Which concepts do inexperienced modelers use to model work? – An exploratory study

Stefan Oppl

Johannes Kepler University of Linz,
Department of Business Informatics – Communications Engineering, Linz, Austria
stefan.oppl@jku.at

Abstract. Modeling languages are generally designed to meet the requirements imposed on them by the domain to be represented and the intended application area of the resulting models. Their comprehensibility and appropriateness for domain experts, who are not necessarily experienced modelers, usually is not considered during language design. While some aspects such as notational properties of languages and their understandability have been subject to extensive research in this area, the semantic appropriateness of language constructs has hardly been a subject of research so far. This paper presents a set of exploratory studies that examine the use of different conceptual modeling languages by inexperienced modelers. The results indicate that the set of available modeling language concepts seems to influence the focus of articulation for inexperienced modelers. Mismatches between modeling language concepts and modelers' conceptualizations of the modeling domain appear to be compensated by ad-hoc re-definitions of language element semantics. Based on these findings we identify implications that need to be considered when designing modeling languages suitable for articulation by inexperienced modelers.

Keywords: Novice Modelers, Modeling Language Semantics, Stakeholder Articulation

1 Introduction

The last years have seen numerous research efforts on involving domain experts in modeling activities on an enterprise level (e.g., [1-3]). Traditionally, domain experts are considered a source of information for process analysts, who are responsible for creating models appropriate for the respective aim of modeling [4]. Active involvement in the modeling process is largely constrained to validation activities, during which the appropriateness of the information represented in the models is checked, e.g., via model walkthroughs [5] or simulations [6]. Research consequently has focused on examining understandability of models [7, 8], which is a prerequisite to enable domain experts to assess whether the information represented in models is correct. Domain experts do not necessarily have modeling skills or experiences. Their involvement in model creation thus has not been addressed to such a large extent. Most existing research either tries to avoid involving domain experts directly in conceptual modeling activities [9] or assumes that they acquire the necessary modeling skills over time [10]. Requirements on modeling languages that can be used by domain experts inexperienced in modeling have hardly been examined.

Modeling by domain experts is referred to by Krogstie et. al [11] as *manual articulation*, during which domain experts externalize their knowledge about the domain into a model. The modeling language here plays a crucial role and needs to be "comprehensible" and "appropriate […] with respect to the domain and participants' knowledge" [11, p. 95]. Whether or not a modeling language meets these requirements consequently cannot be assessed objectively, but is depended on the individual domain experts' conceptualization of how they describe the aspects of the real world that should be represented in the model [12]. *Comprehensibility* is enabled when the semantics of modeling language are described so that participants involved in modeling understand how the language concepts should be used to describe their perceptions of the modeling topic. *Appropriateness* is only achieved if language concepts are comprehensible for a modeling participant, and their semantics match those the participant expects to be available. Mismatches between modeling language semantics and the concepts used by participants to refer to the modeling topic require acts of translation [13]. Such required translations distract from the subject of modeling and interrupt the articulation process, potentially leading to reduced model quality.

To avoid such interruptions, modeling languages designed to be used by inexperienced modelers should allow them to express their expected semantics. How this requirement can be accounted for, however, has hardly been examined empirically in existing research. Zarwin et al. [12, p. 4] state, that "users invent their own notations, decide to break the rules, introduce new conventions, converge to new modes of modeling, and all this on the fly." The present paper sets out to examine how people with no or limited modeling experiences are influenced by modeling language semantics with respect to which concepts they choose to incorporate in their models. The aim is to identify issues that need to be considered when designing modeling languages that should be used by inexperienced modelers to articulate their knowledge in models.

The remainder of this article is organized as follows: in section 2, we briefly review related work in modeling by inexperienced modelers. Section 3 introduces our research setup and describes the deployed methodology. Section 4 summarizes the results of the study we have conducted in the present research. Section 5 discusses the results in the light of the aims of this article and gives an initial account on issues to be considered during language design or selection.

