in line with rightwards reading direction. The models with good layout have between 3 and 9 arcs pointing rightwards, the bad ones between 13 and 25 such arcs.

Participants were then randomly assigned to one of the following settings:

– *Setting A*: M1 (good) - M2 (bad) - M3 (good) - M4 (bad) - M5 (good)
– *Setting B*: M1 (bad) - M2 (good) - M3 (bad) - M4 (good) - M5 (bad)

The participants had to answer six questions per model, i.e. 30 questions altogether, with the questions focusing on aspects such as exclusiveness, concurrency, and optionality of activities, as described above. These aspects are based on the formalization of control flow principles by Kiepuszewski et al. (2003) and comparable to the ones previously used by (Mendling et al. 2012). The comprehension tasks were formulated as a statement about how two or three activities in a process model relate to each other in terms of the aforementioned behavioral relations.

In answering the comprehension tasks, subjects were allowed to look at the process models when answering questions, rather than requiring that they work from memory.

3.4 Measurement

As independent variables, we operationalized an extensive set of MVCs identified from our literature review. Table 2 gives an overview of the measures we consider in our experiment and their previous applications in related studies. In the experiment, we gathered the following data for each participant using self-reports following past practice (Gemino and Wand 2005; Mendling et al. 2012):

Position: This variable captured the primary occupation of the participant in relation to the study on process modeling. Answer choices were "student", "practitioner", or "other".

Theory: This variable captures the sum of correct answers in a process modeling knowledge test as used by (Reijers and Mendling 2011; Mendling et al. 2012). This test allows to capture the a priori knowledge of different control flow concepts relevant to process modeling. Participants are asked to answer seven questions on fundamental concepts of process modeling, including concurrency, gateways, repetition, and choices.

ModelingYears: This variable is a self-reported account of how long ago a participant started with process modeling ("How many years ago did you start process modeling?"). The variable is recorded on a metric scale.

ModelsRead: This variable captures the intensity of working with process models the last twelve months on a metric scale ("How many process models have you analyzed or read within the last 12 months?").

ModelsCreated: This variable defines the intensity of editing process models in the last twelve months ("How many process models have you created or edited within the last 12 months?").

Training: This independent variable records the degree of formal eduction in the last year ("How many work days of formal training on process modeling have you received within the last 12 months?").

SelfEducation: This variable captures self training, which may be acquired through learning-by-doing, or self-study of textbooks or specifications ("How many work days of self education have you made within the last 12 months?").

FAM1-3: This set of metric items captures familiarity with BPMN using a seven point Likert scale (FAM1: "Overall, I am very familiar with the BPMN.", FAM2: "I feel very confident in understanding process models created with the BPMN.", FAM3: "I feel very competent in using the BPMN for process modeling.").

MonthsBPMN: This variable assesses how long ago a participant started using BPMN ("How many months ago did you start using BPMN?").

Table 2 MVCs as a model comprehension variable in the literature

Variable name	Measure	Scale	Relevant literature
theory	Level of theoretical knowledge of modeling	Metric	(Mendling et al. 2012) (Reijers and Mendling 2011) (Khatri et al. 2006)
modelingYears	Years of experience in modeling	Metric	(Recker 2010a)
modelsRead	Number of models read or analyzed within twelve months	Metric	(Recker and Dreiling 2011) (Reijers and Mendling 2011) (Mendling et al. 2012)
modelsCreated	Number of models created or edited within twelve months	Metric	(Reijers et al. 2011b) (Davies et al. 2006)
training	Days of modeling training within twelve months	Metric	(Recker 2010a)
selfEducation	Days of modeling self education within twelve months	Metric	(Recker 2010a)
FAM1, FAM2, FAM3	Level of familiarity with a specific modeling grammar	3-item metric	(Gemino and Wand 2005) (Burton-Jones and Meso 2008)
monthsBPMN	Number of months working with a specific modeling grammar	Metric	(Recker 2010a)

To develop measurements for the dependent variable of *conceptual model comprehension performance*, we automatically recorded the number of correct answers for each of the five process models presented. This provided a measure for *comprehension accuracy*. Each of the comprehension questions had an objectively correct answer, which could be answered based on behavioral semantics of the process model (Weidlich et al. 2011). Accordingly, we define the following dependent variables:

Performance: This variable captures the extent of model comprehension accuracy. It is calculated as the sum of correct answers given by the participant for comprehension questions that relate to a particular model. The maximum value for this variable is six, which results from six yes/no questions for each of the models.

Completion Time: This variable captures the time for completing a specific comprehension task. It is calculated as the sum of completion times of the participant for comprehension questions that relate to a particular model.

The Appendix details the tasks and measures used.

3.5 Procedures

The experiment proceeded through a sequence of three tasks:

1. subjects were to self-assess different MVC-related measures,

2. subjects were to answer a theoretical knowledge test, and

3. subjects were to answer comprehension questions for five process models (layout setting A or B).

All tasks were implemented in an online experimentation system accessible on a website. This implementation allowed us to source participants cross-sectionally and globally. However, it also meant that we had to take specific measures to stimulate participation and mitigate potential cheating. The adoption of BPMN Version 2.0 as an OMG standard in 2010 helped us to direct considerable attention to an experiment, which we hosted on-line as a self-test for BPMN with immediate feedback[2]. A screenshot from the experiment is shown in Fig. 2. As the name suggests, the website was presented as a tool to test one's own understanding of BPMN process models and designed such that feedback was given about errors made as well as relative performance ranking in comparison to others. Specifically, after completion of the test, a participant got feedback in different ways. First, a table was listed with the questions that were answered incorrectly along with an explanation on the correct answer. Second, each participant could see at which position of a high score list he or she ranked. Third, we offered a BPMN education course for those participants who provided their email contact details.

As the high score ranking bore the risk that participants would interpret the selftest as a competition, we had to impose countermeasures in order to avoid getting biased

---

[2]This website was hosted at http://www.bpmn-selftest.org, but is no longer available online.

Fig. 2 A BPMN model with a comprehension question as displayed in the experiment
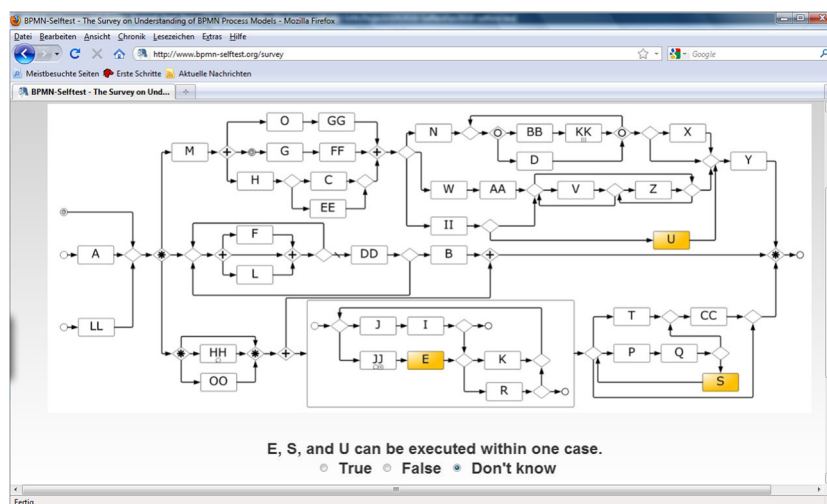
data. First, we included a select bottom before the test, where a participant had to state whether he had done the selftest before. In this way, we aimed to filter out repeated participants. In total, 59 participants were identified as repeaters using this select button. Second, to cover for the case that participants would conduct the selftest a second time without acknowledging repetitive participation, we introduced measures to minimize potential learning effects. Therefore, we used two versions of each model with a slight variation in layout and randomly sampled the 30 questions. Finally, we highlighted the activities corresponding to the comprehension question in order to minimize search time.

The selftest not only recorded results of answering questions, but also demographic data and data on MVCs. We conducted several pilots to make sure that the questionnaire could be completed within 30 minutes, that questions were comprehensible, and that the models could be easily viewed on different screen sizes and with different internet browsers. The website was online from November 2009 until October 2010.

4 Results

We analyzed our data in four steps. First, we checked descriptive statistics and cleansed the data for outliers. Second, we examined the data to investigate the different measures that have been proposed for MVCs. Third, from the data we develop MVCs that characterize student and practitioner subjects. Fourth, we evaluate effects of MVCs on comprehension task performance and on completion time.

