

# A Study into the Factors that Influence the Understandability of Business Process Models

Hajo A. Reijers and Jan Mendling

**Abstract**—Business process models are key artifacts in the development of information systems. While one of their main purposes is to facilitate communication among stakeholders, little is known about the factors that influence their comprehension by human agents. On the basis of a sound theoretical foundation, this paper presents a study into these factors. Specifically, the effects of both *personal* and *model* factors are investigated. Using a questionnaire, students from three different universities evaluated a set of realistic process models. Our findings are that both types of investigated factors affect model understanding, while personal factors seem to be the more important of the two. The results have been validated in a replication that involves professional modelers.

**Index Terms**—Business Process Modeling, Process models, Human information processing, Complexity measures.

## I. INTRODUCTION

SINCE the 1960s, conceptual models are in use to facilitate the early detection and correction of system development errors [1]. In more recent years, the primary focus of conceptual modeling efforts has shifted to *business processes* [2]. Models resulting from such efforts are commonly referred to as *business process models*, or *process models* for short. They are used to support the analysis and design of, for example, process-aware information systems [3], service-oriented architectures [4], and web services [5].

Process models typically capture in some graphical notation what tasks, events, states, and control flow logic constitute a business process. A business process that is in place to deal with complaints may, for example, contain a task to evaluate the complaint, which is followed by another one specifying that the customer in question is to be contacted. Similar to other conceptual models, process models are first and foremost required to be intuitive and easily understandable, especially in IS project phases that are concerned with requirements documentation and communication [6]. Today, many companies design and maintain several thousand process models, often also involving non-expert modelers [7]. It has been observed that such large model collections exhibit serious quality issues in industry practice [8].

Against this background it is problematic that little insights exist into what influences the quality of process models, in particular with respect to their *understandability*. The most important insight is that the *size* of the model is of notable

impact. An empirical study provides evidence that larger, real-world process models tend to have more formal flaws (such as e.g. deadlocks or unreachable end states) than smaller models [9]. A likely explanation for this phenomenon would be that human modelers loose track of the interrelations in large and complex models due to their limited cognitive capabilities (cf. [10]). They then introduce errors that they would not insert in a small model, which will make the model less effective for communication purposes.

There is both an academic and a practical motivation to look beyond this insight. To start with the former, it can be virtually ruled out that size is the sole factor that plays a role in understanding a process model. To illustrate, a purely sequential model will be easier to understand than a model that is similar in size but where tasks are interrelated in various ways. This raises the interest into the other factors that play a role here. A more practical motivation is that model size is often determined by the modeling domain or context. So, process modelers will find it difficult to affect the size metric of a process model towards creating a better readable version: They cannot simply skip relevant parts from the model.

The aim of this paper is to investigate whether factors can be determined *beyond* the size of process model that influence its understanding. In that respect, we distinguish between model factors and personal factors. *Model factors* relate to the process model itself and refer to characteristics such as a model's density or structuredness. *Personal factors* relate to the reader of such a model, for example with respect to one's educational background or the perceptions that are held about a process model. While insights are missing into the impact of any of such factors – beyond size – on *process* model understandability, research has suggested the importance of similar factors in other conceptual *data* models [11], [12].

To investigate the impact of personal and model factors, the research that is presented here takes on a survey design. Using a questionnaire that has been filled out by 73 students from three different universities, hypothetical relations between model and personal factors on the understanding of a set of process models are investigated. Some exploratory findings from this data are reported in [13], [14], which essentially confirm the significance of the two types of factors. The contribution of this paper is quite different. We develop a sound theoretical foundation for discussing individual model understanding that is rooted in cognitive research on computer program comprehension. We use this foundation to establish hypotheses on the connection between understandability on the one hand, and personal and model factors on the other hand. For these tests we use the before mentioned survey

Hajo A. Reijers is with the School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: h.a.reijers@tue.nl.