2 Related Work

The use of modeling languages by inexperienced modelers has been a subject of research in several studies over the last years. Recker et al. [14] have examined how inexperienced modelers design business processes. They have identified five different process design archetypes, which differ in their degree of formalization (abstract vs. concrete) and their form of presentation (text vs. graphics vs. mixed). For their study, they have analyzed models concerning their syntactic, semantic and empirical quality based on existing theoretical frameworks. They found that different aims of modeling are approached via different design archetypes and that graphical representation, in general, seems to be preferable over textual descriptions in most cases. While the aims of the present study are similar to those described by Recker et al. [14], our approach differs in that we explicitly examine how semantic constraints imposed by different modeling languages influence inexperienced modelers when describing their perception of work. Our focus is on the needs of modelers rather than on the semantic quality of models measured against a theoretically backed baseline. This requires a different methodological approach to model analysis that allows extracting the latent semantics represented in a model.

Starting from a similar problem diagnosis as described in section 1, Monsalve et al. [15] set out to specify a method that allows accommodating different stakeholder perspectives in business process modeling. The method achieves this by performing an ex-ante ontological analysis to identify those aspects of a business process a particular group of stakeholders needs to represent their perceptions appropriately. While pursuing a similar aim as this paper, the approach taken is fundamentally different: Their analysis is based on a set of theoretically derived approaches and industry standards and does not directly involve the affected group of stakeholders. The study presented by Patig & Casanova-Brito [16] uses an empirical approach to collect requirements on process modeling languages. They adopt a perspective of industry needs and collect their data from company representatives mainly active in IT and BPM. Such people can be assumed to be more experienced modelers, the study thus present requirements derived from a different target group than addressed here.

The visual notations of modeling languages and their appropriateness for inexperienced modelers have been examined in several works. Genon et al. [17] discuss the notation of BPMN in the light of the Physics of Notation theory [18]. Figl [7] presents a meta-study of aspects impacting the comprehensibility of business process models. Türetken [8] et al. discuss the impact of cognitive style on the understanding of business process models. They all share a focus on people using process models rather than actively designing them. Their findings also indicate that modeling language design and individual characteristics of modelers (or model users) are factors that influence how appropriate models perceived to be for a particular usage scenario. Such influences are also diagnosed by Zur Muehlen & Recker [19], who have examined BPMN models created and deployed in industry settings. They found that modelers usually do not use the whole set of language concepts, but rather focus on subsets depending on modelers’ preferences and the aim of modeling.

3 Methodology

Our research question is to examine the effects of prescribed modeling language semantics on the actual meaning of elements used in the articulation results (i.e., models) created by inexperienced modelers. Based on the findings from related work, our hypothesis is that the degree a chosen modeling language constrains what can be expressed semantically (independent variable) has an impact on the concepts inexperienced modelers use describe their knowledge about the domain (dependent variable). We do not assess the content of the created models themselves, but only examine the semantics of the used language elements.

The data necessary to evaluate this question was generated in real-world modeling workshops to generate models that are based on modeler's actual knowledge about the model subject. The modeling tasks consequently differed in their aims and context. While generating an artificial experimental setup with better controllable conditions and a consistent modeling scenario would be preferable to raise internal validity, modelers in such a case could not rely on real-world knowledge on what to represent in their models. Forcing modelers to explicitly think about what to represent in the models might already be constrained by the semantics of the language to be used for modeling. This would potentially introduce a selection bias in terms of what to represent in the model already before the act of modeling itself and thus threaten external validity. We are aware of the limitations of our chosen approach with respect to the potential impact of the different modeling scenarios, but still have chosen to work with real world cases to avoid this confounding factor. We mitigated the scenario selection problem by selecting workshops that consistently were concerned with representing work processes, thus providing a common scope in the universe of discourse [4].