4.1 Data Cleansing

As a first step we evaluated descriptive statistics with the view to perform different data cleansing and filtering operations. First, we had to drop data from those participants who did not fully complete the experiment. Altogether, 2199 persons started, out of which 778 completed it. The resulting completion rate is 35%, which can be considered reasonable as compared to web-survey rates of often less than 15% (Porter and Whitcomb 2003). Furthermore, we had to filter out suspicious data points. From pilots, we learned that it would require at least five minutes to complete the full set of 30 comprehension questions for a highly skilled process modeler. In case someone was faster, this was seen as an indication of clicking through the experiment without engaging in depth with the tasks. Therefore, we eliminated all participants who completed the experiment in less than five minutes. For this reason, 87 data points were filtered out. We also had to drop cases where participants would be interrupted while working on the selftest. If a participant took longer than 60 minutes, we decided to exclude the case in order to avoid distortions. In this way, we dropped a further 22 cases. Next, we inspected the data for outliers in two steps. First, we conducted an outlier analysis using the Stem-and-Leaf plots for the self-report data in order to identify suspicious data points. For the two independent variables `ModelsRead>` 500 (three participants) and `ModelsCreated>` 200 (two participants), we kept the data points because they showed high theory and performance values as would have been expected for an intensive modeling practice. Second, we followed the recommendation by Wohlin et al. (2000) to judge data points according to whether they are reasonable from a domain perspective. The conditions we define in the following are stricter than those being proposed by the formal outlier analysis. As a result, we excluded data points if one of the following conditions was satisfied:

– `Theory<` 2: We were not interested in answers from participants without any knowledge in process

modeling. We were conservative in that we omitted not only participants that scored 0 out of 7 questions correct but also those that scored one correct answer because this may have been due to chance. This condition was the case for 42 participants. For all remaining participants, we checked whether the theory score could be ascribed to random guessing. However, a one-sample t-test with 3.5 as the test score showed that our population scored significantly higher (p = 0.00, t = 14.81) than what would have been the result of guesswork.

– `ModelingYears> 40`: Assuming a life-long career in process analysis, it is unlikely to have more than 40 years of modeling experience. This condition holds for 6 participants.

– `Training> 75`: Formal training is only realistic for a limited amount of time in a year. We assume a full-time study programm to yield the maximum reasonable value. In case of two semesters of 15 weeks lecturing, each including five days of half-day lectures, this amounts to 75 work days. Three data points are beyond this value.

– `SelfEducation> 180`: An outliers test with SPSS suggested all cases above 36.50 to be outliers and all values of 200 and above to be extreme values. To settle for a not overly restrictive cutoff, we assumed a person starting as a professional process analyst would not reasonably be able to consider more than every second day of the year as self-education. Thus we eliminated eight participants with reported self-education of 180 or more.

Furthermore, 59 cases of repeaters were excluded. Finally, we excluded those 21 participants who stated they were neither student nor practitioner. For instance, a number of academics researching in the field of Business Process Management completed the experiment, and were excluded from the analysis. As some data points were identified by more than one filter, we got a cleansed data sample of 530 participants, which forms the basis for the analyzes reported below.

4.2 Defining MVCs

To determine whether MVC profiles can be defined on the basis of the traits we identified, we performed an exploratory factor analysis of the considered MVC measures. This is an appropriate technique to uncover the underlying factor structure of a large set of variables without a priori specifications of the number of factors and their loadings (Hair et al. 2010). We performed this exploratory factor analysis with four goals in mind:

1. to examine different MVCs,

2. to explore the validity and reliability of the measures used,
3. to reduce the set of variables to appropriately weighted factors resembling the different MVCs, and
4. to explore the effects of these factors on model comprehension performance and completion time below.

First, we explored whether the data distribution assumptions of exploratory factor analysis were met (Hair et al. 2010). The Kaiser-Meyer-Olkin measure of sampling adequacy was above 0.50 (0.75), and Bartlett's test of sphericity was significant at $p = 0.00$ with $df = 36$. Thus, the use of exploratory factor analysis was warranted.

We used a principal component analysis with Varimax rotation to identify factor structures with Eigenvalues greater than 1. Several iterations of the factor analysis were conducted to identify and eliminate problematic measurement items. During this process, it became apparent that one item ("number of months working with a specific modeling grammar") did not load appropriately on any factor. By excluding this item from the analysis, a strong 5-factor solution emerged, which we summarize in Tables 3 (descriptive statistics), 4 (item factor loadings), 5 (properties of the emerging factors), and 6 (factor correlations).

Our analysis yielded five different factors, which we define as follows:

– *Familiarity*: the extent to which individuals perceive themselves to be familiar with process modeling.
– *Intensity*: the extent to which individuals engage in process modeling within a given timeframe.
– *Education*: the extent to which individuals received formal education in process modeling.
– *Knowledge*: the extent to which individuals possess knowledge about process modeling concepts.
– *Duration*: the extent to which individuals have done process modeling in the past.

All items showed adequate reliability: Cronbach's α and p_c exceeded 0.7 for all multiple-item factors except for

Table 3 Descriptive statistics of independent variables

Variable	Mean	St. Deviation	Min	Max
Theory	4.66	1.37	2	7
ModelingYears	4.17	6.62	0	37
ModelsRead	36.36	54.44	0	500
ModelsCreated	17.05	26.66	0	180
Training	5.13	9.51	0	70
SelfEducation	12.57	21.44	0	180
FAM1	3.59	1.67	1	7
FAM2	3.16	1.46	1	7
FAM3	3.58	1.49	1	7

Table 4 Factor loadings

| Variable | Component | | | | |
	Familiarity	Intensity	Education	Knowledge	Duration
Theory	−0.06	−0.06	−0.03	**0.95**	0.01
ModelingYears	0.00	0.01	0.00	0.00	**0.99**
ModelsRead	−0.26	**0.91**	0.01	0.00	0.05
ModelsCreated	−0.14	**0.93**	0.03	−0.07	−0.03
Training	−0.06	−0.12	**0.80**	−0.27	0.00
SelfEducation	0.03	0.15	**0.79**	0.27	−0.01
FAM1	**0.91**	−0.02	−0.02	−0.03	−0.03
FAM2	**0.95**	−0.03	0.03	−0.01	0.01
FAM3	**0.94**	−0.04	−0.01	0.01	0.01

Bold entries are statistically significant at $p < 0.05$.

education. The low value for education is not a surprise because of the formative character of this factor. Education is based on formal training and self education, which can only be partially expected to correlate. For the single-item factors, the commonality h^2 exceeded 0.9 respectively. The standard deviations of all factors were above 1, suggesting adequate variance in the scales. All factors were correlated with each other, with the highest correlations being between *Intensity* and *Familiarity* (-.37), see Table 6). Internal consistency, discriminant validity, and convergent validity were tested by extracting the factor and cross loadings of all indicator items to their respective latent constructs. The results, presented in Table 4 and Table 5, indicate that all items loaded on their respective construct from a lower bound of 0.79 to an upper bound of 0.99, and higher on their respective construct than on any other. Furthermore, each item's factor loading on its respective construct was highly significant (at least at $p < 0.01$). Convergent validity was further supported by all composite reliabilities p_c being 0.83 or higher and AVE of each construct being 0.85 or higher. Discriminant validity was supported by showing that the AVE of each construct was higher than the squared correlation between any two factors (the highest squared correlation was 0.15 between *Intensity* and *Familiarity*). In turn, our analysis yielded five largely disjoint traits of

Table 5 Factor properties

Factor	Reliability	p_c	Communality	AVE
Familiarity	0.93[a]	0.99	0.88	0.94
Intensity	0.73[a]	0.97	0.86	0.93
Education	0.33[a]	0.83	0.73	0.85
Knowledge	0.93[b]	–	0.91	0.95
Duration	1[b]	–	1.00	1.00

[a]Based on Cronbach's α.

[b]Based on Commonality h^2 as suggested by Christophersen and Konradt (2011).

MVCs. In the following, we will thus explore how these traits enable us to explain and predict model interpretation in terms of comprehension performance across students and practitioners.

4.3 MVCs of Students and Practitioners

Having identified five MVC traits, we now seek to explore the profile of participants that belonged to one of two key user groups (viz., students and practitioners). In particular, we aim at examining the typical MVCs associated with these groups and how these profiles differ across the groups. Our interest was to ascertain whether practitioners and students as experimental subjects vary substantially, as suggested (Compeau et al. 2012). If so, it would be interesting to investigate how MVCs, as identified above, might be used to discriminate between students and professionals.

