Information Systems 35 (2010) 467–482

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Activity labeling in process modeling: Empirical insights and recommendations

J. Mendling [a,*], H.A. Reijers [b], J. Recker [c]

[a] Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany
[b] Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[c] Queensland University of Technology, 126 Margaret Street, QLD, 4000 Brisbane, Australia

ARTICLE INFO

Keywords:
Business process modeling
Model quality
Survey
Systems analysis and design

ABSTRACT

Few studies have investigated the factors contributing to the successful practice of process modeling. In particular, studies that contribute to the act of developing process models that facilitate communication and understanding are scarce. Although the value of process models is not only dependent on the choice of graphical constructs but also on their annotation with textual labels, there has been hardly any work on the quality of these labels. Accordingly, the research presented in this paper examines activity labeling practices in process modeling. Based on empirical data from process modeling practice, we identify and discuss different labeling styles and their use in process modeling praxis. We perform a grammatical analysis of these styles and use data from an experiment with process modelers to examine a range of hypotheses about the usability of the different styles. Based on our findings, we suggest specific programs of research towards better tool support for labeling practices. Our work contributes to the emerging stream of research investigating the practice of process modeling and thereby contributes to the overall body of knowledge about conceptual modeling quality.

1. Introduction

In recent years, the conceptual mapping of processes in the form of process models has emerged as a primary reason to engage in conceptual modeling [1] and is considered as a key instrument for the analysis and design of process-aware information systems [2], service-oriented architectures [3], and web services [4] alike. To that end, process models typically describe in a graphical way at least the activities, events, states, and control flow logic that constitute a business process [5]. Additionally, process models may also include information regarding the involved data, organizational and IT resources, and potentially other artifacts such as external stakeholders and performance metrics, see e.g., [6]. Similar to other forms of conceptual modeling, process models are first and foremost required to be intuitive and easily understandable, especially in information systems project phases that are concerned with requirements documentation and communication [7].

Process modeling has been around for some 30 years. However, only of late has research started to examine quality aspects pertaining to process modeling. In fact, quality issues of conceptual modeling in general have only recently been receiving increased attention in academia [8]. Notwithstanding the research findings collected to date, surprisingly little is known about the actual "practice of process modeling" and the factors that contribute to building a "good" process model, for example one that aids human understanding of the depicted business domain [9]. Work has been carried

* Corresponding author. Tel.: +49 7 313 89492.
E-mail addresses: jan.mendling@wiwi.hu-berlin.de (J. Mendling), h.a.reijers@tue.nl (H.A. Reijers), j.recker@qut.edu.au (J. Recker).

0306-4379/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.is.2009.03.009

out, for instance, that examined the impact of process model structure, model user competency and process modeling language on process model understanding. While the impact of structural properties is clearly identified [10], it is also reported that model readers systematically overestimate their ability to draw correct conclusions from a model [9]. It was also found that the choice of languages used for process modeling (e.g., BPMN versus EPCs) has only insignificant effects on process model understanding [11]. Other research has successfully investigated the graphical constructs and their meaning in process models, e.g., [12], the expressiveness and validity of control flow aspects in process models, e.g., [13], or process-related aspects such as data and resources, e.g., [14,15].

This situation raises the question of other antecedents of process model understandability. Most of the previous work has focused on syntactic quality aspects [16]. In contrast, semantic and pragmatic aspects of model quality have mostly been neglected. In particular, little attention has been devoted to a very essential task in process modeling—the *labeling* of the graphical constructs, in particular of the constructs representing "activities" (or "tasks", or "work to be performed") in a process model. This is rather surprising given that—clearly—the true meaning of any construct in a process model is only revealed when model users read and *intuitively* understand the labels assigned to the construct. Current practice indicates that the labeling of activity constructs is a rather arbitrary task in modeling initiatives and one that is sometimes done without a great deal of thought [17]. This can undermine the understandability of the resulting models in cases where the meaning of the labels is ambiguous, not readily understandable, or simply counter-intuitive to the reader.

Accordingly, in our work we seek to address this gap and contribute to the existing line of work towards more understandable process models. The objective of our research is to investigate the styles that are in use to annotate activities in process models and how these styles affect the understandability of such models.[1] More precisely, the *aim of this paper* is to suggest, based on our empirical findings, an imperative style for modelers to create more understandable process models.

We proceed as follows. In Section 2 we discuss the theoretical foundation for our work and investigate current labeling practices in process modeling. In Section 3 we discuss the design of, conduct of, and findings from an experiment with process modelers. In Section 4 we then discuss the implications of our findings and suggest specific programs of research towards better support for process model labeling practices. We conclude in Section 5 by reviewing our contributions, and discussing some conclusions.

[1] We recognize the need to extrapolate our research to other aspects of process models, such as the data, resource and control flow perspective. We deemed the focus on "activity constructs" a suitable starting point for our endeavor due to the centrality of the "activity" concept in process modeling.

2. Background

In presenting the background to our research, we refer to a theory of multimedia learning originating from cognitive science. This theory suggests that labeling practices are indeed significant factors contributing to how well or how poorly process models can be understood by their end users. To determine what a good labeling style is, we then identify different styles of labeling being used in practice. We describe how the exploration of a large number of real-life process models gives us this insight. One of the styles that is encountered is the usage of verb–object labels. As this style is widely promoted in the literature [18–20], we formulate several hypotheses on its presumed superiority over the other styles encountered in our exploration.

2.1. Theoretical foundation

Dual Coding Theory [21] suggests that individuals have two separate channels—visual and auditory—that they use when processing information. The two channels complement each other, such that receiving simultaneous information through each channel improves understanding compared to receiving information through one channel only. In other words, individuals understand informational material better when it is provided through both auditory (i.e., words) and visual (i.e., images) channels.[2]

Based on this observation, the Cognitive Theory of Multimedia Learning (CTML) [23,24] suggests that learning material intended to be received, understood and retained by its recipients should be presented using *both* words and pictures. This sounds conducive to the task of process modeling, where both visual (graphical constructs) and auditory (labels and text annotations) material are available to add information about a business domain in a process model. However, due to the overall limited number of graphical constructs used in a process model—there are typically few if not only one graphical construct for representing activities—most of the critical domain information is contained in the textual labels of the constructs, viz., in *auditory* channels. Based on CTML it can thus be expected that model understanding can be improved if better guidance can be provided for the act of labeling of process model constructs.

The general principle that our expectation builds on is described by Mayer [24] as the "multimedia principal". And indeed, prior research on conceptual modeling has successfully demonstrated that the multimedia principal informs model understanding. Empirically observable differences in model understanding based on the multimedia principal were found, for instance, in the data modeling domain [25,26] as well as in the process modeling domain [11].

2.2. Labeling styles in practice

For business process modeling, the labeling of constructs such as activities is often more art than science. In

[2] Indeed, most people read by speaking out the words of the text in their mind, which even suppresses visual activation [22].

practice, a number of informal guidelines exist that typically suggest a verb–object convention (e.g., approve order, verify invoice) for labeling activities, e.g., [18–20]. This convention is similar to a style that is advocated in guidelines that support the creation of understandable use case descriptions, a widely accepted requirements tool in object-oriented software engineering [27,28]. We will refer to this labeling style of activities as the *verb–object style*. But as much promotion it receives in the process modeling domain, both anecdotal evidence and causal inspection of real process models indicate that this labeling style is neither universally nor consistently applied. Even the practical guide for process modeling with ARIS [29, pp. 66–70] shows models with both actions as verbs and as nouns. Also, one may think that the more information contained in the labels, the clearer the meaning will be to the reader. Recent research, however, uncovered that shorter activity labels improve model understanding [30].

To get a better idea of the variety in labeling styles being applied in practice, we turn to the SAP Reference Model [31]. The development of the SAP reference model started in 1992 and first models were presented at CEBIT'93 [31, p. VII]. Since then, it was developed further until version 4.6 of SAP R/3, which was released in 2000. Overall, the SAP reference model includes 604 business process models depicted using the Event-driven Process Chains (EPC) notation, capturing information about the SAP R/3 functionality to support the business processes in a wide range of organizations. With the SAP solution being the market leading tool in the Enterprise Systems market we feel that the examination of SAP process models gives us a good understanding of the use of process models in real-life business contexts. Amongst other application areas, the SAP reference model denotes a frequently used tool in the implementation of SAP systems [32], and much literature has covered its development and use [31]. Furthermore, it is frequently referenced in research papers as a typical reference model and used in previous examinations of process modeling, e.g., [10,33,34].

Altogether, the 604 EPC models in the SAP reference model include 19,838 activity labels, which we all manually inspected and classified. In 94% of these cases (18,648 instances), the activity labels refer to a certain action that should be undertaken, such as *check billing block* or *order execution*. This is not so for 6% of the labels, because they neither include a verb nor a noun that refers to an action, consider, for instance, *status analysis cash position*. We will refer to this style as the *rest* category.

Note that the EPC models considered were designed based on the functionality and the terminology of the SAP system which might create different biases. On the one hand, system terminology could potentially be less intuitive compared to labeling in conceptual design models. On the other hand, the labels could be more precise than labels in conceptual modeling practice. Yet, neither the high frequency of verb–object styles nor the variety of labeling styles in use directly suggest such bias.