We operationalize the independent variable by varying the modeling languages used in the examined modeling workshops. The modeling languages were selected to represent different degrees of semantic constraints imposed during modeling (for details, see description of the series below). Overall, we conducted three series of workshops, each using a particular modeling language. All three languages were suited for representing work processes (either being explicitly designed for this purpose or leaving the freedom to use according semantics). The participants in the workshops did not have any prior modeling experiences.

The dependent variable is operationalized by analyzing the concepts modelers use when creating their models. We expect the frequency of concept occurrence in the models to be influenced by the used modeling language and its semantic constraints. We hypothesize that languages with less semantic constraints lead to models that more exactly match what the modelers consider relevant to be represented in a model. Our hypothesis would be backed if we could observe the use of different concept types across the workshop series.

In *series 1*, modelers had diverse professional backgrounds. They were asked to describe re-occurring work processes they experience in their daily professional or private lives. They were given no pre-specified semantics to adhere to during modeling but were asked to use equally colored modeling elements for concepts of the same kind. For each used color, they were asked ex-post to describe the concept the color referred to. No further constraints on how to represent work processes or which concepts to use were given. Overall, 26 modelers each created a model, which overall contained 416 elements (min = 6, max = 104, mean = 16.0, std.-dev. = 19.4).

In *series 2*, the modelers were healthcare professionals and industrial production workers participating in vocational training activities. They were asked to describe work processes that they perceived to be problematic in their daily work. They were provided with three different types of modeling elements that were inspired by modern process modeling languages such as BPMN and vaguely defined as representing "What" (i.e., what needs to be done), "Who" (i.e., who is doing the "What") and "Exchange" (i.e., what is required from others or can be provided to others when doing the "What"). Details on the deployed modeling approach can be found in [20]. The exact semantics of the modeling elements were deliberately left vague and provided room for choosing one's preferred level of granularity and meaning of the elements (e.g., describing "Exchange" on the level of messages, documents, materials, etc., or describing "Who" by referring to actual people or abstract organizational roles). Overall, 82 modelers each created a model, which overall contained 1221 elements (min = 3, max = 56, mean = 14.8, std.-dev. = 12.3).

In *series 3*, the modelers were students of a bachelor's curriculum in business information systems in their first year. Neither of them had any background in process modeling. They were asked to use BPMN (Business Process Modeling and Notation) [21] to describe an event agency's process for organizing a customer event. The students did not have prior own knowledge about the process but were provided with information on this process as transcribed interviews, which were deliberately not formulated in a way that would make the mapping to modeling elements straightforward. After familiarizing themselves with the scenario, they were given a 1-hour theory-driven introduction to BPMN and its semantics before starting their task. For modeling, they were provided with a reference chart. Overall, 30 modelers each created a model, which overall contained 3296 elements (min = 32, max = 177, mean = 109.9, std.-dev. = 37.3).

The collected models were analyzed using a variant of thematic content analysis [22] to identify the concepts used for modeling (dependent variable). Thematic analysis is a method for identifying and analyzing patterns (themes) within data. As we were interested in the emerging semantics of the used modeling element types, we used an *inductive* coding approach to identify *latent* themes. Inductive coding here refers to the fact that the set of themes is not derived a priori from a guiding theory, but emerges "bottom-up" directly from the data based on the research question. Thematic analysis on latent level goes beyond the explicit semantics of the data (i.e., the semantics of the model elements as specified in the used modeling language), and examines the "underlying ideas, assumptions, and conceptualizations" [22]. Latent analysis is necessary in our case, as our study requires to examine the actual semantics of elements used in the articulation results.

Interpretation of the models is a subjective process that is prone to coder's bias, in particular when trying to identify latent themes. This was accounted for by having two independent coders assess all models. In the case of divergent codings, the author evaluated the respective element and decided which of the codings was to be used.

Using an initial sample of 10% the of models from all three series (selected from those models sized close to the median of each series), we deductively created a set of initial codes that allowed to classify the model elements used by the inexperienced modelers semantically. Following the procedure proposed by Braun & Clarke [22], the codes were not named a priori, but defined using a qualitative description and examples of the elements that belong to the theme designated by the respective code. When elements were encountered that could not be interpreted using the set of available codes, it was extended accordingly. In case the coders were not able to conclusively assign an element to a theme, the affected theme descriptions were refined.