To that end, we ran a multivariate analysis of variance (MANOVA), with the variable *Position* as an independent factor and the five total factor scores of the identified MVC traits as dependent variables. As a preparatory check, we computed regressed total factor scores, all of which behaved approximately normal and independent samples t-test between the MVC variables by students and practitioners, the results of which were significant for some variables. Table 7 gives the MANOVA results. The data shows that, indeed, the profiles of the two subject groups are significantly different for each modeling MVC dimension, except for the factor *Education*. Table 7 illustrates that students appear to have a higher *Familiarity* with BPMN and a slightly higher *Education* (but not significantly so). By contrast, practitioners score considerably and significantly higher on *Intensity*, *Knowledge*, and *Duration*.

Altogether, it can be seen that the profiles between the two user cohorts are, except for *Education*, largely and significantly different, as visualized in Fig. 3. Notably, practitioners have longer and more intensive engagement in

Table 6 Factor correlations

Factor	Familiarity	Intensity	Education	Knowledge	Duration
Familiarity	1.00				
Intensity	−0.37	1.00			
Education	−0.14	0.09	1.00		
Knowledge	−0.13	0.16	−0.02	1.00	
Duration	−0.19	0.14	−0.03	0.05	1.00

process modeling work (*Intensity* and *Duration*), whereas students have higher perceptions of *Familiarity*. Differences in *Knowledge* and *Education* appear marginal.

4.4 Effects of MVCs

At this point, we have seen that MVC profiles of students and practitioners are largely different. In our final analysis, we now examine how well the identified five MVCs allow the explanation of model comprehension performance and completion time.

To that end, we first computed the weighted total factor scores for the five MVCs emerging from our factor analysis. Table 8 summarizes descriptive statistics about the factors. All factors were normalized to a mean of 0.00 (and thus have a standard deviation 1.00).

Next, we performed two MANOVA, one with comprehension performance (*comprehension*) and one with completion time (*time*) as dependent variable. In both analyses, we included *Position* and *Layout* as independent variables and the five identified MVC factors as covariates. Box's test of equality of covariance matrices had a value of 0.175 indicating that the assumptions are met. Levene's test of equality of error variances was indicating a violation of the assumption except for Model 4.

4.4.1 Effects of MVCs on Task Performance

Tables 9 and 10 illustrate the results of the comprehension performance analysis. The results show that particularly *Layout* is a significant factor for predicting comprehension performance. Both students and practitioners perform significantly better when a model is laid out well. While this effect is not significant for Model 5, we still observe that the Model 5 comprehension performance of the students is higher for the well laid out model and that the Model 5 comprehension performance for good and bad layout of the practitioners 5 is almost equal. The importance of *Layout* is further highlighted by the fact that the variable *Position* cannot explain the differences in comprehension performance for any of the models. While we do see a slightly better performance of the practitioners, this difference is statistically insignificant.

With respect to the identified MVCs, we observe that *Familiarity*, *Intensity*, and *Knowledge* can be used to explain performance differences. Both *Familiarity* and *Knowledge* are significant for all 5 models. *Intensity* is significant for models 1, 3, 4, and 5. Interestingly, *Education* and *Duration* are not significant for any of the models. This shows that neither the formal education the participant has received nor the extent to which the participant has done process modeling in the past are good

Table 7 Multivariate analysis of variance on MVCs

Dependent Var.	Group	N	Mean	St. Dev.	F (Sig.)	η_p^2
Familiarity	Students	333	0.26	1.00	**68.28**	0.12
	Practitioners	197	-0.44	0.82	**(0.00)**	
Intensity	Students	333	-0.24	0.56	**58.44**	0.09
	Practitioners	197	0.41	1.38	**(0.00)**	
Education	Students	333	0.02	1.01	0.62	0.01
	Practitioners	197	-0.04	0.98	(0.43)	
Knowledge	Students	333	-0.08	1.00	**6.17**	0.01
	Practitioners	197	0.14	0.97	**(0.01)**	
Duration	Students	333	-0.17	0.82	**28.31**	0.05
	Practitioners	197	0.29	1.18	**(0.00)**	

Bold entries are statistically significant at $p < 0.05$.
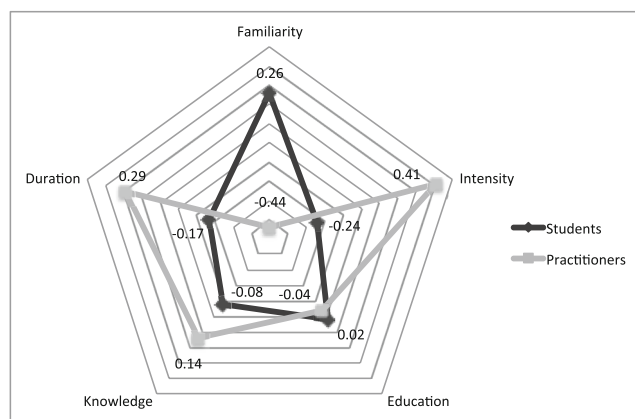
Fig. 3 Average values of MVCs of students and practitioners

Table 9 Descriptive statistics about comprehension performance

Model	Layout	Students		Practitioners	
		Mean	STD	Mean	STD
Model 1	good	4.54	1.43	5.02	1.35
	poor	4.20	1.48	4.48	1.28
Model 2	good	4.07	1.56	4.75	1.29
	poor	3.96	1.37	4.29	1.43
Model 3	good	4.00	1.33	4.19	1.28
	poor	3.56	1.49	3.75	1.38
Model 4	good	3.41	1.50	4.19	1.48
	poor	3.15	1.41	3.75	1.62
Model 5	good	4.10	1.55	4.53	1.51
	poor	3.63	1.69	4.57	1.48

predictors for performance in this experiment. By contrast, self-reported familiarity (*Familiarity*), actual knowledge about the notation (*Knowledge*), and intensity with which process modeling was used (*Intensity*) are good predictors for comprehension performance. Importantly, interactions hardly contribute to explaining performance differences.

Altogether, the MANOVA for comprehension performance shows that layout and a particular set of MVCs, i.e. *Familiarity*, *Intensity*, and *Knowledge*, turn out to be significant factors for comprehension performance. Interestingly, the mere fact that a participant is a student or practitioner is not significant in explaining performance differences.

4.4.2 Effects of MVCs on Completion Time

Tables 11 and 12 illustrate the results of the completion time analysis. They show that *Layout* is also an important factor for predicting completion time. The analysis reveals that all models with good layout had significantly faster completion times. This holds for both practitioners as well as students. As opposed to the comprehension performance analysis, the variable *Position* is significant for Model 1, 3, 4, and 5. Interestingly, students completed the comprehension task significantly faster than practitioners. A possible explanation for this observation may be general familiarity of students with problem solving tasks due to regular exams.

With respect to the MVCs, there is no central factor emerging from the analysis. In line with comprehension

Table 8 Descriptive statistics about factors

Factor	Range	Min	Max
Familiarity	4.73	−2.05	2.68
Intensity	6.66	−1.10	5.56
Education	6.74	−0.91	5.84
Knowledge	6.25	−2.51	3.74
Duration	4.90	−0.95	4.95

performance, we observe significant effects of *Familiarity* and *Knowledge* for Model 3 and 5, and *Intensity* for Model 4, but not for the other models. In addition, the MVC *Duration* is significant for Model 1. Altogether, it appears as if none of the MVCs is a good predictor for completion time. For interactions, there are hardly any consistent patterns that emerge. Apparently, interactions are of minor relevance also for completion time.

5 Discussion

This section discusses the implications of our research findings. Section 5.1 summarizes the results. Section 5.2 discusses implications for research. Section 5.3 clarifies potential threat to validity of this research.

5.1 Summary of Results

We set out to empirically examine how modeling expertise relates to model comprehension performance and, based on the results, develop and explore a multi-dimensional profile of modeling expertise. Table 13 summarizes the results. Note that η_p^2 have been suggested to indicate small, medium and large effects for values of 0.01, 0.06 and 0.14, respectively (Kirk 1996; Field 2013). It shows that layout plays a significant role for both comprehension performance as well as completion time. The position of a participant (student versus practitioner), by contrast, can only explain differences in completion time. As for the five MVCs, we observe that *Familiarity*, *Intensity*, and *Knowledge* can be used to explain performance differences. *Education* as well as *Duration* were not significant in this context. With respect to completion time, none of the MVCs appears to be a good predictor.