Despite the wide proliferation of 18,648 "action-oriented" labels of the 19,838 activity labels in the SAP reference model overall, this situation does not imply that the verb–object style is strictly enforced within this subset. Rather, it is applied to only about two-third of the "action-oriented" labels (60% of *all* activity labels). The remaining subset of the "action-oriented" labels (34% of *all* activity labels) denote labels where the action is grammatically captured *as a noun*. This noun can be either a gerund of the verb or a noun that is derived from a verb, like *order processing* or *invoice verification*. We will refer to this style of labeling as the *action-noun style*. The overall result from classifying all 19,838 activity labels can be seen in Table 1.

Table 1
Distribution of activity label styles in the SAP reference model.

Verb–object labels	Action-noun labels	Rest	Sum
11,830	6808	1201	19,838
60%	34%	6%	100%

We will now consider these data in more detail. More precisely, for each of the labeling styles found, we perform a grammatical analysis using the lexical database WordNet [35] to identify potential types of interpretation ambiguity. This grammatical analysis builds on the identification of syntactic categories such as *noun* and *verb*. Further categories like *adjective* and *adverb* could also be used but do not pertain to activity labeling in process modeling, which is why we excluded these categories from our analysis. For many words, the syntactic category can be identified purely syntactically, as for instance with the word *grammar*, which is a noun. Some words, however, are *ambiguous* regarding the category they belong to (when analyzed in isolation). Consider the word *design*, which can be a verb (*to design*) or a noun (*the design*) depending on the grammatical context. As these examples from natural language processing show, ambiguity can be a significant impediment to ease of understanding. In light of this observation we thus argue that those labeling styles should be considered in process modeling that are least susceptible to ambiguity. We illustrate our argument with examples from the SAP Reference Model:

Verb–object labels: Most of the verb–object labels seem intuitively understandable to us. Still, there are some cases that are ambiguous from a grammatical point of view: The English language allows for a so-called *zero derivation* beyond the suffix *-ize* and the suffix *(i)fy* derivation of verbs from nouns [36]. As a consequence, the same word can both be a noun and a verb. Consider, for example, the labels *measure processing*, *export license check*, and *process cost planning*. They have in common that the first word can be a verb, but reading it as an object describing an action is also possible. *Measure processing* could potentially refer to the processing of a measure or to the measurement of a processing. The same observation holds for the other labels. Some of these ambiguities can be resolved by considering context information, such as the labels of the other activities in the same process model. If the verb–object style was consistently used as a standard throughout a process model, it would be clear to interpret the first term as a verb.

Action-noun labels: With respect to action-noun labels, some of these can be easily interpreted, but again there can be cases of grammatical ambiguity. Consider, for instance, *notification printing*. Again, there are two potential interpretations: a notification is printed, or someone is notified of a printing job. Alternatively, the verb could just have been forgotten by the modeler. This interpretation is likely in cases where the action noun could also be an object, like *order*, which can refer to both an action or an object. We call this type of ambiguity the *action-object ambiguity*. In such cases, the model reader might be tempted to infer the action by considering the context of the activity. Syntactically, the label could be easily extended with such semantically diverse verbs as *start*, *stop*, or *schedule*. Using a verb–object style would have avoided the problem of action-object ambiguity and the necessity of having to infer a verb to establish the appropriate meaning.

Rest labels: Some of the rest labels clearly point to a specific business object, for instance *status analysis cash position*, such that a verb could potentially be inferred from the context. Yet there are also activity labels like *DEÜV* and *Jamsostek* that are altogether difficult to understand. Presumably, the first one refers to the German regulation for data storage and transmission (DEÜV Datenerfassungs- und Übertragungsverordnung) and the second to the Indonesian social security system. Clearly labels of the "rest" category require crystal clear context information, otherwise an inference of the action to be performed is a highly problematic task due to the occurrence of *verb-inference ambiguity*, i.e., the problem of inferring from the context of the label the type of action to be performed as part of the considered process task.

In conclusion, the three different classes exhibit different types of ambiguities. For the verb–object style, we found instances of *zero-derivation ambiguity* in the SAP reference model. Altogether, we identified exactly 600 labels with such ambiguity; these labels contained 23 different verbs including *change*, *design*, *process*, and *report*. For the action-noun style, this problem class is relevant, too. Furthermore, this style is susceptible to *action-object ambiguity*, if an action noun can also refer to an object. We counted 615 cases of such ambiguities. Finally, the rest group of labels, which do not mention an action at all, faces *verb-inference ambiguity* (1190 cases). These three ambiguity classes differ in occurrence frequency: while the *zero-derivation ambiguity* requires the unlikely combination of a verb and an action object, the *action-object ambiguity* is found more often since many documents in a business context are synonymous to an action noun (e.g., *order*, *receipt*, *confirmation*). The *verb-inference ambiguity* is the most significant one, since all labels of the rest group suffer from it.

2.3. Hypotheses

On basis of the findings discussed above, our contention is to conjecture about the influence of choice of labeling styles on the pragmatic quality of process models in terms of unambiguously facilitating action [16] and usage [37]. We summarize our expectations as follows. First, we formulated and grounded our expectation that model understanding can be improved by guiding the act of labeling following the theory of multimedia learning. In search for candidate guidelines for labeling activities, anecdotal evidence, the study of the SAP reference model, and our literature review suggest the verb–object labeling style to be the strongest candidate style. Our empirical exploration of the SAP reference model indeed confirmed the wide application of this style in practice. Yet, we also found that this style is not the only style being applied: a large fraction of activity labels follows an action-noun style, and there are also other (rest) styles to be found in process models. Our grammatical analysis of the three modeling styles, as described in the previous section, suggested that the verb–object style appears to be the least susceptible to various types of interpretation ambiguity, indicating its superiority in terms of clarity of specification.

In light of these observations, we suggest the following primary conjecture that we seek to test in our study. Based on our grammatical analysis, we theorize that process modelers perceive the verb–object style to be superior to the action-noun and rest labeling style alongside two dimensions:

- *perceived ambiguity*: the degree to which an individual believes that a label is ambiguous, and
- *perceived usefulness* (PU): the degree to which an individual believes that a label is useful for understanding the process modeled.

This conjecture rests on the observation that the verb–objective style is less prone to result in misinterpretation and confounding complexity. After all, our grammatical analysis showed that it is least susceptible to ambiguity. We thus advance the following two primary hypotheses we seek to test in this study. First, we theorize that users working with process models have a clear preference for labeling styles that avoid ambiguity:

H1. Verb–object style labels are least frequently perceived as being ambiguous, followed by action-noun style labels, and finally rest labels.

Second, we theorize that end users working with process models have different perceptions of the usefulness of the labels for understanding the process modeled, dependent on the labeling style in which the label is articulated. More specifically:

H2a. Verb–object style labels are perceived as more useful for understanding the process model than action-noun style labels.

H2b. Verb–object style labels are perceived as more useful for understanding the process model than rest style labels.

H2c. Action-noun style labels are perceived as more useful for understanding the process model than rest style labels.

Hypotheses H2a–H2c rest on the assumption that the perceived usefulness of a label is negatively influenced by the perceived ambiguity of the labeling style used, based on the contention that the grammatical style of a labeling type can lead to misinterpretation and confounding complexity. To gather empirical evidence for this contention, we advance the following, additional hypothesis that we will test:

H3. Perceived ambiguity of a labeling style is negatively associated with the perceived usefulness of the label.

In our study, we also need to consider that differences in the perceptions about the ambiguity and usefulness of a process model label can also stem from differences between the study participants. Recent experimental research on conceptual modeling, most notably [26,38,39], has indicated significant differences in the understanding of conceptual models stemming from two characteristics of the conceptual model readers, these being *knowledge of the application domain* (e.g., [38]) and *familiarity with the technique or notation used for conceptual modeling* (e.g., [26]). CTML [24] suggests that previous knowledge of the domain covered in the conceptual modeling lowers the cognitive load required to develop a mental model of the information displayed in the conceptual model, and hence, model understanding will be easier. This is because readers can bring to bear an understanding of the semantics, relevant entities or procedures that make up the application domain depicted in a model. Similarly, expertise or knowledge of the conceptual modeling artifact (i.e., the method, technique or notation used) has been shown to increase the quality of the models produced (e.g., [40,41]), and sometimes to increase the understanding of the models produced [38]. The noted interaction effects of notation familiarity are speculated to stem from a modeler's self-perception about his or her modeling skills. In other words, a modeler that deems himself or herself to be experienced, may approach modeling tasks and outcomes differently to someone that believes oneself to be a novice.

In light of these findings we thus advance the following, additional exploratory hypotheses that seek to investigate how knowledge about the application domain and familiarity with the process modeling notation used act as moderating variables to the propositions outlined above:

H4a. Knowledge about the application domain moderates the strength of the relationship between labeling style and perceived usefulness of the label.

H4b. Familiarity with the process modeling notation moderates the strength of the relationship between labeling style and perceived usefulness of the label.

3. Research method

3.1. Research design and conduct

To test the hypotheses advanced in the previous section, we developed a (self-administered) questionnaire to gather quantitative insights. With this questionnaire we asked participants about the perceived ambiguity of certain activity labels, as well as their perceived usefulness. Along with the questionnaire, we presented to the participant a number of activity labels as part of a specific process model. This has been done for several reasons. First, a label in a business process model is never interpreted in isolation. Various other labels in the model and the control flow relationship between the activities establish a context against which a single label is interpreted. Since we do not aim to gain insight into labels *per se* but in their use in process models, we have to present all the labels that are discussed in the questionnaire in the context of a model. Second, we had to choose a model from practice; otherwise there would have been the risk that we would (unconsciously) tailor it to meet our hypotheses. Third, this process model had to show a substantial variation in the labeling styles being used so that we can limit potential bias in our research design.