Based on this set of codes, the full data set from all three series were coded independently by each of the two coders. Consistent coding was assessed using Cohen‘s Kappa as a metric, as suggested by Trickett & Trafton [23]. Overall, Cohen‘s Kappa was 0.943, which indicates a set of clearly defined and delimited themes and a good understanding of these themes by the coders. The set of codes thus was not modified any further. In a final step, the codes were named to designate the different themes and be able to refer to them in the following description of the results. In the following section, we summarize the results of our analysis and give an initial account of patterns that emerge from the available data.

4 Results

The results of the thematic analysis of the data collected in the three series of modeling workshops are presented in this section. We first describe the results of the inductive development of a set of codes used for describing the latent themes of the elements contained in the collected models. We then progress with an overview of the coding results, showing the distribution of identified themes for each of the examined series.

4.1 Concepts used for modeling

As described in the methodology section, we have derived the set of codes to designate latent themes used for modeling elements from a sample of 10% of the available models. These codes were validated in a double-coding process using the whole body of data. The set of codes was not altered in the validation process. After the coding process, the identified themes were revisited, named and described. Also, they were grouped into thematic clusters, which refer to the aspects of organizational work represented by elements assigned to the contained themes. The descriptions of the themes in Table 1 are grouped in these clusters:

Themes referring to WHAT-aspects describe those aspects of the work that can be executed by somebody and directly or indirectly contribute to achieving the aim of the work represented in the model (indirect contributions, e.g., would be coordinative activities [24]). **Themes referring to WHO-aspects** describe aspects of who is responsible for, or involved in, (parts of) the described work. **Themes referring to WITH WHAT-aspects** refer to those aspects that are required for, produced by, or exchanged in the course of carrying out (parts of) the described work. This also includes tools required for carrying out (parts of) the described work. **Themes referring to WHEN-aspects** refer to those aspects that describe under which conditions (parts of) the work can be executed. Such aspects are not constrained to temporal or causal aspects, but also include other types of constraints or conditions. The only aspect that could not be assigned to any of the clusters above are generic comments, which consequently were introduced as a separate theme. Table 1 shows the identified themes and the descriptions used during coding.

Table 1. Themes identified during latent coding

WHAT	
Activity	Elements describing aspects of the work that can be executed by any instance of the themes for WHO and directly or indirectly contribute to achieving the aim of the work.
Communication	Elements describing aspects of the work that are executed by at least two instances of the themes listed under WHO collaboratively and designate an exchange of information among them
WHO	
Individual Actor	Elements representing a specific person (i.e., somebody who can be distinctively identified and named ad personam).
Role	Elements representing abstract organizational positions or responsibilities, that can be taken by an individual actor.
Organization / Unit	Elements describing whole organisations or organizational units that do not give any explicit information on how many actors or roles are contained in them, but usually would contain more than one actor or role.
WITH WHAT	
Resource	Elements describing anything that is need to perform an instance of the themes listed under WHAT and is independent of the concrete work instance (i.e., its content or nature remains unaltered between work instances). This, e.g., includes tools, machinery, but also procedural guidelines.
Physical Object	Elements describing physical objects that are needed to perform an instance of the themes listed under WHAT and are consumed or altered in the process thereof. This, e.g., includes raw materials.
Document	Elements describing documents that are needed to perform an instance of the themes listed under WHAT and contain data specific for the particular work instance. This, e.g., includes filled paper forms.
Information	Elements describing information that is needed to perform an instance of the themes listed under WHAT. The way of delivery of this information is not specified (i.e., the information could be conveyed verbally as well as via paper).
Skill	Elements describing skills that are required from an actor / role / organization(al unit) to perform an instance of the themes listed under WHAT.
Communication Channel	Elements describing ways to exchange information between instances of the themes listed under WHO. This, e.g., includes email, phone calls, but also informal coffee talk.
WHEN	
Temporal Constraint	Elements describing temporal constraints for executing instances of the themes listed under WHAT. This, e.g., includes absolute and relative time constraints, such as events or causal relationships between WHAT-elements.
Spatial Constraint	Elements describing spatial constraints for executing instances of the themes listed under WHAT. This, e.g., includes locations or proximity to mobile instances of WHO.
Availability Constraint	Elements describing constraints for executing instances of the themes listed under WHAT that are dependent on the availability of particular WITH WHAT elements. This, e.g., includes elements representing waiting for incoming information, or the availability of a particular resource.
Alternative Activity flow	Elements representing choices on which of several potential instances of WHAT should be executed. This, e.g., includes exclusive or event-based gateways.
Parallel Activity flow	Elements representing that several potential instances of the themes listed under WHAT call be executed in parallel. This, e.g., includes parallel or inclusive gateways.
OTHER	
Comments	Elements containing unspecific comments, i.e., comments that cannot be classified under one of the themes described above.
Not coded	Elements that could not be coded for reasons independent of this coding scheme. This, e.g., includes elements with unreadable designators.