Jan Mendling is with the School of Business and Economics, Humboldt-Universität zu Berlin, Germany, e-mail jan.mendling@wiwi.hu-berlin.de

Manuscript submitted April 19, 2009.

data. Beyond that, we provide an extensive validation of our findings and our instruments addressing two major challenges. First, there has been little research on construct validity of understandability measures. We use Cronbach’s alpha to check the consistency of our questions that are used in calculating the understandability score. Furthermore, we address potential threats to external validity. We report on a replication of our survey with practitioners and investigate if the results differ from that of the students.

The rest of this paper is organized in accordance with the presented aims. Section II introduces the theoretical foundations of process model understanding. We identify matters of process model understanding and respective challenges. This leads us to factors of understanding. Section III describes the setup of our survey design and the motivations behind it. Section IV then presents the data analysis and the interpretation. Section V discusses threats to validity and how we addressed them. Section VI concludes the article. We use Appendix A to summarize our survey design.

## II. BACKGROUND

This section introduces the theoretical background of our empirical research. Section II-A gives a brief overview of the information content of a process model, defines a notion of understandability, and summarizes related work on process model understanding. Section II-B investigates potential factors of understandability. We utilize insights from cognitive research into computer program comprehension in order to derive propositions about the significance of personal and model factors for understanding.

### A. Matters of Process Model Understanding

Before considering a notion of understandability we first have to discuss matters that can be understood from a process model. We are focusing on so-called activity-based or control-flow-based process models (in contrast to goal-oriented [15] and choreography-oriented languages [16]). Figure 1 shows an example of such a process model in a notation that we will use throughout this paper. This notation essentially covers the commonalities of Event-driven Process Chains (EPCs) [17], [18] and the Business Process Modeling Notation (BPMN) [19], which are two of the most frequently used notations for process modeling. Such a process model describes the control flow between different activities (A, B, I, J, K, L, M, N, and O in Figure 1) using arcs. So-called connectors (XOR, AND, OR) define complex routing constraints of splits (multiple outgoing arcs) and joins (multiple ingoing arcs). XOR-splits represent exclusive choices and XOR-joins capture respective merges without synchronization. AND-splits introduce concurrency of all outgoing branches while AND-joins synchronize all incoming arcs. OR-splits define inclusive choices of a 1-to-all fashion. OR-joins synchronize such multiple choices, which requires a quite sophisticated implementation (see [18], [20]). Furthermore, there are specific nodes to indicate start and end.

In this paper we consider formal statements that can be derived about the behavior described by such a process model, ignoring the (informal) content of activity labels. This formal