Following these considerations we selected a model of a complaint process from a department of a Dutch governmental agency, which is concerned with complaint handling (see Fig. 1). The model follows the EPC notation, which is one of the most popular modeling techniques in industry [1]. Indeed, it is the same technique as applied in the SAP reference model. In an EPC, so-called *functions* (rectangles) correspond to the various tasks that may need to be executed (e.g., *register receipt date of complaint letter*). *Events* (hexagons) describe the situation before and after a function is executed (e.g., "customer at desk"). *Logical connectors* (circles) define routing rules. In particular, there are three types of connectors: the logical AND for concurrency, XOR for exclusive choices, and OR for inclusive choices. Functions, events, and connectors are the classical elements of control flow modeling. These routing elements are also included in other modeling languages like BPMN, YAWL, and UML Activity Diagrams, which supports generalizability and repeatability of our procedure.

The given model roughly describes the following procedure to handle the complaints that the agency receives. A new case is opened if a new complaint is received—be it by means of a phone call, personal contact, or letter. In some situations, the complaint must be referred to another party, either internal or external to the agency involved. Internal referrals have to be put on a so-called *incident agenda*, while external referrals always require a confirmation. In both cases the referral is archived in parallel. As a final step in this procedure, the complainant is informed. If no referral is required, a complaint analysis is conducted. Later, the complaint is archived and the complainant is contacted, with an optional follow up (see Fig. 1).

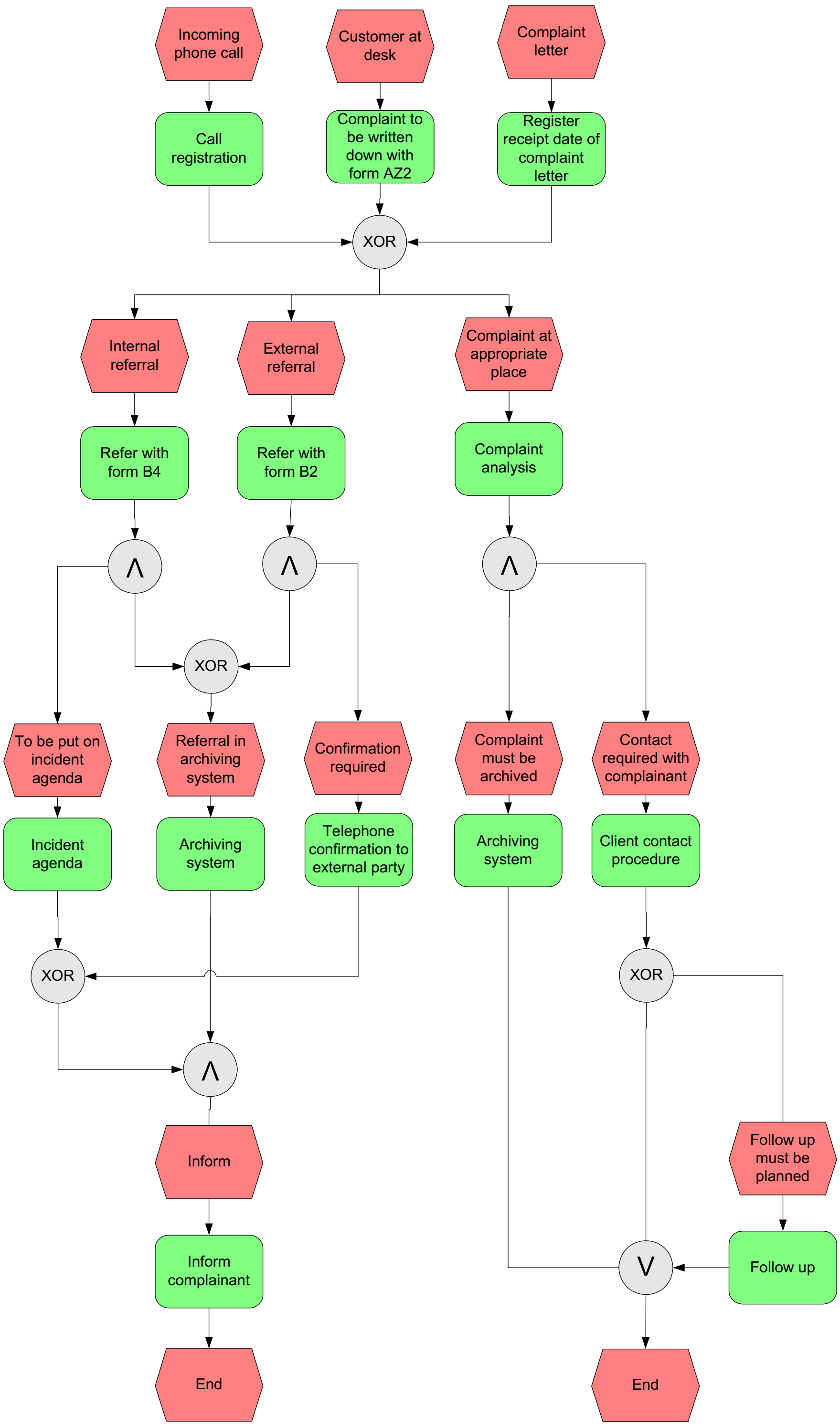

Fig. 1. The complaint handling process.

The complaint process model in Fig. 1 is at the heart of our questionnaire, which is subdivided into three parts. In the first part we recorded demographic information about the participants including gender, years of tertiary education, preliminary knowledge of process modeling, and number of EPC models created. These questions were used to gather information about the demographic distribution of the study participants.

In order to measure knowledge about the application domain (in our case: complaints handling), we asked participants whether they had previous experience with complaints handling processes (yes/no). Since we did not expect much domain knowledge in a student population, the use of a more extensive scale (like the one described in [26]) was not considered. In order to measure respondents' familiarity with the EPC notation, we adapted a three-item scale for notation familiarity developed by Recker [42], which is based on Gemino and Wand's three pre-test questions about the *familiarity*, *competence*, and *confidence* of respondents with respect to an analysis method (see Appendix A and [26]). Accordingly, the three-item familiarity scale assesses familiarity with the (EPC) process modeling notation in a sense of generally felt familiarity (Fam1), self-perceived competence with the notation (Fam2) and self-perceived confidence in using the notation (Fam3). Appendix A lists all items used in the questionnaire.

The second part of the questionnaire shows the process model as depicted in Fig. 1. In order to gather data to examine hypothesis H1, the participants were asked to identify the top three activity labels that they consider to be the most ambiguous. In the third part, we sought to gather data to examine hypotheses H2a–H2c. In order to evaluate usefulness perceptions, we developed a two-item measurement scale that stresses the act of understanding. Specifically, we used the Perceived Usefulness scales developed by Maes and Poels [43] as a basis for our measurement development. The motivation is that their PU measures were developed specifically for the conceptual modeling context. Our scales were worded *Overall*, *I found* [label] *useful for understanding the process modeled* and *Overall*, *I think* [label] *improves my performance when understanding the process modeled*.[3] We asked the participants for their perception in these terms of six activity labels from the process model, using a 7-point Likert scale with the anchor points "Disagree strongly" and "Agree strongly".

We chose not to measure perceived usefulness for each of the 12 distinct labels shown in Fig. 1 but instead to record these measures for six labels only. We have done so for the pragmatic reason of not making our data collection instrument—and the conduct of the experiment—unnecessarily long. Considering six labels allowed us to obtain 6 (labels) $\times$ 29 (number of responses) $=$ 174 data points for hypothesis testing, which we deemed sufficient for our analysis. We arbitrarily selected two labels for each of the three styles we identified in the previous section, these being *register receipt date of complaint letter* and *inform complainant* as verb–object labels, *registration* and *follow up* that follow the action-noun style, as well as *archiving system* and *incident agenda* for the rest group. We consider our selection strategy sufficiently randomized based on the observation that neither our research objectives nor our hypotheses address the choice of word items or the specificity of the word items used within these labels. Hence, there was no motivation for us to prefer any particular label over another.

3.2. Results

Demographics: The questionnaire of our survey was filled out by 29 students who were at that time following a post-graduate course on process modeling at Eindhoven University of Technology in the Netherlands. Participation was voluntary, and as a reward we offered the students a copy of the study results. Twenty-five participants were male, while 4 were female. While some of the participants only had followed university courses for one year, most of them had done so for three years or more, with 3.8 years of study being the mean value. Half of the population had preliminary experience with business process modeling, either professionally or through previous courses. Four persons had not yet worked with EPCs, but the average participant had known them for three months and created 10 models so far. Altogether, 25 out of the 29 participants self-assessed their familiarity with EPCs as better than 3 (average total factor score), with the median being 4.5. We included a brief description of the EPC notation similar to [45, p. 36] such that the participants would in any case have the necessary background to understand the process model. Finally, there were six persons who had some preliminary knowledge of complaint handling processes.

Overall, the study population contained individuals with some application domain knowledge and familiarity of the EPC notation, but without high levels of either. Studies using students have been often criticized for lack of external validity. Despite this criticism, we agree with Gemino and Wand [26,46], Recker and Dreiling [11] as well as Batra et al. [47] that the selection of students over practitioners in this type of research can in fact be advisable. Results from both domain understanding and problem solving tasks could have been confounded by participants that are able to bring to bear prior application domain knowledge in one of the areas [48]. Also, post-graduate students (like the one participating in our study) have been found to be adequate proxies for analysts with low to medium expertise levels [46,49].