The set of themes gives an overall impression of which categories the inexperienced modelers used to describe the work in their models. It is important to note that the presence of certain themes and, in particular the absence thereof has potentially been influenced not only by the used modeling languages but also by the modeling tasks. Only the models of series 3 are based on an explicitly described scenario, which still was formulated in a way that did not impose any “obvious” modeling elements to be used. It can thus be assumed that the presence of certain themes is mainly dependent on the modelers’ way of describing their work in a conceptual model and the deployed modeling language. Overall, the themes are in line with what Fjuk el al. [25] in the context of CSCW refer to as the “salient dimensions of work — who, what, when, how” (“how” referring to aspects we have clustered under “with what”).

4.2 Coding Results

The set of codes described above was used to analyze the model elements collected in all three series of workshops. Table 2 shows the detailed numerical distribution of the observed codes. The data reveals several striking differences between models generated in the different series. We discuss them in the following paragraphs per series.

Table 2. Distribution of themes per series

	Theme	*Series 1*	*Series 2*	*Series 3*
WHO	*individual actor*	0,48%	2,21%	0,76%
	role	9,62%	11,96%	4,28%
	organization / unit	4,81%	6,47%	3,43%
WHAT	*activity*	12,26%	25,63%	21,00%
	communication	4,33%	9,42%	15,02%
WITH WHAT	*resource*	4,09%	0,08%	0,18%
	physical object	10,58%	0,00%	0,24%
	document	4,33%	1,31%	3,49%
	information	28,13%	23,67%	6,98%
	skill	10,34%	0,08%	0,00%
	communication channel	0,00%	1,15%	1,40%
WHEN	*spatial constraint*	0,48%	0,00%	0,09%
	availability constraint	0,00%	0,00%	0,00%
	temporal constraint	2,40%	9,50%	24,15%
	alternative activity flow	0,00%	0,00%	12,01%
	parallel activity flow	0,00%	0,00%	3,64%
OTHER	*comment*	0,00%	0,00%	1,24%
	not coded	8,17%	8,52%	2,09%

In *series 1*, where modelers could define their own semantic concepts when describing their work, 57.5 % of the elements represent aspects of the WITH WHAT cluster. This is a remarkable amount when compared to the other series, where the share of this cluster is around 25% in series 2 and only around 11% in series 3. Series 1 furthermore is the only series that contains elements categorized in the theme “skills” to a relevant amount (it is also present in series 2 with 1 element, which accounts for 0.08% of the elements). When specifying their own concepts, inexperienced modelers thus seem to strongly focus on the WITH WHAT aspect, where the largest shares of elements belong to the themes “information” (28.1%), “physical object” (10.6%), and “skill” (10.3%). In contrast, the themes belonging to the WHEN aspect were hardly used in series 1 and only account for 2.9% of the elements. This is remarkable, as this cluster contains all elements describing temporal and causal relationships. Consequently, the absence of elements in this cluster means that modelers have hardly adopted a flow-oriented perspective when describing their work processes.