Table 10 Multiple analysis of variance on comprehension performance

Factor	Model 1		Model 2		Model 3		Model 4		Model 5	
	F (Sig.)	η_p^2	F (Sig.)	η_p^2	F (Sig.)	η_p^2	F (Sig.)	η_p^2	F (Sig.)	η_p^2
Between-Subjects										
Layout	**9.03 (0.00)**	0.02	**7.14 (0.01)**	0.01	**11.56 (0.00)**	0.02	**10.20 (0.00)**	0.02	2.75 (0.10)	0.01
Position	0.38 (0.84)	0.00	1.01 (0.32)	0.00	2.20 (0.14)	0.00	2.25 (0.13)	0.00	2.69 (0.10)	0.01
Within-Subjects										
Familiarity	**13.02 (0.00)**	0.03	**24.29 (0.00)**	0.05	**19.93 (0.00)**	0.04	**32.17 (0.00)**	0.06	**19.22 (0.00)**	0.04
Intensity	**5.21 (0.02)**	0.01	1.13 (0.29)	0.00	**10.71 (0.00)**	0.02	**5.97 (0.02)**	0.01	**4.79 (0.03)**	0.01
Education	0.00 (0.95)	0.00	0.20 (0.66)	0.00	2.81 (0.09)	0.01	0.02 (0.88)	0.00	3.09 (0.08)	0.01
Knowledge	**31.27 (0.00)**	0.06	**51.16 (0.00)**	0.06	**67.45 (0.00)**	0.09	**33.44 (0.00)**	0.06	**51.60 (0.00)**	0.09
Duration	0.30 (0.58)	0.00	0.14 (0.70)	0.00	2.61 (0.11)	0.01	0.09 (0.77)	0.00	0.01 (0.92)	0.00
Interactions										
Layout*Familiarity	**4.84 (0.03)**	0.01	0.13 (0.72)	0.00	1.31 (0.25)	0.00	1.09 (0.30)	0.00	0.56 (0.45)	0.00
Layout*Intensity	0.10 (0.75)	0.00	**4.12 (0.04)**	0.00	0.07 (0.79)	0.01	2.17 (0.14)	0.00	1.28 (0.26)	0.00
Layout*Education	0.01 (0.92)	0.00	0.07 (0.80)	0.00	0.10 (0.75)	0.00	0.05 (0.82)	0.00	3.48 (0.06)	0.01
Layout*Knowledge	2.39 (0.12)	0.00	2.04 (0.15)	0.00	0.09 (0.77)	0.00	0.03 (0.86)	0.00	1.86 (0.17)	0.00
Layout*Duration	0.12 (0.73)	0.00	1.44 (0.23)	0.00	0.36 (0.55)	0.00	0.46 (0.50)	0.00	0.49 (0.48)	0.00

Bold entries are statistically significant at $p < 0.05$.

Table 11 Descriptive statistics about completion time

Model	Layout	Students		Practitioners	
		Mean	STD	Mean	STD
Model 1	good	225.11	156.05	249.34	135.30
	poor	284.37	173.22	369.76	172.21
Model 2	good	219.70	201.90	226.93	93.09
	poor	283.03	169.95	351.69	204.83
Model 3	good	204.05	97.16	261.65	173.98
	poor	245.36	145.36	328.22	161.01
Model 4	good	270.61	223.97	313.81	198.47
	poor	293.10	158.30	398.54	252.88
Model 5	good	216.58	115.89	265.85	138.27
	poor	268.91	165.99	351.43	197.76

5.2 Implications

Our specific aim was to explore three key assumptions prevalent in the literature in this domain, viz., (1) the significance of MVCs for predicting model comprehension performance, (2) potential implications of the (often) uni-dimensional measurement of MVCs, and (3) the use of students as adequate proxies for modeling practitioners.

With respect to (1), our results show that MVCs are key factors contributing to an accurate understanding of process models. One interpretation is to discuss the utilized MVCs and their connection with expertise. Our findings support extant literature, which presumes that the cognitive load of understanding an external schema can indeed be reduced by having expertise in terms of effective storage and processing strategies for these models (Sweller and Chandler 1994). It must be noted that the overall concept of expertise is insufficiently covered by the MVCs used in this study and preceding model comprehension experiments. The complexity of expertise (Chi et al. 2014) and the diversity of task facets (Spence and Brucks 1997; Reuber 1997; Jacoby et al. 1986) call for theoretical research to complement our empirics-driven approach.

With respect to (2), we find evidence that an appropriate measurement of MVCs requires at least the recording of several factors. This is interesting because general research on expertise emphasizes its task specificity and continuous deliberate practice (Ericsson et al. 2007). Furthermore, the different MVCs used in prior literature show different effects on the development of model understanding. We empirically found that MVCs relate to different factors, some of which are connected with *comprehension performance* (*Intensity*, *Knowledge*, *Familiarity* and *Duration*) and some of which may be irrelevant (*Education*), while apparently the impact on *task completion time* is quite different (primarily driven by secondary notation, specifically layout here). Research including (Topi

and Ramesh 2002; Batra et al. 1990) can serve as a starting point to integrate these findings into a theoretical model.

With respect to (3), we find that students are distinctively different from practitioners. Specifically, our analysis shows that these groups have significantly different MVC profiles. Their performance results may differ as well. While the difference in performance appears to be more related to MVCs, the difference in completion time turns out to be significantly connected with the student or practitioner position. These results call for more research to investigate these diverging patterns. Potentially additional factors will have to be included, such as cognitive abilities (Recker et al. 2014), limitations of sight (Permvattana et al. 2013) or risk aversion (Cox et al. 2014). What becomes clear though is the fact that the discussion of the student versus practitioner dichotomy only touches the problem at the surface. The underlying factors of performance differences of these two groups appear to be related to MVC profiles.

Clearly, these points emphasize the need to develop new measurement instruments for MVCs from scratch. The measures of *Familiarity*, *Intensity*, and *Knowledge* might be building blocks to be integrated. We also see the potential of not only developing better measurement items for surveys, but for using alternative means of objective measurement. For example, recent studies on process model comprehension by Petrusel et al. (2016) and Petrusel et al. (2017) utilize eye-tracking devices and find that measures of visual cognition explain comprehension performance well. The more broader potential of neuroscience in this area is highlighted by Davis et al. (2017). As long as such new measures are not available, experimenters are advised to record *Familiarity*, *Intensity*, and *Knowledge* as used in this study and use them as covariates in their analysis. Furthermore, prior studies should be replicated with these measures being added in the data analysis. The AIS Transactions on Replication Research are an excellent outlet where these studies can be reported.

Table 12 Multiple analysis of variance on completion time

Factor	Model 1 F (Sig.)	η_p^2	Model 2 F (Sig.)	η_p^2	Model 3 F (Sig.)	η_p^2	Model 4 F (Sig.)	η_p^2	Model 5 F (Sig.)	η_p^2
Between-Subjects										
Layout	**35.41 (0.00)**	0.06	**30.47 (0.00)**	0.06	**19.28 (0.00)**	0.04	**5.96 (0.01)**	0.01	**26.82 (0.00)**	0.05
Position	**6.02 (0.01)**	0.01	3.14 (0.08)	0.01	**13.54 (0.00)**	0.03	**5.08 (0.02)**	0.01	**8.23 (0.00)**	0.02
Within-Subjects										
Familiarity	1.47 (0.23)	0.00	1.87 (0.17)	0.00	**7.21 (0.01)**	0.01	2.69 (0.10)	0.01	**4.83 (0.03)**	0.01
Intensity	0.00 (0.96)	0.00	0.10 (0.76)	0.00	0.05 (0.83)	0.00	**5.97 (0.01)**	0.01	1.85 (0.17)	0.00
Education	0.01 (0.94)	0.00	0.09 (0.76)	0.00	0.19 (0.66)	0.00	2.89 (0.09)	0.01	1.45 (0.23)	0.00
Knowledge	0.03 (0.86)	0.00	0.19 (0.67)	0.00	**5.02 (0.03)**	0.01	0.03 (0.87)	0.00	**5.50 (0.02)**	0.01
Duration	**4.69 (0.03)**	0.01	1.00 (0.32)	0.00	0.04 (0.84)	0.00	0.11 (0.74)	0.00	0.18 (0.67)	0.00
Interactions										
Layout*Familiarity	**4.06 (0.04)**	0.01	2.86 (0.09)	0.01	0.80 (0.37)	0.00	**7.76 (0.01)**	0.01	1.13 (0.29)	0.00
Layout*Intensity	0.37 (0.55)	0.00	2.99 (0.08)	0.01	0.18 (0.67)	0.00	**6.12 (0.01)**	0.01	0.25 (0.62)	0.00
Layout*Education	0.11 (0.74)	0.00	1.15 (0.28)	0.00	0.24 (0.63)	0.00	3.32 (0.07)	0.01	0.50 (0.48)	0.00
Layout*Knowledge	1.71 (0.19)	0.00	**12.03 (0.00)**	0.02	0.44 (0.51)	0.00	**8.90 (0.00)**	0.02	0.11 (0.74)	0.00
Layout*Duration	**7.53 (0.01)**	0.01	0.09 (0.76)	0.00	0.44 (0.51)	0.00	1.50 (0.22)	0.00	0.19 (0.66)	0.00

Bold entries are statistically significant at $p < 0.05$.