Perceived ambiguity: The second part of the questionnaire focused on the relationship between label types and perceived ambiguity, as stated in hypothesis H1. We asked the participants to identify those three activity labels that they consider to be the most ambiguous. Since there are 12 distinct labels in the model and 29 participants, we received 348 assessments whether a particular label (belonging to a certain label type) was considered to be among the three most ambiguous ones. The labels *incident agenda*, *complaint analysis*, and *archiving system* were mentioned most frequently (14, 13, and 12 times). Note that the first and third labels belong to the rest group, while *complaint analysis* follows the action-noun style. In contrast, the most ambiguous label following the

[3] We chose not to adapt the PU1 item from [43]. This item cannot be reasonably applied to text labels. The item would have read *Overall, I think the* [label] *would be an improvement to a textual description of the business process*, which essentially is a tautology. Also note that we focus on perceived usefulness in our experiment for its importance as a key antecedent to actual usage [44]. The research by Maes and Poels [43] is much broader in its goal to reveal the contribution of different dimensions to the quality of conceptual models.

verb–object style—*inform complainant*—received only two counts overall. The estimated probability of a label for being mentioned among the three most ambiguous ones was 0.13 for verb–object labels, 0.24 for action-noun labels, and 0.45 for the rest group. The 95% confidence intervals show little overlap: 0.08–0.19 for verb–object label, 0.17–0.31 for action-noun labels, and 0.32–0.58 for the rest, which correspond to our expectations. To calculate reliability of the assessments made by the study participants, we calculated Cohen's Kappa [50] statistic to examine the level of agreement between study participants on which labels were most ambiguous. The Kappa statistic measures inter-rater reliability whilst controlling for change agreement, and is the generally agreed to be the most adequate tool to measure inter-rater reliability [51]. We obtained a Kappa value of 0.607, which can be classified as substantial or good [51].

As per our hypothesis H1, we were interested in testing whether the differences between the label types as noted are significant. An analysis of variance (ANOVA) test was not applicable, since the variance of the variable values is not homogeneously distributed and because the dependent variable is not on scale level. Instead, we applied Friedman's two-way analysis of variance by ranks [52]. For each participant, we determined an individual ranking of the three label types. This was achieved as follows. For each label type, we determined its relative proportion among the labels that were rated as most ambiguous by that participant. This gives us 29 matched evaluations, leading to rank totals for the three label types as shown in Table 2. As can be seen, verb–object labels receive the lowest rank total, which means that this type is least often considered as containing ambiguous labels. We advance the null hypothesis that there are no differences in individual rankings of the three label types, i.e., that each label type would be mentioned similarly in the top three lists in each of the 29 evaluations. In seeking to refute this null hypothesis, we computed the Friedman statistic χ_r^2. Note that the Friedman statistic χ_r^2 is distributed approximately as chi square [52, p. 168]. For this case, it turns out that $\chi_r^2 = 6.28$ with $df = 2$, which means a significant difference in the rankings of the three labeling styles at a 95% confidence level. This result lends support to hypothesis H1. We conclude that verb–object style labels are indeed least frequently perceived as being ambiguous, followed by action-noun style labels, and finally rest labels.

Perceived usefulness: In the third part of the questionnaire, we recorded the perceived usefulness of six activity labels, two for each label type. We used two measures for PU as described above. More specifically, the used scales measure the extent to which a label is *useful for understanding* and *improves the performance when understanding*. We received 174 responses (6×29) that we were able to link to label types. Based on these data, we examined the hypotheses H2a–H2c.

Before proceeding with hypothesis testing, we first examined reliability and validity of the PU measures used. Reliability refers to the internal consistency of scales. The most widely used test for internal consistency is Cronbach's α, which should be higher than 0.8 [53].

Table 2
Rank totals for the three label types.

	Verb–object labels	Action-noun labels	Rest
Observed ranked total	49	57	68
Expected ranked total	58	58	58

A second test uses the composite reliability measure p_c, which represents the proportion of measure variance attributable to the underlying trait. Scales with p_c values greater than 0.5 are considered to be reliable [44]. For the PU measures, we obtained a Cronbach's α value of 0.857, and a p_c value of 0.884, suggesting adequate reliability of the measures. To establish validity of the measures, we examined convergent and discriminant validity of the PU measures. Convergent validity can be tested using three criteria suggested by Fornell and Larcker [54]:

(1) All indicator factor loadings should be significant and exceed 0.6.
(2) Construct composite reliabilities p_c should exceed 0.8.
(3) Average variance extracted (AVE) by each construct should exceed the variance due to measurement error for that construct (i.e., AVE should exceed 0.50).

Factor loadings for the two PU measures were 0.936 and 0.936 and significant at $p = 0.000$. Composite reliability of the PU construct was estimated to be 0.884, and average variance extracted was computed to be 0.936. These results suggest adequate convergent validity. To check for discriminant validity, we considered whether measures used for the PU construct would cross-load on other constructs considered (in our case, measures for notation familiarity). The test for discriminant validity is met when the AVE for each construct exceeds the squared correlation between that and any other construct considered in the factor correlation matrix. The squared correlation between the PU and the familiarity factor was computed to be 0.030, which shows that the AVE measures for both PU (0.936) and notation familiarity (0.927) well exceeded the squared correlation between the factors. Appendix B summarizes factor loadings, communalities, and correlations.

Next, to test the hypotheses, we first constructed a box-plot for the average total factor scores for the PU variable, and examined the rank correlations as well as the differences in variance between the average total factor scores for the different label types. Fig. 2 gives the box plots.

As illustrated by the box-plot in Fig. 2, verb–object labels were found to be best in terms of their perceived usefulness, followed by action-noun labels, and then the rest group. Perusal of Table 3 further shows that the reported 95% confidence intervals around the means hardly overlap between the label types. In particular, the verb–object style can easily be distinguished from the action-noun style: the upper bounds of the confidence intervals for the action-noun style are strictly lower than

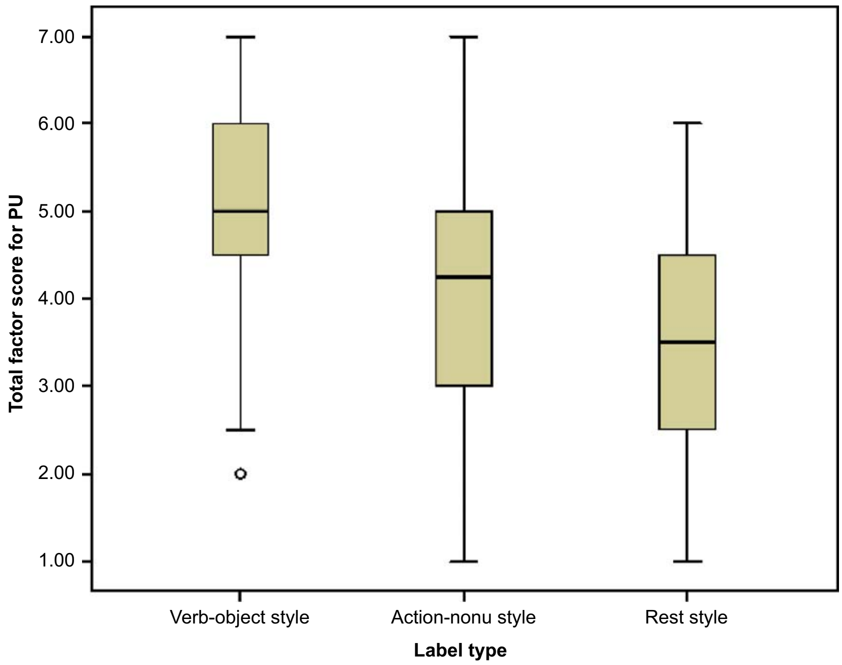

Fig. 2. Box-plot of perceived usefulness rankings, by label type.

Table 3
Perceived usefulness of label types.

	Perceived usefulness (avg. total factor score)
Verb–object	
95% upper bound	5.304
Mean	5.000
95% lower bound	4.696
Action-noun	
95% upper bound	4.480
Mean	4.121
95% lower bound	3.761
Rest	
95% upper bound	3.905
Mean	3.552
95% lower bound	3.199

the lower bounds for the verb–object style. These results lend initial support to hypotheses H2a–H2c.

As a next step, we examined whether the noted differences are statistically significant. In the data, we identified a significant negative Spearman rank correlation between the label style and its perceived usefulness (−0.430 at 99% significance level). This finding suggests that a deviation from the verb–object style to any of the other two is connected with lower usefulness perceptions, hence lending further support to hypotheses H2a–H2c. Additionally, based on the data displayed in Table 3 we performed an analysis of variance test implemented in SPSS 16.0 [55] to further examine the differences in the average total factors scores for PU. Between-group differences across the different label styles were statistically significant with $F = 18.495$, $p = 0.000$, thereby confirming our test results.

To test whether there are significant pair-wise differences between the label types (verb–object versus action-noun, verb–object versus rest, and action-noun versus rest), we repeated the ANOVA analysis using the Contrast function [55] to detect pair-wise differences. For perceived usefulness, the contrast between verb–object and action-noun style was significant at $contrastValue = 0.879$, $t = 3.665$, $p = 0.000$, while the contrast between verb–object and rest style was significant $contrastValue = 1.448$, $t = 6.036$, $p = 0.000$. Finally, the contrast between action-noun and rest style was significant at $contrastValue = 0.569$, $t = 2.371$, $p = 0.019$. These results further lend strong support to hypotheses H2a–H2c. In summation, the reported findings support our hypotheses H2a–H2c that verb–object styles are regarded more useful than action-noun styles, and rest styles.