While the modeling language deployed in *series 2* is only vaguely specified and would allow representing elements stemming from nearly all identified themes, it still poses some constraints on what can be represented in models. The language, e.g., does not provide constructs to represent communication channels or communication activities. Still, both themes have been used in the collected models to a limited extent (communication channels: 1.2%, communication: 9.4%). The prescribed semantics of the modeling language obviously has been re-defined by the modelers implicitly during modeling for nearly 10% of the used model elements to fit better the modelers' perceptions of how the represented work is organized. This redefinition has mainly happened for elements concerned with communication among people during the work process and might indicate too limited expressiveness of the used modeling language.

The models in *series 3* show a high number of elements belonging to the *WHEN*-cluster (39.9% overall). This is remarkable when compared to the other series, where the share of this cluster is around 23.8% in series 3, 9.5% in series 2 and only 2.9% in series 1. Looking at this cluster in more detail, 24.2% of the overall elements belong to the "temporal constraint" theme, which contains elements representing causal and temporal relationships. A total of 15.6% of the elements belongs to the themes concerned with representing alternatives and parallelization in the flow of activities. While this is not surprising – given the representational scope of BPMN – it still poses a major difference from the other series, where such themes were not present at all. This phenomenon might be attributable to the scenario description provided in series 3 as a basis for modeling, which contained indications that the process might exist in different variants. When not nudged towards such aspects, however, inexperienced modelers seem to focus on single cases and thus largely refrain from including alternatives or potentially parallel execution of activities. The low number of elements in the WITH WHAT-cluster reflects the semantic focus of BPMN. Some concepts that allow representing information handling or exchange are present in the models and account for the share of elements assigned to themes in this cluster.

5 Discussion

In this section, we first review the results presented above in the light of the research question formulated in section 3. We then address the preliminary implications for selecting or designing modeling languages for use by inexperienced modelers. Overall, the results of the analysis provide indications that the semantic constraints of modeling languages influence how inexperienced modelers select the aspect to be included in their models when describing their work processes. If semantics are left for the modelers to be specified (as in series 1), they seem not necessarily to organize their models along temporal or causal relationships, but also use other constraints to organize their models (such as assigning activities to a place or the availability of resources).

The analysis of the coding results has shown that some models from series 2 and 3 contain latent themes beyond the explicit semantics provided by the deployed modeling languages. This seems to indicate that inexperienced modelers re-define modeling element semantics on the fly to represent aspects of work they consider relevant, but are not covered by the modeling language. We hypothesize that these acts of redefinition hardly are deliberate decisions, but mostly are re-interpretations of modelers' understandings of the semantics of an existing language concept (i.e., "intuitively" using it). We do not follow up on this hypothesis here for reasons of space.

The results of our study on the latent themes used in modeling (cf. section 4.1) overall show that inexperienced modelers in general use concepts to describe their perception of work that do not fundamentally differ from the semantics offered by most conceptual modeling language designed for representing work processes [26, 27]. Some of the concepts observed for constraining the execution of activities (like spatial constraints), however, are hardly considered in common modeling languages (ibid.). This diagnosis is confirmed in ontological analyses [28] of process modeling languages, which found a lack of constructs being able to represent lawful states in process modeling languages – a category that is predominantly applicable to the "constraint"-themes identified during coding. More importantly, no single modeling language covers all aspects identified in coding [28]. A specific deployed language thus constraints which aspects of work can be represented at all, and might prompt modelers to think mainly about those aspects that are explicitly considered in the language.