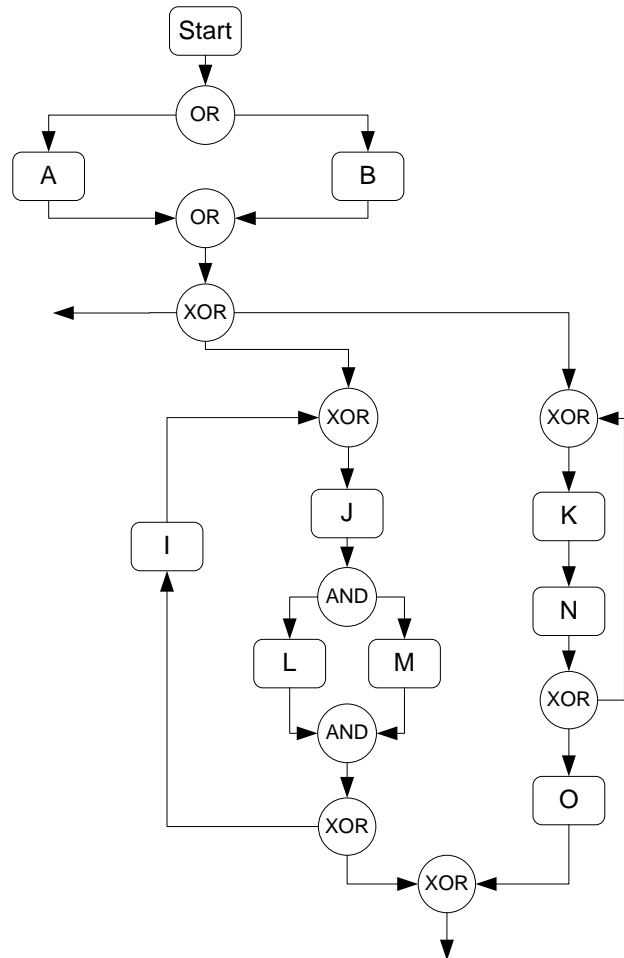


Fig. 1. Part of a process model

focus has the advantage that we can unambiguously evaluate whether an individual has grasped a particular model aspect correctly. In particular, we focus on binary relationships between two activities in terms of execution order, exclusiveness, concurrency, and repetition. These relationships play an important role for reading, modifying, and validating the model.

- *Execution Order* relates to whether the execution of one activity  $a_i$  eventually leads the execution of another activity  $a_j$ . In Figure 1, the execution of J leads to the execution of L.
- *Exclusiveness* means that two activities  $a_i$  and  $a_j$  can never be executed in the same process instance. In Figure 1, J and K are mutually exclusive.
- The *concurrency* relation covers two activities  $a_i$  and  $a_j$  if they can potentially be executed concurrently. In Figure 1, L and M are concurrent.
- A single activity  $a$  is called *repeatable* if it is possible to execute it more than once for a process instance. In Figure 1, among others, K, N, and I are repeatable.

Statements such as “Executing activity  $a_i$  implies that  $a_j$  will be executed later” can be easily verified using the reachability graph of the process model. A reachability graph captures all states and transitions represented by the process

model and it can be (automatically) generated from it. For some classes of models, several relationships can be calculated more efficiently without the reachability graph [21]. For instance, these relations can be constructed for those process models that map to free-choice Petri nets in  $O(n^3)$  time [22], [23].

### B. Factors of Process Model Understanding

Throughout this paper, we use the term *understandability* in order to refer to the degree to which information contained in a process model can be easily understood by a reader of that model. This definition already implies that understandability can be investigated from two major angles: personal factors related to the model reader and factors that relate to the model itself. We discuss the relevance of both categories using the cognitive dimensions framework as a theoretical foundation.

The cognitive dimensions framework is a set of aspects that have empirically been proven to be significant for the comprehension of computer programs and visual notations [24]. There are two major findings that the framework builds upon: A representation always emphasizes a certain information at the expense of another one, and there has to be a fit between the mental task at hand and the notation [25], [26]. The implications of these insights are reflected by cognitive dimensions that are relevant for process model reading.

- *Abstraction Gradient* refers to the grouping capabilities of a notation. In a single process model, there is no mechanism to group activities. Therefore, flow languages are called abstraction-hating [24]. As a consequence, the more complex the model gets the more difficult it becomes for the model reader to identify those parts that closely relate. Presumably, expert model readers will be more efficient in finding the related parts.
- *Hard Mental Operations* relates to an over-proportional increase in difficulty when elements are added to a representation. This is indeed the case for the behavioral semantics of a process model. In the general case, calculating the reachability graph for a process model is NP-complete [27]. Therefore, a larger process model is over-proportionally more difficult to interpret than a simple model. On the other hand, experts are more likely to know decomposition strategies, e.g. as described in [28], to cope with complexity.
- *Hidden Dependencies* refer to interdependencies that are not fully visible. In process models such hidden dependencies exist between split and join connectors: each split should have a matching join connector of the same type, e.g. to synchronize concurrent paths. In complex models, distant split-join pairs can be quite difficult to trace. In general such interdependencies can be analyzed using the reachability graph, but many analyses can be performed also structurally (see [29]). Experts modelers tend to use structural heuristics for investigating the behavior of a process model.
- *Secondary Notation* refers to any piece of extra information that is not part of the formalism. In process models secondary notation is an important matter, among others

in terms of labeling conventions [30] or layout strategies [31]. For models of increasing complexity, secondary notation also gains in importance for making the hidden dependencies better visible. On the other hand, it has been shown that experts' performance is less dependent on secondary notation as that of novices [32].