Perceived ambiguity's effect on perceived usefulness: As discussed in the hypothesis development section, our study rests on the assumption that ambiguity of textual labels is an impediment to the perceived usefulness of the label for understanding the process modeled. To test this assumption as specified in hypothesis H3, we once again performed an ANOVA test.

Support for hypothesis H3 exists if there are significant differences in the average total factor scores for perceived usefulness for labels that are either considered ambiguous, or not, with the expectation that the average total factor score will be lower for the group that considered a particular labeling style to be ambiguous. Prior to conduct, ANOVA assumptions were tested and showed no violation. Table 4 provides the results.

Table 4
Average perceived usefulness scores for ambiguous versus unambiguous label types.

	Unambiguous label, $N = 132$		Ambiguous label, $N = 42$		ANOVA	
	Mean	StDev	Mean	StDev	*F*	Sig.
Perceived usefulness	4.538	1.241	3.238	1.495	31.553	0.000

The results displayed in Table 4 confirm our assumption and lend strong support for hypothesis H3. The average total factor score for perceived usefulness was higher for those label types that were not listed as ambiguous by the participants (reported average total factor scores are 4.538 in contrast to 3.238). The ANOVA test showed these differences to be statistically significant at $p = 0.000$.

Moderating effects: As discussed in the demographics section, the participants ranged in terms of their familiarity with the EPC notation used in the process model, as well as in their knowledge of the chosen application domain (complaints handling). More precisely, six participants brought to bear experience with complaints handling domain, and 17 out of 29 participants were above the median in notation familiarity.

Again we first established reliability and validity of the measure "familiarity with the EPC notation". Cronbach's α for the familiarity scale was computed to be 0.914, and composite reliability was computed to be 0.859. Factor loadings for the three familiarity measures were 0.919, 0.930 and 0.931, all significant at $p = 0.000$. Average variance extracted of the familiarity construct was estimated to be 0.927. As described above, AVE also exceeded the squared correlation between the PU and the familiarity construct. Altogether, these result suggest adequate reliability and validity. Appendix B summarizes factor loadings, communalities, and correlations.

In order to test hypotheses H4a and H4b, we examined the differences in the average total factor scores for perceived usefulness of the labels between two sets of two groups of participants (high/low application domain knowledge and high/low familiarity with the EPC notation). Support for the hypotheses would then exist if the differences in the dependent variables between the groups would be significant. We used an analysis of covariance (ANCOVA) test implemented in SPSS 16.0 to test the hypotheses. ANCOVA is an appropriate analysis technique because it allows to control for potential effects of covariates in the examination of dependent variable scores between two treatment groups [55]. ANCOVA assumptions of equal slopes were tested prior to conduct, and showed no violation of normality.

We used two covariates in the analysis of the effect on labeling type on perceived usefulness. The first is the binary variable "Knowledge of the complaints handling domain", which simply establishes the existence of any relevant knowledge in this domain. As a second covariate, we used the median of the total factor score of the three item "Familiarity" scale, to separate the respondents pool in two groups using a dummy variable (high familiarity/low familiarity). Both variables have been described in Section 3.1. Appendix A lists all items used in the questionnaire. We obtained the following results:

- Application domain knowledge does not show a significant interaction effect on the relationship between label type and perceived usefulness ($F = 1.363$, $p = 0.245$, *partial eta square* $= 0.008$). Accordingly, hypothesis H4a must be refuted.
- Notation familiarity does not show a significant interaction effect on the relationship between label type and perceived usefulness ($F = 1.334$, $p = 0.239$, *partial eta square* $= 0.006$). Accordingly, hypothesis H4b must be refuted.

These results are similar to those reported in [11,26], which also did not indicate significant moderation effects of their measures of application domain knowledge or familiarity with the notation on understanding of conceptual models—and contrary to those reported in [38,39], both of which reported some spurious effects on a number of the dependent variables they considered. In the context of the study reported in this paper, the results indicate that understanding of textual labels contained in process models is independent from any expertise gained from previous notation usage or from previous knowledge of the considered domain. In light of the other results presented above, the findings suggest that a label's usefulness is indeed dependent on the grammatical style of the labels itself.

3.3. Discussion

The support for our hypotheses strongly suggests that a verb–object labeling style is rightfully proposed as a preferred way of activity labeling. Indeed, our results indicate strong and favorable perceptions towards a superiority of the verb–object labeling style. Given the key role that usage beliefs (such as perceived ambiguity or perceived usefulness) play in informing actual usage behavior [44,56,57], we deem this finding instrumental to explaining, and supporting, process model understandability. However, whilst process modelers tend to favor verb–object styles, this situation does not necessarily reflect actual usage for activity labeling. In fact, our exploration of the usage frequency of activity labels in the SAP reference model indicates that a large proportion of labels found in practice *cannot* be interpreted as genuine implementations of this style (see Section 2). In contrast, our results indicate that there is wide variety in labeling.

We would argue that this situation can largely be attributed to a lack of operationalized guidance in the proper use of "good" labeling styles (such as the verb–object style). We further argue that the results from our empirical investigations have implications for both research and practice on the quest towards guiding process modelers towards the consequent and consistent usage of labeling styles. In the next section, we will address this issue in more detail.

4. Implications

In this section we highlight some implications of this research. We first discuss implications for research in Section 4.1 and then implications for practice in Section 4.2.

4.1. For research

Our research has strong implications for research into process modeling. While there are some works that describe process models and the information they reveal in a holistic way, e.g., [16], most contributions consider a process model as a structural design artifact. This holds for the whole stream of formal analysis techniques, such as those based on Petri nets. The latter stream has dominated, for example, papers presented at the recent Business Process Management conferences (see [58,59]). In that type of work, textual labels are usually little more than identifiers to the activity concepts in a process model. Our research, in contrast, shows the relevance of labeling for perceived ambiguity and perceived usefulness—that is, to key beliefs informing actual usage behavior. While this does not directly allow a statement on the relative importance of structure and labeling for the pragmatic value of a process model, it suggests that both aspects are complementary and hence deserve closer attention in process modeling research.

Furthermore, in our research we designed a measurement for perceived ambiguity for the textual content of activity labels. It is worth noting that this concept can be equally adapted for investigating the structural elements of a process model. For instance, the formal semantics of OR-join elements have been debated extensively in process modeling research (see e.g., [60,61] for an overview of the discussion). The problem with the OR-join is that it is meant to synchronize only those branches that are still active—which may lead to contradictions when multiple OR-joins wait for each other. Up until now, no generally accepted formalization has been found for this element. It would be an interesting matter for future research to investigate in how far this formal ambiguity materializes in user perceptions on ambiguity.

We see another research implication stemming from the fact that the use of textual labels in conceptual models addresses aspects of linguistics. Through our grammatical analysis we were able to show how some linguistic principles inform information systems practice in the conceptual modeling space. This situation suggests the field of linguistics to be a fruitful reference discipline from which theories, research strategies as well as empirical measurements could be drawn that advance our understanding of the role and practice of conceptual modeling. In our work, we relied in part on existing measurements adopted from information systems research to measure usefulness and ambiguity of textual labels. Clearly, these types of evaluations also pertain to the study of language in general. Future research could examine to what extent knowledge advances from the field of linguistics could inform practices and outcomes in information systems and conceptual modeling. And indeed, a number of scholars have already established that linguistics contributes to informing the body of knowledge around conceptual modeling, e.g., [62–65].

4.2. For practice

In addition to this work's academic merits we also identify a number of implications that pertain directly to process modeling practice. Most notably, our research confirms the suggestion—hitherto largely unreflected—that verb–object styles are an appropriate labeling convention.

In order to lend better support for practitioners working with verb–object labeling styles, it is important to remember that the labeling of graphical model constructs refers to the act of annotating the model construct with information about the intended real-world *domain semantics* that the graphical element is purported to articulate in the model. Domain semantics define the real-world meaning, or essence thereof, of the terms used in any conceptual model, that is, of words and phrases used to label constructs [17]. The delicate part is that some of these semantics are well-known and unambiguous while others may vary with context, i.e., they can be subject to multiple interpretations. Furthermore, the words used to annotate semantics (e.g., the verb and the object in the verb–object labeling style) are typically selected by the model developer, sometimes without a great deal of thought spent on finding the one that best reflects the intended real-world semantics. This can lead to problems when reading and interpreting the model, especially—as often found in modeling practice—model developer and reader audiences do not overlap. This situation is further complicated by the vast amount of terms found in a natural language such as English. For example, the online catalogue WordNet contains over 21,000 different verb word forms alone [35].

In essence, our research results imply, and highlight, a need for a closer integration of process model and structured content. This would have to be supported by process modeling tools. We now discuss this issue in more detail. In particular, in Section 4.2.1 we analyze *label parsing* as a quality measure mechanism to enforce the verb–object style. Sections 4.2.2 and 4.2.3 investigate how controlled vocabularies for *business activities* and *business objects*—the two central terms in the verb–object style—can be supported in process modeling. The key challenge in this area is to aid the process modeler in determining a precise label. In linguistic terms, this challenge closely relates to the problem of synonyms and homonyms. In order to avoid interpretation problems syntactically different terms should be used for referring to semantically distinct verbs or object (avoid synonyms) and syntactically equivalent terms should represent equivalent semantics (avoid homonyms).