The prompting effect of the deployed modeling language, i.e. causing modelers to focus on certain aspects of their work, could be indicatively confirmed in our coding results. The distribution of latent themes in the models was highly dependent on the used language. Deploying a language with specific process-oriented semantics such as BPMN appears to prompt modelers to emphasize how work is organized temporally and causally. Leaving open the semantic categories appeared to lead cause a tendency to focus on required resources, skills and work results, which was not observed in any other series. This difference indicates that the appropriateness [11] of a modeling language deployed in workshops involving manual articulation indeed is a relevant issue during modeling. It needs to be explicitly considered when selecting a language aiming at capturing knowledge of inexperienced modelers to avoid the need for translating mental concepts to the semantics of an external model representation [13].

These findings lead to several implications with respect designing or selecting modeling languages to be used by inexperienced modelers: First, the coding process has shown that intersubjective understandability is rarely a problem. The latent semantics of modeling elements was identified with only few inconsistent codings. Consequently, the use of standardized modeling languages seems to be necessary only if the IT-supported processing of models is required (e.g., for workflow execution [29]). Second, it might be useful to provide a set of pre-specified modeling elements to prompt thinking about particular aspects of work. Using a subset of elements provided by an existing modeling language as a foundation might provide a gateway towards IT-supported processing (e.g., as suggested for BPMN [21]).

6 Conclusion

In this paper, we have presented a study to examine the practical use of different modeling languages by inexperienced modelers. We have aimed to identify, whether the semantics of a prescribed language has an impact on the aspects modelers choose to represent in their models of work. We could confirm that modelers focus on different aspects depending on the constraints imposed by the modeling languages. The data suggest that orientation on the flow of activities is not necessarily an intuitively chosen way of representing work by inexperienced modelers. Rather, they have focused on required resources and work results in our study. Based on these results, we suggest to further examine the use of flexibly deployable modeling languages to achieve appropriateness for people without conceptual modeling experiences.

The present study has several limitations. First, data generated from real-world workshops is only of limited validity when aiming at comparing the modeling results produced under different languages constraints. Such settings usually lead to confounding variables that can hardly be observed and accounted for systematically during data analysis. Now that indicative results provide starting points for an informed formulation of more specific hypotheses, studies conducted under more strictly controlled settings, such as (quasi-)experiments, could allow addressing this issue. Second, the absence of conceptual modeling experience has only been considered in our study as a binary value and surveyed through participants' self-declarations. This might be an oversimplification, as our data seems to indicate, that there might be latent influences of participants' earlier acquired skills, such as spreadsheet design (related to data modeling [30]) or software development (related to process modeling [29]).

We currently conduct a follow-up study to the present work, in which we examine prior knowledge of inexperienced modelers and its relation to deployed modeling strategies and used semantics more thoroughly. This will allow a more detailed and grounded discussion of this issue in future work.

Acknowledgements

The author would like to thank Marlene Sonnleitner and Stefan Haselböck for conducting the coding activities necessary to perform the present study.

References

1. Front, A., Rieu, D., Santorum, M., Movahedian, F.: A participative end-user method for multi-perspective business process elicitation and improvement. SoSyM. 1–24 (2015).
2. Antunes, P., Simões, D., Carriço, L., Pino, J.A.: An end-user approach to business process modeling. Journal of Network and Computer Applications. 36, 1466–1479 (2013).
3. Korhonen, J.J., Kaidalova, J.: Enterprise Modeling Facilitating Business and IT Alignment Along the Social Dimension. Proc. of IEEE CBI (2015).
4. Frederiks, P.J.M., van der Weide, T.P.: Information modeling: The process and the required competencies of its participants. Data & Knowledge Engineering. 58, 4–20 (2006).
5. Herrmann, T., Loser, K.U., Jahnke, I.: Sociotechnical walkthrough: a means for knowledge integration. The Learning Organization. 14, 450–464 (2007).