Table 13 Summary of results

Factor	Comprehension η_p^2 Range	Completion Time η_p^2 Range
Layout	0.01 - 0.02	0.01 - 0.06
Position	not significant	0.01 - 0.03
Familiarity	0.03 - 0.06	mainly not significant
Intensity	0.00 - 0.02	mainly not significant
Education	not significant	mainly not significant
Knowledge	0.06 - 0.12	mainly not significant
Duration	not significant	mainly not significant

5.3 Threats to Validity

Several threats to validity exist in our study. First, our analysis was based on a study with process modeling students and practitioners. While process modeling is one key approach in conceptual modeling, the external validity of our implications regarding the use of data or object-oriented models requires further investigation.

Second, our examination of MVCs was based on dimensions and measures found in the literature. While this gives us the ability to relate our findings to the work to date, there is a need for research on dedicated construct development to identify alternative or more suitable MVCs. Several recommendations exist in the literature, e.g., Lewis et al. (2005), which can guide such work. Additionally, we note that many of the measures we found relied on self-reporting and that other, more objective metrics could inform different results. One example of such a measure worth revisiting is related to our factor *Education*, viz., self-reported training and self-reported learning-by-doing, both of which may be open to interpretation bias.

Third, we note that our experimental results may be biased by the chosen design for our experimental tasks. We chose our tasks because they had been shown to be highly valid to study model comprehension and are also well-utilized in the current literature. Our tasks belong to a class of schema-based problem solving tasks (Gemino and Wand 2003), which are different from inferential problem solving tasks (Bodart et al. 2001; Recker and Dreiling 2011; Gemino and Wand 2005). In turn, the external validity of our findings might be bounded to comprehension and schema-based problem solving on the basis of conceptual models.

Fourth, while we took care in eliminating confounding factors in our experiment design we still note that the chosen grammar, BPMN, as well as the chosen notational elements (Figl et al. 2013a) and complexity levels of the models (Recker 2013) are impacting how well end users comprehend the model, in turn defining a boundary of internal validity to our work. We selected a traditional left-to-right direction of our process models. However, this may have been more cumbersome to read for some participants than expected. Interestingly, recent experimental work (Figl and Strembeck 2015) suggests that flow direction is generally not a substantial influence on model comprehension. So, bias, if any, all should be marginal.

Fifth, we utilized an interactive website in order to motivate people to participate. It is known that people might not be equally effective working with a website as compared to standard-alone tools or printouts of models (Polančič et al. 2015). More specifically, we addressed potential learning effects of repeated participation by a respective filter strategy. We addressed adverse self-selection by emphasizing the interactive feedback and the value proposition of the website to work as a learning resource.

Sixth, we note that data collection occurred mainly in 2010, so one might conceive the data and thus the results to be dated. However, the adoption of BPMN since 2010 has further increased since then (Recker 2010b; Chinosi and Trombetta 2012), reinforcing the relevance of our findings. Also, conceptual modeling studies continue to be published in top journals to this day but, as summarized in Table 1, even recent work has not yet devoted substantial attention to MVCs. Finally, even modern textbooks on process modeling education (Dumas et al. 2013) remain similar in their treatment of main concepts as education guidelines available at the time of our data collection (Recker and Rosemann 2009). The results, in sum, are timely and relevant.

6 Conclusions

We designed an experiment as an empirical review of model viewer characteristics (MVCs) and their impact on comprehending conceptual process models. We did so by collecting MVC data, using measures reported in the literature, and using good and bad layout as a treatment in an experiment with 333 students and 197 practitioners. We recorded significant differences in comprehension performance that we could link back to differences in MVCs that characterize user differences.

Our results affirm our contention that experiments in conceptual modeling literature would benefit from a more developed understanding of which MVCs need to be included in experimental designs. Our research is a first empirical exploration of this area and we hope future studies will further extend these ideas. Two avenues are particularly important in our view. First, the development of a more sound theoretical basis to conceptualize MVCs and ideally also other elements of conceptual modeling pragmatics -

the study of the contexts in which conceptual models are used. Second, the execution of more rigorous and systematic measurement development work to operationalize MVC as the multi-dimensional construct we found it to be. Third, prior studies should be replicated with MVCs being explicitly integrated into the data analysis. We hope that other colleagues will join us in these endeavors.

Appendix: Study Materials

Measurements Related to Expertise

`ModelingYears:` How many years ago did you start process modeling? (integer)

`ModelsRead:` How many process models have you analyzed or read within the last 12 months? (integer)

`ModelsCreated:` How many process models have you created or edited within the last 12 months? (integer)

`Training:` How many work days of formal training on process modeling have you received within the last 12 months? (integer)

`SelfEduction:` How many work days of self education have you made within the last 12 months? (integer)

`FAM1:` Overall, I am very familiar with the BPMN. (1-7)

`FAM2:` I feel very confident in understanding process models created with the BPMN. (1-7)

`FAM3:` I feel very competent in using the BPMN for process modeling. (1-7)

`MonthsBPMN:` How many months ago did you start using BPMN? (integer)

Knowledge Questions for Variable `Theory`

– For exclusive choices, exactly one of the alternative branches is activated (yes/no).
– Exclusive choices can be used to model a repetition (yes/no).
– In BPMN, synchronization is modeled by an AND-join (yes/no).
– If two activities are concurrent, then they must be executed at the same time (yes/no).
– If an activity is modeled to be part of a loop, then it has to be executed at least once (yes/no).
– For correctly joining multiple paths coming from the same OR split, you can use either XOR or AND gateways (yes/no).
– An OR gateway activates either one or all outgoing paths (yes/no).

Comprehension Questions on Model 1

1. If L is executed for a case, then H might have been executed for the same case.
2. Z and AA are exclusive to each other.
3. If A is executed for a case, then O and BB must also be executed for this case.
4. Both A and FF can be executed more than once.
5. After O has been executed, and the default path is taken at the next gateway, then Z must always be executed.
6. If X has been executed for a case, then it is not possible to execute N.

Comprehension Questions on Model 2

1. S, T, and U can be executed within one case.
2. If Z is executed, then C and J must have been executed before.
3. If X is executed for a case, then both BB and V can be executed for the same case.
4. If Y has been executed for a case, then at least W or AA are executed for this case, too.

Fig. 4 Model 1 of the BPMN-Selftest (good layout)

Fig. 5 Model 1 of the
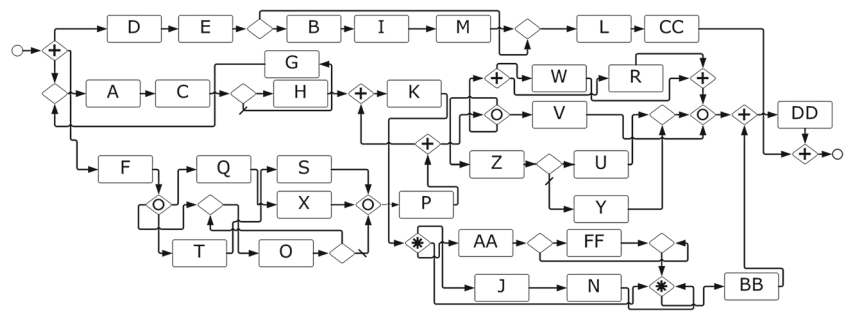BPMN-Selftest (bad layout)

Fig. 6 Model 2 of the
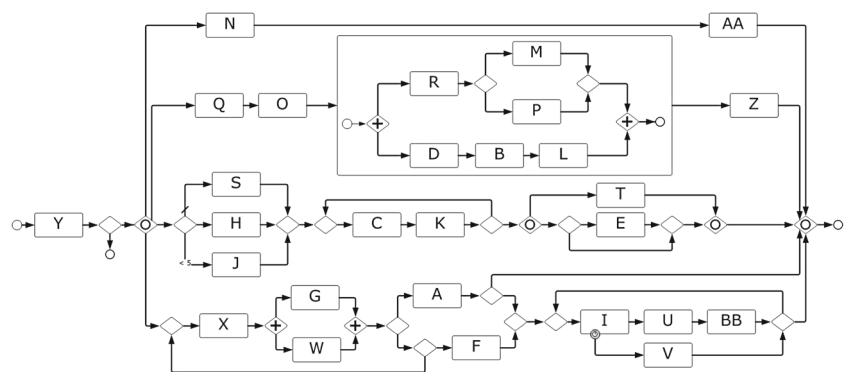BPMN-Selftest (good layout)