4.2.1. *Parsing of labels*

Current approaches to establishing syntactical correctness of process models (e.g., [66–68]) only consider properties of the process graph. The results of our empirical research suggest extending these approaches to also consider the labeling style of activities. This requires a grammatical analysis of the activity labels. The respective parsing can be facilitated in two levels of sophistication. In the simple case, the different words of the text label string are identified using standard programming facilities such as the Java String Tokenizer. Tools like WordNet [35] then check whether the first word is a verb or not, and whether some of the remaining words are nouns. A business process modeling tool can use such an analysis for pointing modelers to labels that do not follow the verb–object style. This approach can be extended by using verb phrase parsing techniques such as [69] to identify the grammatical role of each word in the text label. This way, a process modeling tool can help the user not only to use verb–object style, but also to avoid grammatical errors.

The enforcement of the verb–object style in this way might help to close the gap between natural language and formal language processing. And indeed, the relationship between process models and natural language has been discussed and utilized in various works. In [70] the authors investigate in how far the three steps of building a conceptual model (linguistic analysis, component mapping, and schema construction) can be automated using a model for pre-design. Further text analysis approaches have been used to link activities in process models to document fragments [71] and to compare process models from a semantic perspective [72]. Most beneficiary is the verb–object style for model verbalization and paraphrasing, see [73,74]. Such verbalization is an important step in model and requirements validation [75]. For instance, verb–object style labels can easily be verbalized using the "You have to" prefix which yields natural language sentences like "You have to reject order". This way, verb–object labeling and automatic parsing enables a better validation of process models.

4.2.2. *Controlled object vocabulary*

The parsing of activity labels can then be used to introduce further measures of quality assurance. In this context, it is important that the entities referenced in the labels of the process model relate to relevant concepts of the organization and its environmental context. In research and practice it is widely acknowledged that an agreed set of key terms is an essential prerequisite for modeling business processes [76]. The existence of a repository of such terms and entities, and its integration of the process modeling tool is essential for supporting respective quality measures. In the following paragraph, we discuss two options for integrating such a repository: by modeling and by reuse.

Different authors, e.g., [76], recommend a preparatory step called technical term modeling before modeling the actual process. Technical term models capture the key entities involved in a business process and delineate their hierarchy and semantic relationships. Often, entity-relationship diagrams or class diagrams are employed as a modeling language by practitioners. Some niche tools like Semtalk already support an integration of ontologies and process models for this purpose [77], but major tools such as ARIS or Telelogic System Architect do not.

Instead of modeling from scratch there is also the option to reuse existing data dictionaries and data models. These do not necessarily have to be company-specific. Domain standards and ontologies like the XML Common Business Library (xCBL), or Health Level 7 (HL7) are well suited. Also, some authors, e.g., [78] have suggested general ontological models to guide the act of conceptual modeling by defining key terms and concepts and their relationships. In addition to this work, tools such as the WordNet catalogue can be used to resolve homonym and synonym conflicts between the users data input for the label and the data dictionary. Furthermore, a side effect of this concept could be a better integration of process modeling and semantic business process management (SBPM) technologies [79]. And indeed, previous research has already delivered stimulating work towards a better integration between organizational objects and data concepts, and their role in dynamic processes. Wagner [80], for instance, describes how static, dynamic and deontic aspects of organizations can conceptually be captured on the basis of a set of 19 ontological principles. It is on the basis of work like this where future research can deliver relevant guidance to process modeling practice.

4.2.3. *Controlled verb vocabulary*

While class diagrams and data dictionaries can be easily used to control the object part of verb–object labels, the case is more difficult for verbs. Some work in data modeling has shown that the variety in relationship semantics is much smaller than the potential set of relationship labels [17]. Consequently, Storey argues for, and develops, an ontology for the semantic classification of relationship-type constructs in data models based on dictionaries and business taxonomies. We argue that a similar idea is applicable for activity labeling in process modeling.

Some work is available as a foundation for such an endeavor. The MIT process handbook [20], for instance, discusses a wide range of action terms to be used in business contexts. Building on the lexical database WordNet [35], the MIT handbook defines an inheritance hierarchy that originates from eight generic verbs (viz., *create*, *modify*, *preserve*, *destroy*, *combine*, *separate*, *decide*, *manage*). Verb classifications and verb ontologies have been proposed before. The systematic work by Levin is an important contribution in this area. It defines 49 semantic classes of verbs and categorizes more than 3000 English verbs [81]. A formal approach towards a verb ontology is reported in [82]. Yet, there are several problems, in particular with the classifications of the MIT process handbook and that by Levin when applied to process modeling. Recent research has shown that both schemes cover only a limited amount of verbs found in real-world process model collections (44% and 68%) [83]. Furthermore, neither of the hierarchies is a tree, a problem that stems from synonyms and homonyms.

Clearly, future work is required to address these shortcomings. More precisely, to lend further support to labeling practices, a verb hierarchy should be constructed that defines generic verb terms of pertinence to business process contexts whilst avoiding homonyms and synonyms at all. Such a verb hierarchy could then become an integral part of process modeling tools, as much as thesauri are used in word processing tools. Facilities to extend this hierarchy with domain-specific verbs could be implemented in as simple a way as defining user extensions to the general verb dictionary.

5. Conclusions

This section concludes the paper by summarizing the contributions, the limitations, and by giving an outlook on future research.

5.1. Contributions

In this paper we discussed an essential yet under-researched aspect of process modeling practice, namely that of labeling the graphical activity elements in a process model. In this way, we complement the existing streams of research investigating other dimensions of process modeling (e.g., the data, resource, or control-flow perspectives). Our line of research is based on the assumption that process model understanding can be improved if a more systematic way of labeling constructs can be found. Based on Dual Coding Theory and CTML we argued that understanding can be improved if more consideration is given to the *style* and *choice of terms* for labeling activities in process models. We examined over 600 process models and considered data on the user perceptions of labeling styles to lend support for our arguments. We then explored the implications of our empirical findings and suggested three programs of research towards better, and more stringent, support for process model labeling practices.

5.2. Limitations

Clearly, our research has its limitations and is not yet complete. First, even though we examined a considerably large number of process models, we only considered EPCs of the SAP reference model. This may limit the extent to which our results can be generalized. However, the way we described the design of our experiments, and the inclusion of our data collection instrument (see Appendix A), will allow researchers to replicate our study in other process modeling contexts, e.g., using different model sets, or different process modeling notations such as BPMN or UML Activity Diagrams. It will be an interesting topic of future research to examine whether our findings can be directly transferred to other activity-based process modeling languages, for instance, those that do not explicitly label events (e.g., BPMN, YAWL).

Second, in our study we examined the general labeling style used in process modeling. Clearly, not only the style of labeling but also the specificity of the word items used within these styles (e.g., the actual verb or object terms used in the verb–object style) will have an influence on the quality of the model produced. Future research should thus more closely investigate how the choice or specificity of terms influence process model quality. We outlined some suggestions for such research in Section 4.1.

Third, as with any other research studying perceptual beliefs, our measurement strategy is a potential source of limitations. For some of the aspects (e.g., ambiguity) we considered in our study, we had to develop new measures. In the case of perceived ambiguity, the chosen operationalization only allowed us to calculate Cohen's Kappa as a measure of reliability. In effect, we cannot rule out potential validity issues with this measure. For other aspects (e.g., perceived usefulness, familiarity), in part, we took inspiration from existing measures used in other studies of Information Systems or conceptual modeling phenomena, to propose modified measurement items that we deemed appropriate to our problem context. While reliance on existing measurements in instrument development may be an efficient research practice, it should not be considered superior to rigorously developing new measures [84]. For instance, an alternative to our approach may have been the development of a specialized scale for measuring the understandability of labels, which may have given insights beyond the ones presented in this paper. Furthermore, our approach to *modify* existing measures to make them fit to our research context clearly hampers reliance on their earlier validations. For this reason, we have given careful consideration to testing potential reliability and validity issues (see the analyses reported in Appendix B). While our test results indicate the adequateness of the selected operationalizations, we can imagine that further studies will be useful for a proper reflection on our measurement strategy, and we would like to invite our fellow scholars to join in this endeavor.

5.3. Outlook

Some of the future research streams we consider will be as follows. Aside from seeking to validate our findings on a more general level by considering various other process modeling notations in use today (e.g., BPMN, YAWL), we also aim to examine empirically the suitability of different verb classification schemes for classifying activity tasks in process models. Similar to the experiment described in [17], we will have respondents classify activity tasks in a number of process models as per the verb classification schemes to establish the viability of these schemes.

In a related stream of research we will then aim to establish in an empirical setting whether the consistent usage of the operationalized verb–object style in process models does *in fact* warrant improved model understandability. CTML suggests three outcomes of understanding—retention, recall and transfer—that can be used as measures in a related empirical study. In conducting such a study we can refer to the works of Gemino and Wand [26] and Recker and Dreiling [11] that both used these measures for examining understanding generated through data [26] and process modeling [11], respectively. The empirical results reported in this paper show that label styles have an effect on user perceptions of usefulness, and in our future work we are keen to examine the effect of labeling styles on actual measures of usability, understanding and performance.