Personal factors have also been recognized as important factors in engineering and design [33], [34]. In particular, the matter of expertise is clearly established by prior research on human-computer interaction. While research on perceptual quality and perceptual expertise is only emerging recently in conceptual modeling (see [35], [36]), there are some strong insights into the factors of expert performance in different areas. A level of professional expertise is assumed to take at least 1,000 to 5,000 hours of continuous training [37, p.563]. In this context, it is not only important that the expert has worked on a certain matter for years, but also that practicing has taken place on a daily basis [38]. Such regular training is needed to build up experience, knowledge, and the ability to recognize patterns [39]. Furthermore, the way information is processed by humans is influenced by cognitive styles, which can be related to personality. There are persons who prefer verbal over image information and who rather grasp the whole instead of analytically decomposing a matter, or the other way round [40]. As models enable reasoning through visualization, perceptual capabilities of a person are also relevant [41]. Clearly, these capabilities differ between persons with different process modeling expertise.

We conclude for this theoretical discussion that model features and personal characteristics are indeed likely to be relevant factors of process model understandability.

### C. Related work

In this section we present related work grouped into three categories: model factors, personal factors, and other factors.

The importance of model characteristics was intuitively assumed by early work into process model metrics. Such metrics quantify structural properties of a process model, inspired by prior work in software engineering on lines of code, cyclomatic number, or object-oriented metrics [42]–[44]. Early contributions by Lee and Yoon, Nissen, and Morasca [45]–[47] focus on defining metrics. More recently, different metrics have been validated empirically. The work of Cardoso is centered around an adaptation of the cyclomatic number for business processes he calls *control-flow complexity (CFC)* [48]. This metric was validated with respect to its correlation with perceived complexity of process models [49]. The research conducted by a group including Canfora, Rolón, and García analyzes understandability as an aspect of maintainability. They include different metrics of size, complexity, and coupling in a set of experiments, and identify several correlations [50], [51]. Further metrics take their motivation from cognitive research, e.g. [14], and based on concepts of modularity, e.g. [52], [53]. Most notably, an extensive set of metrics has been validated as factors of error probability [9], a symptom of bad understanding. The different validations clearly show that size is an important model factor for understandability, but does

not fully determine phenomena of understanding: additional metrics like structuredness help to improve the explanatory power significantly [18].

Personal factors have been less intensively researched as factors of process model understanding. The experiment by Recker and Dreiling operationalizes the notion of process modeling expertise by a level of familiarity of a particular modeling notation [54]. In a survey by Mendling, Reijers, and Cardoso participants are characterized based on the number of process models they created and the years of modeling experience they have [13]. Mendling and Strembeck measure theoretical knowledge of the participants in another survey using six yes/no questions [55]. Most notable are two results that point to the importance of theoretical process modeling knowledge. In the Mendling, Reijers, and Cardoso survey the participants from TU Eindhoven with strong Petri net education scored best and in the Mendling and Strembeck survey, there was a high correlation between theoretical knowledge and the understandability score.

There are other factors that also might have an impact on process model understanding. We briefly discuss model purpose, problem domain, modeling notation, and visual layout.

*Model purpose:* The understanding of a model may be affected by the specific purpose the modeler had in mind. The best example is that some process models are not intended to be used on a day-to-day basis by people but instead are explicitly created for automatic enactment. In such a case, less care will be given to make them comprehensible to humans. The differences between process models as a result of different modeling purposes are mentioned, for example, in [6]. Empirical research into this factor is missing.