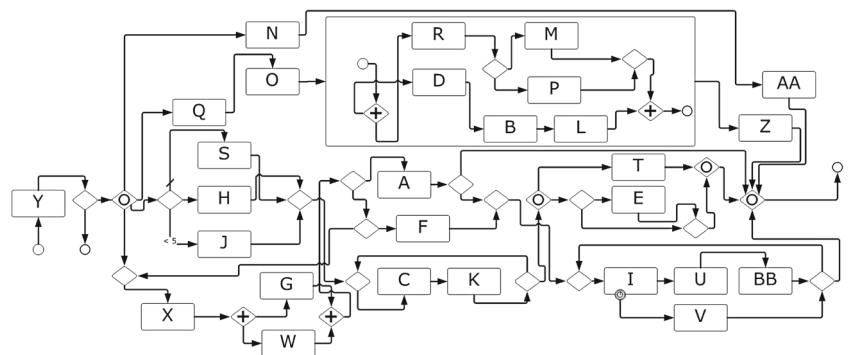
Fig. 7 Model 2 of the
BPMN-Selftest (bad layout)

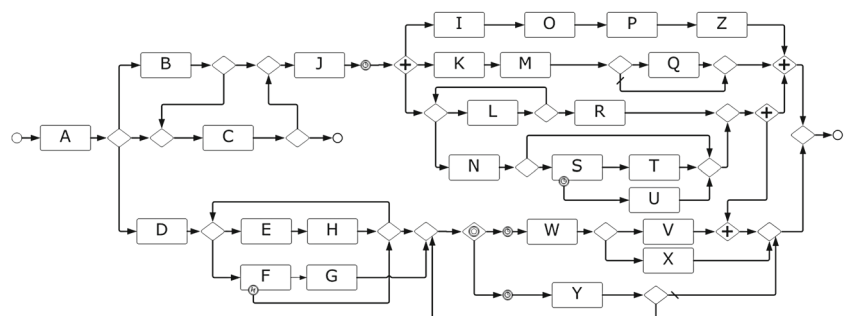**Fig. 8** Model 3 of the
BPMN-Selftest (good layout)

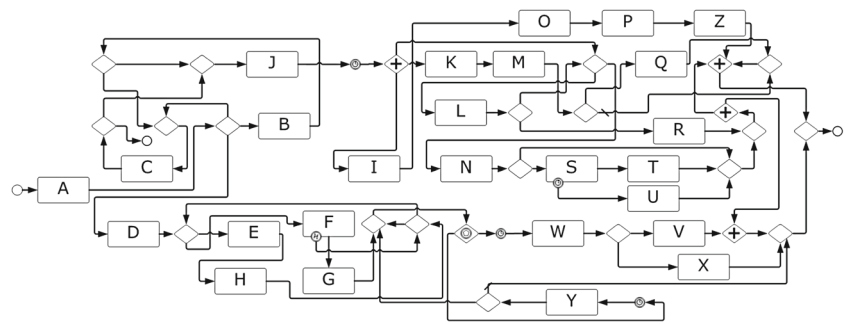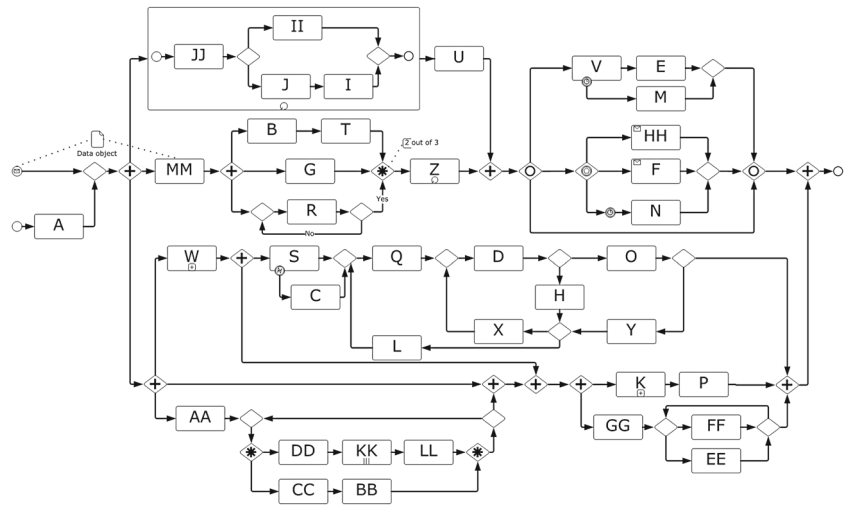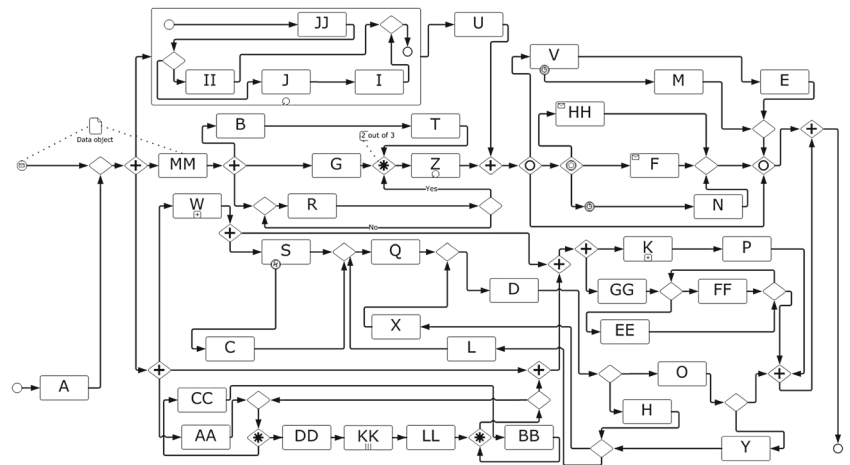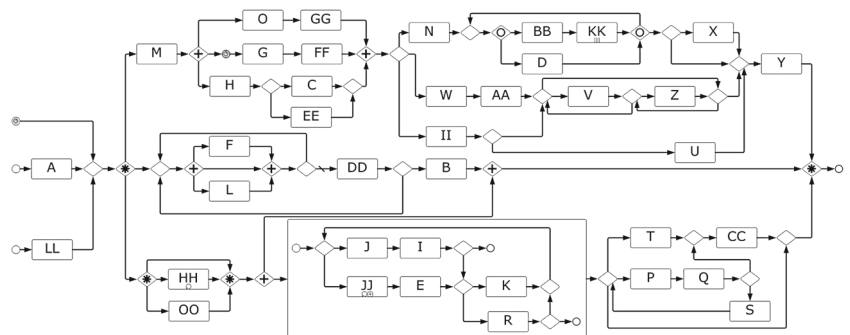**Fig. 9** Model 3 of the
BPMN-Selftest (bad layout)

Fig. 10 Model 4 of the
BPMN-Selftest (good layout)

Fig. 11 Model 4 of the
BPMN-Selftest (bad layout)

Fig. 12 Model 5 of the
BPMN-Selftest (good layout)

Fig. 13 Model 5 of the
BPMN-Selftest (bad layout)
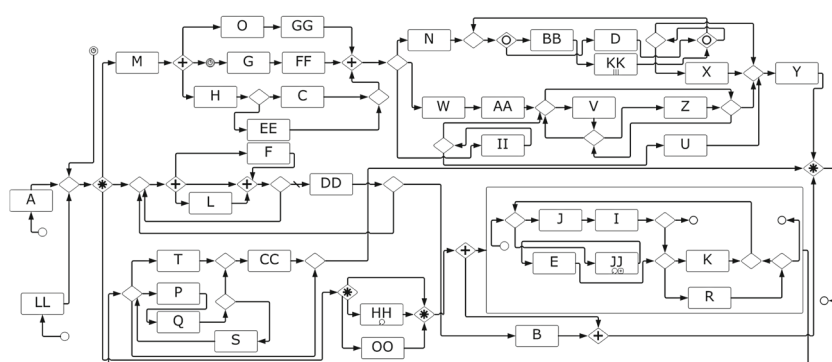
5. If Q is executed for a case, then P and X are executed as well for this case.
6. If J and O are executed, then L must be executed for that case too.