6. Smeds, R.: Simulation for accelerated learning and development in industrial management. Production Planning & Control. 14, 107–110 (2003).
7. Figl, K.: Comprehension of procedural visual business process models. Business & Information Systems Engineering. 59, 41–67 (2017).
8. Türetken, O., Vanderfeesten, I., Claes, J.: Cognitive style and business process model understanding. Proc. of CAISE 2017 (2017).
9. Simões, D., Antunes, P., Cranefield, J.: Enriching knowledge in business process modelling: a storytelling approach. In: Innovations in KM. pp. 241–267. Springer (2016).
10. Hjalmarsson, A., Recker, J.C., Rosemann, M., Lind, M.: Understanding the behavior of workshop facilitators in systems analysis and design projects: Developing theory from process modeling projects. Communications of the AIS. 36, 421–447 (2015).
11. Krogstie, J., Sindre, G., Jørgensen, H.D.: Process models representing knowledge for action: a revised quality framework. eJ. 15, 91–102 (2006).
12. Zarwin, Z., Bjekovic, M., Favre, J.M., Sottet, J.S., Proper, E.: Natural modelling. Journal of Object Technology. 13, 4:1–36 (2014).
13. Falkenberg, E.D., Hesse, W., Lindgreen, P., Nilsson, B.E., Oei, J.L.H., Rolland, C., Stamper, R.K., van Assche, F.J.M., Verrijn-Stuart, A.A., Voss, K.: A Framework of Information System Concepts. The FRISCO Report. IFIP WG 8.1 (1998).
14. Recker, J.C., Safrudin, N., Rosemann, M.: How novices design business processes. Information Systems. 37, 557–573 (2012).
15. Monsalve, C., April, A., Abran, A.: Representing unique stakeholder perspectives in BPM notations. Proc. of SERA 2010. 42–49 (2010).
16. Patig, S., Casanova-Brito, V.: Requirements of process modeling languages--results from an empirical investigation. Proc. of Int. Conf. WI, AIS (2011).
17. Genon, N., Heymans, P., Amyot, D.: Analysing the Cognitive Effectiveness of the BPMN 2.0 Visual Notation. In: Software Lang. Eng.. pp. 377–396. Springer (2011).
18. Moody, D.: The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Trans. on Software Eng. 35, 756–779 (2009).
19. Muehlen, Zur, M.: Class Notes: BPM Research and Education—A Little Knowledge is a Dangerous Thing. BPTrends, January. 1–5 (2008).
20. Oppl, S.: Articulation of work process models for organizational alignment and informed information system design. Information & Management. 53, 591–608 (2016).
21. Silver, B.: BPMN Method and Style: A levels-based methodology for BPM process modeling and improvement using BPMN 2.0. Cody-Cassidy Press, US. (2009).
22. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative research in psychology. 3, 77–101 (2006).
23. Trickett, S.B., Trafton, J.G.: A primer on verbal protocol analysis. In: The PSI Handbook of Virtual Environments for Training and Education. pp. 332–346 (2009).
24. Fujimura, J.H.: Constructing “Do-Able” Problems in Cancer Research: Articulating Alignment. Social Studies of Science. 17, 257–293 (1987).
25. Fjuk, A., Nurminen, M.I., Smørdal, O.: Taking Articulation Work Seriously: an activity theoretical approach. (1997).
26. Soderstrom, E., Andersson, B., Johannesson, P., Perjons, E., Wangler, B.: Towards a framework for comparing process modelling languages. Proceedings of CAiSE 2002. 600–611 (2002).
27. List, B., Korherr, B.: An evaluation of conceptual business process modelling languages. SAC '06: Proceedings of the 2006 ACM symp. on Applied computing. 1532–1539 (2006).
28. Recker, J.C., Rosemann, M., Indulska, M., Green, P.: Business process modeling-a comparative analysis. Journal of the Association for Information Systems. 10, 1 (2009).
29. Curtis, B., Kellner, M.I., Over, J.: Process modeling. CACM. 35, 75–90 (1992).
30. Carlsson, S.A.: Why Johnny can't or won't spreadsheet. ScIS. 1, 118–142 (1989).