*Problem domain:* People may find it easier to read a model about the domain they are familiar with than other models. While this has not been established for process models, it is known from software engineering that domain knowledge affects the understanding of particular code [56].

*Modeling notation:* In the presence of many different notations for process models, e.g. UML Activity diagrams, EPCs, BPMN, and Petri nets, it cannot be ruled out that some of these are inherently more suitable to convey meaning to people than others. Empirical research that has explored this difference is, for example, reported in [57]. According to these publications, the impact of the notation being used is not very high, maybe because the languages are too similar. Other research that compares notations of a different focus identify a significant impact on understanding [58], [59].

*Visual layout:* Semantically equivalent models can be arranged in different ways. For example, different line drawing algorithms can be used or models may be split up into different submodels. The effect of layout on process model understanding was already noted in the early 1990s [60]. With respect to graphs, it is a well-known result that edge crosses negatively affect understanding [61]. Also, for process models, the use of modularity can improve understanding [62].

Given that, as we argued, the insights into the understanding of process models are limited, this is probably not a complete set of factors. But even at this number, it would be difficult to investigate them all together. In this study, we restrict ourselves

to the first two categories, i.e. personal and model factors. In the definition of this survey, which will be explained in the next section, we will discuss how we aimed to neutralize the potential effects of the other categories.

#### D. Summary

From cognitive research into program understanding we can expect that personal and model factors are likely to be factors of process model understandability. The impact of size as an important metric has been established by prior research. Yet, it only partially explains phenomena of understanding. Personal factors also appear to be relevant. Theoretical knowledge was found to be a significant factor, but so far only in student experiments. Furthermore, research into the relative importance of personal and model factors are missing. In the following sections, we present a survey to investigate this question and analyze threats to validity.

### III. DEFINITION, PLANNING, AND OPERATION OF THE SURVEY DESIGN

This section explains the definition, the planning and the operation of a survey design in personal and model related factors of understanding.

#### A. Definition

According to the theoretical background we provided, both the characteristics of the *reader* of a process model and those of the *process model* itself affect the understanding that such a reader may gain from studying that model. Both types of characteristics can be considered as independent variables, while the understanding gained from studying a process model constitutes the dependent variable. Beyond this, there are other potential factors of influence which we wish to neutralize, i.e. model purpose, problem domain, modeling notation, and visual layout (see Section II-C). To explore the relations that interest us, the idea is to expose a group of respondents to a set of process models and then test their understanding of these models using a self-administered questionnaire. Such a design shares characteristics with a survey where personal and model parameters are recorded, but without predefined factor levels. We use a convenience sample of students. From the analysis perspective it can be classified as a correlational study that seeks to identify statistical connections between measurements of interest. Similar designs have been used to investigate metrics in software engineering, e.g. in [63]. Conclusions on causality are limited, though.

We strived to neutralize the influence of other relevant factors. First of all, a set of process models from practice was gathered that was specifically developed for documentation *purposes*. Next, to eliminate the influence of *domain* knowledge all the task labels in these process models were replaced by neutral identifiers, i.e. capital letters *A* to *W*. In this way, we also prevent a potential bias stemming from varying length of natural activity label (see [55]). Based on the observation by [57] that EPCs appear to be easier to understand than Petri nets, we chose for an EPC-like *notation* without events.

The participants received a short informal description of the semantics similar to [64, p. 25]. Finally, all models were graphically presented on one page, without use of modularity, and drawn in the same top-to-bottom *layout* with the start element at the top and end element at the bottom.

Furthermore, in our exploration we wish to exclude one particular process model characteristic, which is *size*. As we argued in the introduction of this paper and our discussion of related work, process model size is the one model characteristic of which its impact on both error proneness and understanding is reasonably well understood. Because it is our purpose to look beyond the impact of this particular aspect, we have controlled the number of tasks in the process model. Each of the included process models has the same number of tasks. However, to allow for variation across the other model characteristics, two additional variants were constructed for each of the real process models. The variations were established by changing one or two routing elements in each of these models (e.g. a particular XOR-split in a AND-split).