Comprehension Questions on Model 3

1. Once V is executed for a case, Y can no longer be executed for that case.
2. If an error occurs at F, then G can no longer be executed.
3. After M is executed, the next gateway along the path activates its default path. Then it is no more possible to execute T.
4. V can be the last activity before the process terminates.
5. If I has been executed for a case, B must have been executed before.
6. C, E, and G can be executed several times for a case.

Comprehension Questions on Model 4

1. I can be executed several times for a case.
2. E, F, and H can be executed in parallel for a case.
3. If M has been executed for a case, it is possible to execute K for that case.
4. L, P, and KK are executed at most once for all cases.
5. If DD is executed for a case, then Y can be executed for the same case.
6. For any case, A, O, and MM must be executed at least once.

Comprehension Questions on Model 5

1. If HH is executed, then K must be executed for the same case.
2. B and KK can run in parallel.
3. If H is executed, then Y and Z must also be executed for that case.

4. If X is executed for a case, then EE must always be executed before.
5. After O has been executed, JJ can be executed several times.
6. E, S, and U can be executed within one case.

References

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

Allen, G., & Parsons, J. (2010). Is query reuse potentially harmful? anchoring and adjustment in adapting existing database queries. *Information Systems Research* **21**(1) 56–77.

Batra, D., Hoffler, J.A., Bostrom, R.P. (1990). Comparing representations with relational and eer models. *Communications of the ACM* **33**(2) 126–139.

Becker, J., Delfmann, P., Dietrich, H.-A., Steinhorst, M., Eggert, M. (2016). Business process compliance checking – applying and evaluating a generic pattern matching approach for conceptual models in the financial sector. *Information Systems Frontiers* **18**(2) 359–405. https://doi.org/10.1007/s10796-014-9529-y.

Bera, P. (2012). Does cognitive overload matter in understanding bpmn models? *Journal of Computer Information Systems* **52**(4) 59–69.

Bera, P., Burton-Jones, A., Wand, Y. (2014). Research note-how semantics and pragmatics interact in understanding conceptual models. *Information Systems Research* **25**(2) 401–419.

Bodart, F., Patel, A., Sim, M., Weber, R. (2001). Should optional properties be used in conceptual modelling? a theory and three empirical tests. *Information Systems Research* **12**(4) 384–405.

Bowen, P.L., O'Farrell, R.A., Rohde, F. (2009). An empirical investigation of end-user query development: The effects of improved model expressiveness vs. complexity. *Information Systems Research* **20**(4) 565–584.

Burton-Jones, A., & Meso, P. (2008). 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) 784–802.

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

Chi, M.T.H., Glaser, R., Farr, M.J. (2014). *The nature of expertise*. Psychology Press.

Chinosi, M., & Trombetta, A. (2012). Bpmn: An introduction to the standard. *Computer Standards & Interfaces* **34**(1) 124–134.

Christophersen, T., & Konradt, U. (2011). Reliability, validity, and sensitivity of a single-item measure of online store usability. *International Journal of Human-Computer Studies* **69**(4) 269–280.

Claes, J., Vanderfeesten, I., Gailly, F., Grefen, P., Poels, G. (2015). The structured process modeling theory (spmt) a cognitive view on why and how modelers benefit from structuring the process of process modeling. *Information Systems Frontiers* **17**(6) 1401–1425.

Cognini, R., Corradini, F., Gnesi, S., Polini, A., Re, B. (2016). Business process flexibility - a systematic literature review with a software systems perspective. *Information Systems Frontiers* https://doi.org/10.1007/s10796-016-9678-2.

Compeau, D., Marcolin, B., Kelley, H., Higgins, C. (2012). Generalizability of information systems research using student subjects – a reflection on our practices and recommendations for future research. *Information Systems Research* **23**(4) 1093–1109.

Cox, J.C., Sadiraj, V., Schmidt, U. (2014). Paradoxes and mechanisms for choice under risk. *Experimental Economics* **18**(2) 215–250.

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

Davis, C.J., Hevner, A.R., Weber, B. (2017). Studying the creation of design artifacts. *Information Systems and Neuroscience*. Springer, 115–122.

Dehnert, J., & van der Aalst. W.M.P. (2004). Bridging The Gap Between Business Models And Workflow Specifications. *International J. Cooperative Inf. Syst.* **13**(3) 289–332.

Dijkman, R.M., Dumas, M., Ouyang, C. (2008). Semantics and analysis of business process models in bpmn. *Information and Software Technology* **50**(12) 1281–1294.

Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A. (2013). *Fundamentals of Business Process Management*. Springer.

Ericsson, K.A., Prietula, M.J., Cokely, E.T. (2007). The making of an expert. *Harvard business review* **85**(7/8) 114.

Evermann, J., & Wand, Y. (2005). Toward formalizing domain modeling semantics in language syntax. *IEEE Transactions on Software Engineering* **31**(1) 21–37.

Fettke, P. (2009). How conceptual modeling is used. *Communications of the Association for Information Systems* **25**(43) 571–592.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Figl, K. (2017). Comprehension of procedural visual business process models - A literature review. *Business & Information Systems Engineering* **59**(1) 41–67. https://doi.org/10.1007/s12599-016-0460-2.

Figl, K., Mendling, J., Strembeck, M. (2013a). The influence of notational deficiencies on process model comprehension. *Journal of the Association for Information Systems* **14**(6) 312–338.

Figl, K., Recker, J., Mendling, J. (2013b). A study on the effects of routing symbol design on process model comprehension. *Decision Support Systems* **54**(2) 1104–1118.

Figl, K., & Strembeck, M. (2015). Findings from an experiment on flow direction of business process models.

Gemino, A., & Wand, Y. (2003). Evaluating modeling techniques based on models of learning. *Commun. ACM* **46**(10) 79–84.

Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering* **9**(4) 248–260.

Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering* **55**(3) 301–326.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. (2010). *Multivariate Data Analysis*. 7th ed. Prentice Hall, Upper Saddle River, New Jersey.

Houy, C., Fettke, P., Loos, P. (2012). Understanding understandability of conceptual models - what are we actually talking about? Paolo Atzeni, David W. Cheung, Sudha Ram, eds., *Conceptual Modeling - 31st International Conference ER 2012, Florence, Italy, October 15-18, 2012. Proceedings*, *Lecture Notes in Computer Science*, vol. 7532. Springer, 64–77. https://doi.org/10.1007/978-3-642-34002-4.

Indulska, M., Recker, J., Rosemann, M., Green, P. (2009). *Process Modeling: Current Issues and Future Challenges*, *Lecture Notes in Computer Science*, vol. 5565. Springer, Amsterdam, The Netherlands, 501–514.

Jacoby, J., Troutman, T., Kuss, A., Mazursky, D. (1986). Experience and expertise in complex decision making. *Advances in consumer research* **13**(1).

Khatri, V., & Vessey, I. (2016). Understanding the role of is and application domain knowledge on conceptual schema problem solving: A verbal protocol study. *Journal of the Association for Information Systems* **17**(12) 759–803.

Khatri, V., Vessey, I., Ramesh, V., Clay, P., Sung-Jin, P. (2006). Understanding conceptual schemas: Exploring the role of application and is domain knowledge. *Information Systems Research* **17**(1) 81–99.

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

Kindler, E. (2006). On the semantics of EPCs: Resolving the vicious circle. *Data & Knowledge Engineering* **56**(1) 23–40.

Kirk, R.E. (1996). Practical significance: A concept whose time has come. *Educational and psychological measurement* **56**(5) 746–759.

Kock, N., Gray, P., Hoving, R., Klein, H.K., Myers, M.D., Rockart, J.F. (2002). Is research relevance revisited: Subtle accomplishment, unfulfilled promise, or serial hypocrisy? *Communications of the Association for Information Systems* **8**(23) 330–346.

Kummer, T.-F., Recker, J., Mendling, J. (2016). Enhancing understandability of process models through cultural-dependent color adjustments. *Decision Support Systems* **87** 1–12.

Lewis, B.R., Templeton, G.F., Byrd, T.A. (2005). A methodology for construct development in mis research. *European Journal of Information Systems* **14**(4) 388–400.

Lukyanenko, R., Parsons, J., Wiersma, Y.F. (2014). The impact of conceptual modeling on dataset completeness: A field experiment. *35th International Conference on Information Systems*. Association for Information Systems.