Finally, our research into the labeling of graphical elements should lead to specific guidelines that can be effectively used by modelers. Even if a verb–object style of modeling is preferable over other styles, clearly more perspectives, e.g., data, resource, and control-flow, should be considered to create an overall understandable model. Earlier research, for example, has shown the impact of a process model's size, structure, and modularity on its overall understandability [9,85]. Based on such insights, a preliminary, broad set of guidelines is presented in [86]. This so-called 7PMG set includes a guideline on using the verb–object style, as well as guidelines on the number of elements in a model, the application of structured modeling, and the decomposition of a process model. Aside from the challenge to generate guidelines from emerging research on process model quality, a whole new venue of research emerges with respect to establishing the effectiveness of such a set. A potential source of inspiration is the field of use case writing, which we referred to earlier (see Section 2.2). In various papers in this field, experiments are described to assess the impact of modelers' usage of guidelines on the quality of use case descriptions and how alternative guideline sets compare with each other in this respect [27,28,87].

Acknowledgments

The authors thank the anonymous reviewers for their valuable feedback, which improved the paper significantly, as well as the students that filled out the questionnaire.

Appendix A. Questionnaire material used

Demographics

- Gender (male/female)
- Years of tertiary education
- Working experience in process modeling (yes/no)
- Experience in EPC modeling (months)
- Number of EPC models (created/read)
- Training received in EPCs (formal/internal/university/on the job/auto-didact/reading/other)

Familiarity with the EPC notation (*7-point scale from "Strongly disagree" to "Strongly agree"*)

- Overall, I am very familiar with EPCs.
- I feel very confident in understanding process models created with EPCs.
- I feel very competent in using EPCs for process modeling.

Application domain knowledge

- Knowledge of claims handling processes (yes/no)

Perceived ambiguity of a label

- Please list the three function labels from the model that you consider to be the most *ambiguous* ones, i.e., they are most open for alternative interpretations:

Perceived usefulness of a label (*7-point scale from "Strongly disagree" to "Strongly agree"*)

- Overall, I found label *X* useful for understanding the process modeled.
- Overall, I think label *X* improves my performance when understanding the process modeled.

Appendix B. Reliability and validity results (Table B.1)

Table B.1

	Perceived usefulness		Notation familiarity	
	Factor loadings	Communalities	Factor loadings	Communalities
PU1	0.936	0.877		
PU2	0.936	0.877		
Fam1			0.919	0.845
Fam2			0.930	0.866
Fam3			0.931	0.867
Cronbach's α	0.857		0.914	
Composite reliability	0.884		0.868	
Average variance extracted	0.936		0.927	
	1.000		0.030	
Correlation	−0.030		1.000	

References

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

[2] M. Dumas, W.M.P. van der Aalst, A.H.M. ter Hofstede (Eds.), Process Aware Information Systems: Bridging People and Software Through Process Technology, Wiley, Hoboken, NJ, 2005.

[3] T. Erl, Service-oriented Architecture: Concepts, Technology, and Design, Prentice-Hall, Upper Saddle River, NJ, 2005.
[4] C. Ferris, What are web services?, Communications of the ACM 46 (6) (2003) 31–32.
[5] B. Curtis, M.I. Kellner, J. Over, Process modeling, Communications of the ACM 35 (9) (1992) 75–90.
[6] A.-W. Scheer, ARIS—Business Process Modeling, third ed., Springer, Berlin, Germany, 2000.
[7] J. Dehnert, W.M.P. van der Aalst, Bridging the gap between business models and workflow specifications, International Journal of Cooperative Information Systems 13 (3) (2004) 289–332.
[8] D.L. Moody, Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions, Data and Knowledge Engineering 15 (3) (2005) 243–276.
[9] J. Mendling, H.A. Reijers, J. Cardoso, What makes process models understandable?, in: G. Alonso, P. Dadam, M. Rosemann (Eds.), Business Process Management—BPM 2007, Lecture Notes in Computer Science, vol. 4714, Springer, Brisbane, Australia, 2007, pp. 48–63.
[10] J. Mendling, G. Neumann, W.M.P. van der Aalst, Understanding the occurrence of errors in process models based on metrics, in: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2007, Lecture Notes in Computer Science, vol. 4803, Springer, Vilamoura, Portugal, 2007, pp. 113–130.
[11] J. Recker, A. Dreiling, Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education, in: M. Toleman, A. Cater-Steel, D. Roberts (Eds.), 18th Australasian Conference on Information Systems, The University of Southern Queensland, Toowoomba, Australia, 2007, pp. 356–366.
[12] M. Rosemann, J. Recker, M. Indulska, P. Green, A study of the evolution of the representational capabilities of process modeling grammars, in: E. Dubois, K. Pohl (Eds.), Advanced Information Systems Engineering—CAiSE 2006, Lecture Notes in Computer Science, vol. 4001, Springer, Luxembourg, Grand-Duchy of Luxembourg, 2006, pp. 447–461.
[13] W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, A.P. Barros, Workflow patterns, Distributed and Parallel Databases 14 (1) (2003) 5–51.
[14] N. Russell, W.M.P. van der Aalst, A.H.M. ter Hofstede, D. Edmond, Workflow resource patterns: identification, in: O. Pastor, J. Falcao e Cunha (Eds.), Advanced Information Systems Engineering—CAiSE 2005, Lecture Notes in Computer Science, vol. 3520, Springer, Porto, Portugal, 2005, pp. 216–232.
[15] N. Russell, A.H.M. ter Hofstede, D. Edmond, W.M.P. van der Aalst, Workflow data patterns: identification, representation and tool support, in: L.M.L. Delcambre, C. Kop, H.C. Mayr, J. Mylopoulos, O. Pastor (Eds.), Conceptual Modeling—ER 2005, Lecture Notes in Computer Science, vol. 3716, Springer, Klagenfurt, Austria, 2005, pp. 353–368.
[16] J. Krogstie, G. Sindre, H.D. Jorgensen, Process models representing knowledge for action: a revised quality framework, European Journal of Information Systems 15 (1) (2006) 91–102.
[17] V.C. Storey, Comparing relationships in conceptual modeling: mapping to semantic classifications, IEEE Transactions on Knowledge and Data Engineering 17 (11) (2005) 1478–1489.
[18] L.D. Miles, Techniques of Value Analysis and Engineering, McGraw-Hill, New York, 1961.
[19] A. Sharp, P. McDermott, Workflow Modeling: Tools for Process Improvement and Application Development, Artech House, Boston, MA, 2001.
[20] T.W. Malone, K. Crowston, G.A. Herman, Organizing Business Knowledge: The MIT Process Handbook, The MIT Press, Cambridge, MA, 2003.
[21] A. Paivio, Dual coding theory: retrospect and current status, Canadian Journal of Psychology 45 (3) (1991) 255–287.
[22] L.R. Brooks, The suppression of visualization by reading, The Quarterly Journal of Experimental Psychology 19 (4) (1967) 289–299.
[23] R.E. Mayer, Models for understanding, Review of Educational Research 59 (1) (1989) 43–64.
[24] R.E. Mayer, Multimedia Learning, Cambridge University Press, Cambridge, MA, 2001.
[25] F. Bodart, A. Patel, M. Sim, R. Weber, Should optional properties be used in conceptual modelling? A theory and three empirical tests, Information Systems Research 12 (4) (2001) 384–405.
[26] A. Gemino, Y. Wand, Complexity and clarity in conceptual modeling: comparison of mandatory and optional properties, Data and Knowledge Engineering 55 (3) (2005) 301–326.
[27] K. Cox, K.T. Phalp, Replicating the crews use case authoring guidelines experiment, Empirical Software Engineering 5 (3) (2000) 245–267.
[28] K.T. Phalp, J. Vincent, K. Cox, Improving the quality of use case descriptions: empirical assessment of writing guidelines, Software Quality Journal 15 (4) (2007) 383–399.
[29] R.B. Davis, Business Process Modelling with ARIS: A Practical Guide, Springer, London, England, 2001.
[30] J. Mendling, M. Strembeck, Influence factors of understanding business process models, in: W. Abramowicz, D. Fensel (Eds.), Business Information Systems—BIS 2008, Lecture Notes in Business Information Processing, vol. 7, Springer, Innsbruck, Austria, 2008, pp. 142–153.
[31] G. Keller, T. Teufel, SAP R/3 Process Oriented Implementation: Iterative Process Prototyping, Addison-Wesley, Boston, MA, 1998.
[32] M. Daneva, Erp requirements engineering practice: lessons learned, IEEE Software 21 (2) (2004) 26–33.
[33] J. Mendling, H.M.V. Verbeek, B.F. van Dongen, W.M.P. van der Aalst, G. Neumann, Detection and prediction of errors in epcs of the sap reference model, Data and Knowledge Engineering 64 (1) (2008) 312–329.
[34] A. Dreiling, M. Rosemann, W.M.P. van der Aalst, W. Sadiq, From conceptual process models to running systems: a holistic approach for the configuration of enterprise system processes, Decision Support Systems 45 (2) (2008) 189–207.
[35] G.A. Miller, Wordnet—a lexical database for English, Communications of the ACM 38 (11) (1995) 39–41.
[36] R.M.W. Dixon, Deriving verbs in English, Language Sciences 30 (1) (2008) 31–51.
[37] R. Price, G. Shanks, A semiotic information quality framework: development and comparative analysis, Journal of Information Technology 20 (2) (2005) 88–102.
[38] V. Khatri, I. Vessey, P.C.V. Ramesh, P. Sung-Jin, Understanding conceptual schemas: exploring the role of application and is domain knowledge, Information Systems Research 17 (1) (2006) 81–99.
[39] K. Masri, D.C. Parker, A. Gemino, Using iconic graphics in entity-relationship diagrams: the impact on understanding, Journal of Database Management 19 (3) (2008) 22–41.
[40] D. Batra, J.G. Davis, Conceptual data modelling in database design: similarities and differences between expert and novice designers, International Journal of Man–Machine Studies 37 (1) (1992) 83–101.
[41] G. Shanks, Conceptual data modelling: an empirical study of expert and novice data modellers, Australasian Journal of Information Systems 4 (2) (1997) 63–73.
[42] J. Recker, Understanding process modelling grammar continuance: a study of the consequences of representational capabilities, Ph.D Thesis, Queensland University of Technology, 2008.
[43] A. Maes, G. Poels, Evaluating quality of conceptual modelling scripts based on user perceptions, Data and Knowledge Engineering 63 (3) (2007) 769–792.
[44] A. Bhattacherjee, Understanding information systems continuance: an expectation-confirmation model, MIS Quarterly 25 (3) (2001) 351–370.
[45] J. Mendling, Detection and prediction of errors in epc business process models, Ph.D. Thesis, Vienna University of Economics and Business Administration, 2007.
[46] A. Gemino, Y. Wand, A framework for empirical evaluation of conceptual modeling techniques, Requirements Engineering 9 (4) (2004) 248–260.
[47] D. Batra, J.A. Hoffler, R.P. Bostrom, Comparing representations with relational and eer models, Communications of the ACM 33 (2) (1990) 126–139.
[48] K. Siau, P.-P. Loo, Identifying difficulties in learning uml, Information Systems Management 23 (3) (2006) 43–51.
[49] J. Parsons, L. Cole, What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques, Data and Knowledge Engineering 55 (3) (2005) 327–342.
[50] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1) (1960) 37–46.
[51] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (2) (1977) 159–174.
[52] S. Siegel, Nonparametric Statistics for The Behavioral Sciences, McGraw-Hill Series in Psychology, McGraw-Hill Kogakusha, Tokyo, Japan, 1956.
[53] J.C. Nunnally, I.H. Bernstein, Psychometric Theory, McGraw-Hill Series in Psychology, third ed., McGraw-Hill, New York, 1994.

[54] C. Fornell, D.F. Larcker, Evaluating structural equations with unobservable variables and measurement error, Journal of Marketing Research 18 (1) (1981) 39–50.
[55] J.P. Stevens, Applied Multivariate Statistics for the Social Sciences, fourth ed., Applied Multivariate STATS, Lawrence Erlbaum Associates, Hillsdale, NJ, 2001.
[56] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly 13 (3) (1989) 319–340.
[57] G.C. Moore, I. Benbasat, Development of an instrument to measure the perceptions of adopting an information technology innovation, Information Systems Research 2 (3) (1991) 192–222.
[58] G. Alonso, M. Rosemann, P. Dadam (Eds.), Business Process Management—BPM 2007, Lecture Notes in Computer Science, vol. 4714, Springer, Brisbane, Australia, 2007.
[59] M. Dumas, M. Reichert, M.-C. Shan (Eds.), Business Process Management—BPM 2008, Lecture Notes in Computer Science, vol. 5240, Springer, Milan, Italy, 2008.
[60] E. Kindler, On the semantics of epcs: resolving the vicious circle, Data and Knowledge Engineering 56 (1) (2005) 23–40.
[61] J. Mendling, in: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction and Guidelines for Correctness, Lecture Notes in Business Information Processing, vol. 6, Springer, Berlin, Germany, 2008.
[62] D.C. Sutton, Linguistic problems with requirements and knowledge elicitation, Requirements Engineering 5 (2) (2000) 114–124.
[63] N. Guarino, Concepts, attributes, and arbitrary relations. Some linguistic and ontological criteria for structuring knowledge bases, Data and Knowledge Engineering 8 (3) (1992) 249–261.
[64] J.F.M. Burg, R.P. van de Riet, The impact of linguistics on conceptual models: consistency and understandability, Data and Knowledge Engineering 21 (2) (1997) 131–146.
[65] P.J. Agerfalk, O. Eriksson, Action-oriented conceptual modelling, European Journal of Information Systems 13 (1) (2004) 80–92.
[66] B. Kiepuszewski, A.H.M. ter Hofstede, W.M.P. van der Aalst, Fundamentals of control flow in workflows, Acta Informatica 39 (3) (2003) 143–209.
[67] D. Gasevic, V. Devedzic, Petri net ontology, Knowledge-Based Systems 19 (4) (2006) 220–234.
[68] J. Mendling, W.M.P. van der Aalst, Formalization and verification of epcs with or-joins based on state and context, in: J. Krogstie, A.L. Opdahl, G. Sindre (Eds.), Advanced Information Systems Engineering—CAiSE 2007, Lecture Notes in Computer Science, vol. 4495, Springer, Trondheim, Norway, 2007, pp. 439–453.
[69] D. Hardt, Verb phrase ellipsis: form, meaning, and processing, Ph.D. Thesis, University of Pennsylvania, 1993.
[70] G. Fliedl, C. Kop, H.C. Mayr, From textual scenarios to a conceptual schema, Data and Knowledge Engineering 55 (1) (2005) 20–37.
[71] J.E. Ingvaldsen, J.A. Gulla, X. Su, H. Ronneberg, A text mining approach to integrating business. Process models and governing documents, in: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2005: OTM Workshops, Lecture Notes in Computer Science, vol. 3762, Springer, Agia Napa, Cyprus, 2005, pp. 473–484.
[72] M. Ehrig, A. Koschmider, A. Oberweis, Measuring similarity between semantic business process models, in: J.F. Roddick, A. Hinze (Eds.), Conceptual Modelling 2007, Conferences in Research and Practice in Information Technology, vol. 67, Australian Computer Science Communications, Ballarat, Australia, 2007, pp. 71–80.
[73] T. Halpin, M. Curland, Automated verbalization for orm 2, in: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2006: OTM year Workshops, Lecture Notes in Computer Science, vol. 4278, Springer, Montpellier, France, 2006, pp. 1181–1190.
[74] P.J.M. Frederiks, T.P. van der Weide, Information modeling: the process and the required competencies of its participants, Data and Knowledge Engineering 58 (1) (2006) 4–20.
[75] B. Nuseibeh, S. Easterbrook, Requirements engineering: a roadmap, in: C. Ghezzi, M. Jazayeri, A.L. Wolf (Eds.), 22nd International Conference on Software Engineering, ACM, Limerick, Ireland, 2000, pp. 35–46.
[76] M. Rosemann, Preparation of process modeling, in: J. Becker, M. Kugeler, M. Rosemann (Eds.), Process Management: A Guide for the Design of Business Processes, Springer, Berlin, Germany, 2003, pp. 41–78.
[77] C. Fillies, G. Wood-Albrecht, F. Weichardt, Pragmatic applications of the semantic web using semtalk, Computer Networks 42 (5) (2003) 599–615.
[78] G. Guizzardi, H. Herre, G. Wagner, On the general ontological foundations of conceptual modeling, in: S. Spaccapietra, S.T. March, Y. Kambayashi (Eds.), Conceptual Modeling—ER 2002, Lecture Notes in Computer Science, vol. 2503, Springer, Tampere, FL, 2002, pp. 65–78.
[79] M. Hepp, F. Leymann, C. Bussler, J. Domingue, A. Wahler, D. Fensel, Semantic business process management: using semantic web services for business process management, in: IEEE International Conference on e-Business Engineering, IEEE, Beijing, China, 2005, pp. 535–540.
[80] G. Wagner, The agent-object-relationship metamodel: towards a unified view of state and behavior, Information Systems 28 (5) (2003) 475–504.
[81] B. Levin, English Verb Classes and Alternations: A Preliminary Investigation, University of Chicago Press, Chicago, IL, 1993.
[82] C.P. Menzel, M. Grüninger, A formal foundation for process modeling, in: N. Guarino, B. Smith, C. Welty (Eds.), 2nd International Conference on Formal Ontology in Information Systems, ACM, Oguinquit, Maine, 2001, pp. 256–269.
[83] J. Mendling, J. Recker, Towards systematic usage of labels and icons in business process models, in: T. Halpin, H.A. Proper, J. Krogstie (Eds.), 12th International Workshop on Exploring Modeling Methods in Systems Analysis and Design, CEUR Workshop Proceedings Series, CEUR, Montpellier, France, 2008, pp. 1–13.
[84] M. Boudreau, D. Gefen, D. Straub, Validation in information systems research: a state-of-the-art assessment, MIS Quarterly 25 (1) (2001) 1–16.
[85] H.A. Reijers, J. Mendling, Modularity in process models: review and effects, in: M. Dumas, M. Reichert, M.-C. Shan (Eds.), Business Process Management—BPM 2008, Lecture Notes in Computer Science, vol. 5240, Springer, Milan, Italy, 2008, pp. 20–35.
[86] J. Mendling, H.A. Reijers, W.M.P. van der Aalst, Seven process modeling guidelines (7pmg), QUT ePrints Report 12340, Queensland University of Technology, 2008.
[87] C.B. Achour, C. Rolland, N.A.M. Maiden, C. Souveyet, Guiding use case authoring: results of an empirical study, in: IEEE International Symposium on Requirements Engineering, IEEE, Limerick, Ireland, 1999, pp. 36–43.