Mayer, R.E. (2009). *Multimedia learning*. Cambridge university press.

Mendling, J. (2008). *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, *Lecture Notes in Business Information Processing*, vol. 6. Springer.

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

Mendling, J., Strembeck, M., Recker, J. (2012). Factors of process model comprehension findings from a series of experiments. *Decision Support Systems* **53**(1) 195–206.

Miller, D. (2007). Paradigm prison, or in praise of atheoretic research. *Strategic Organization* **5**(2) 177–184.

OMG (2010). Business process model and notation (bpmn) - version 2.0.

Parsons, J. (2011). An experimental study of the effects of representing property precedence on the comprehension of conceptual schemas. *Journal of the Association for Information Systems* **12**(6) 401–422.

Permvattana, R., Armstrong, H., Murray, I. (2013). E-learning for the vision impaired: A holistic perspective. *International Journal of Cyber Society and Education* **6**(1) 15–30.

Petre, M. (1995). Why looking isn't always seeing: Readership skills and graphical programming. *Commun. ACM* **38**(6) 33–44. https://doi.org/10.1145/203241.203251.

Petre, M. (2006). Cognitive dimensions 'beyond the notation'. *Journal of Visual Languages & Computing* **17**(4) 292–301. https://doi.org/10.1016/j.jvlc.2006.04.003. Ten Years of Cognitive Dimensions Ten Years of Cognitive Dimensions.

Petrusel, R., Mendling, J., Reijers, H.A. (2016). Task-specific visual cues for improving process model understanding. *Information & Software Technology* **79** 63–78. https://doi.org/10.1016/j.infsof.2016.07.003.

Petrusel, R., Mendling, J., Reijers, H.A. (2017). How visual cognition influences process model comprehension. *Decision Support Systems* **96** 1–16. https://doi.org/10.1016/j.dss.2017.01.005.

Polančič, G., Jošt, G., Heričko, M. (2015). An experimental investigation comparing individual and collaborative work productivity when using desktop and cloud modeling tools. *Empirical Software Engineering* **20**(1) 142–175.

Porter, S.R., & Whitcomb, M.E. (2003). The impact of contact type on web survey response rates. *The Public Opinion Quarterly* **67**(4) 579–588.

Purchase, H. (1997). Which aesthetic has the greatest effect on human understanding? *Graph Drawing*. Springer, 248–261.

Purchase, H.C. (2014). Twelve years of diagrams research. *J. Vis. Lang. Comput.* **25**(2) 57–75. https://doi.org/10.1016/j.jvlc.2013.11.004.

Purchase, H.C., Carrington, D., Allder, J.-A. (2002). Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering* **7**(3) 233–255.

Purchase, H.C., Cohen, R.F., James, M.I. (1997). An experimental study of the basis for graph drawing algorithms. *ACM Journal of Experimental Algorithmics* **2** 4. https://doi.org/10.1145/264216.264222.

Purchase, H.C., McGill, M., Colpoys, L., Carrington, D. (2001). Graph drawing aesthetics and the comprehension of uml class diagrams: an empirical study. *Proceedings of the 2001 Asia-Pacific symposium on Information visualisation-Volume 9*. Australian Computer Society, Inc., 129–137.

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

Recker, J. (2010b). Opportunities and constraints: The current struggle with bpmn. *Business Process Management Journal* **16**(1) 181–201.

Recker, J. (2013). Empirical investigation of the usefulness of gateway constructs in process models. *European Journal of Information Systems* **22**(6) 673–689.

Recker, J., & Dreiling, A. (2011). 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) 65–84.

Recker, J., Reijers, H.A., van de Wouw, S.G. (2014). Process model comprehension: The effects of cognitive abilities, learning style, and strategy. *Communications of the Association for Information Systems* **34**(9) 199–222.

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

Recker, J., Rosemann, M., Green, P., Indulska, M. (2011). Do ontological deficiencies in modeling grammars matter? *MIS Quarterly* **35**(1) 57–79.

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

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

Reijers, H.A., Mendling, J., Dijkman, R.M. (2011b). Human and automatic modularizations of process models to enhance their comprehension. *Information Systems* **36**(5) 881–897.

Reuber, R. (1997). Management experience and management expertise. *Decision Support Systems* **21**(2) 51–60.

Sarshar, K., & Loos, P. (2005). Comparing the control-flow of epc and petri net from the end-user perspective. W.M.P. van der 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, 434–439.

Schrepfer, M., Wolf, J., Mendling, J., Reijers, H.A. (2009). The impact of secondary notation on process model understanding. *The Practice of Enterprise Modeling*. Springer, 161–175.

Shanks, G., Moody, D.L., Nuredini, J., Tobin, D., Weber, R. (2010). Representing classes of things and properties in general in conceptual modelling: An empirical evaluation. *Journal of Database Management* **21**(2) 1–25.

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., Weber, R. (2008). Representing partwhole relations in conceptual modeling: An empirical evaluation. *MIS Quarterly* **32**(3) 553–573.

Sidorova, A., Evangelopoulos, N., Valacich, J.S., Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly* **32**(3) 467–482.

Spence, M.T., & Brucks, M. (1997). The moderating effects of problem characteristics on experts' and novices' judgments. *Journal of marketing Research* 233–247.

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

Topi, H., & Ramesh, V. (2002). Human factors research on data modeling: a review of prior research, an extended framework and future research directions. *Journal of Database Management (JDM)* **13**(2) 3–19.

Turetken, O., & Schuff, D. (2007). The impact of context-aware fisheye models on understanding business processes: An empirical study of data flow diagrams. *Information & management* **44**(1) 40–52.

Venable, J.R. (2007). Relevance vs. rigour or relevance and rigour? contingence and invariance in standards for is research. *Wirtschaftsinformatik* **49**(5) 407–409.

Wand, Y., & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda. *Information Systems Research* **13**(4) 363–376.

Ware, C., Purchase, H.C., Colpoys, L., McGill, M. (2002). Cognitive measurements of graph aesthetics. *Information Visualization* **1**(2) 103–110. https://doi.org/10.1057/palgrave.ivs.9500013.

Weber, R. (1997). *Ontological Foundations of Information Systems*. Coopers & Lybrand and the Accounting Association of Australia and New Zealand, Melbourne, Australia.

Weber, R. (2006). *Like Ships Passing in the Night: The Debate on the Core of the Information Systems Discipline*. John Wiley & Sons, Chichester, England, 292–299.

Weidlich, M., Mendling, J., Weske, M. (2011). Efficient consistency measurement based on behavioral profiles of process models. *IEEE Trans. Software Eng.* **37**(3) 410–429.

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

Xiao, L., & Zheng, L. (2012). Business process design: Process comparison and integration. *Information Systems Frontiers* **14**(2) 363–374. https://doi.org/10.1007/s10796-010-9251-3.

Jan Mendling is a full professor with the Institute for Information Business at Wirtschaftsuniversitt Wien (WU Vienna), Austria. His research interests include various topics in the area of business process management and information systems. He has published more than 300 research papers and articles. He is member of the board of the Austrian Society for Process Management, co-founder of the Berlin BPM Community of Practice, and member of the IEEE Task Force on Process Mining. His Ph.D. thesis has won the Heinz-Zemanek-Award of the Austrian Computer Society and the German Targion-Award for dissertations in the area of strategic information management.

Jan Recker is full professor for information systems and systems development in the Faculty of Management, Economics and Social Sciences at the University of Cologne and adjunct professor in the QUT Business School at Queensland University of Technology. His research focuses on systems analysis and design, digital innovation and environmental sustainability. He is presently editor-in-chief of the Communications of the Association for Information Systems.

Hajo A. Reijers is a full professor at the Vrije Universiteit Amsterdam, where he heads the Information Management & Software Engineering group of the Computer Science department. He is also a part-time, full professor at the Department of Mathematics & Computer Science of Eindhoven University of Technology. His expertise is in enterprise systems, business process management, process mining, conceptual modeling, and workflow technology. Reijers has published over 150 scientific papers, chapters in edited books, and articles in professional journals.

Henrik Leopold is an Assistant Professor at the Department of Computer Science at the Vrije Universiteit Amsterdam. His research interests include business process analysis using natural language processing and machine learning techniques, process model matching, and conformance checking. He has published over 50 research papers and articles, among others, in Data & Knowledge Engineering, Decision Support Systems, IEEE Transactions on Software Engineering, and Information Systems. His doctoral thesis received the German Targion Award 2014 for the best dissertation in the field of strategic information management.