Having taken care of the various factors we wish to control, at this point we can refine what personal and model factors are taken into account and how these are measured in the questionnaire. Note that a summary of the questionnaire is presented in Appendix A.

For the personal factors, we take the following variables into consideration:

- **THEORY:** A person's theoretical knowledge on process modeling. This variable is measured as a self-assessment by the respondents on a 5-point ordinal scale, with anchor points "I have weak theoretical knowledge" and "I have strong theoretical knowledge".
- **PRACTICE:** A person's practical experience with process modeling. This variable is a self-assessment by the respondents. It is measured on a 4-point ordinal scale. The scale has anchor points "I never use business process modeling in practice" and "I use business process modeling in practice every day".
- **EDUCATION:** A person's educational background. This categorical variable refers to the educational institute that the respondents is registered at.

For the model factors, several variables are included. These variables are all formally defined in [18, pp. 117-128], with the exception of the cross-connectivity metric that is specified in [14]. The model factors can be characterized as follows:

- **#NODES, #ARCS, #TASKS, #CONNECTORS, #AND-SPLITS, #AND-JOINS, #XOR-SPLITS, #XOR-JOINS, #OR-SPLITS, #OR-JOINS:** These variables all relate to the number of a particular type of elements in a process model. These include counts for the number of arcs (**#ARCS**) and nodes (**#NODES**). The latter can be further subdivided into **#TASKS** on the one hand and **#CONNECTORS** on the other. The most specific counts are subcategorizations of the different types of logical connectors, like **#AND-SPLITS** and **#OR-JOINS**.
- **DIAMETER:** The length of the longest path from a start node to an end node in the process model.
- **TOKEN SPLITS:** The maximum number of paths in a

process model that may be concurrently initiated through the use of AND-splits and OR-splits.

- **AVERAGE CONNECTOR DEGREE, MAXIMUM CONNECTOR DEGREE:** The **AVERAGE CONNECTOR DEGREE** expresses the average of the number of both incoming and outgoing arcs of the connector nodes in the process model; the **MAXIMUM CONNECTOR DEGREE** expresses the maximum sum of incoming and outgoing arcs of those connector nodes.
- **CONTROL FLOW COMPLEXITY:** A weighted sum of all connectors that are used in a process model.
- **MISMATCH:** The sum of connector pairs that do not match with each other, e.g. when an AND-split is followed up by an OR-join.
- **DEPTH:** The maximum nesting of structured blocks in a process model.
- **CONNECTIVITY, DENSITY:** While **CONNECTIVITY** relates to the ratio of the total number of arcs in a process model to its total number of nodes, **DENSITY** relates to the ratio of the total number of arcs in a process model to the theoretically maximum number of arcs (i.e. when all nodes are directly connected).
- **CROSS-CONNECTIVITY:** The extent to which all the nodes in a model are connected to each other.
- **SEQUENTIALITY:** The degree to which the model is constructed of pure sequences of tasks.
- **SEPARABILITY:** The ratio of the number of cut-vertices on the one hand, i.e. nodes that serve as bridges between otherwise disconnected components, to the total number of nodes in the process model on the other.
- **STRUCTUREDNESS:** The extent to which a process model can be built by nesting blocks of matching split and join connectors.
- **CONNECTOR HETEROGENEITY:** The extent to which different types of connectors are used in a process model.

To illustrate these factors, we refer the reader to Figure 2. Shown here is a model of a loan request process expressed in the EPC modeling notation, which is elaborated in [18, pp. 19-20]. In addition to the standard EPC notational elements, tags are added to identify sequence arcs, cut vertices, and cycle nodes. Additionally, the numbers of incoming and outgoing arcs are given for each node, as well as a bold arc that provides the diameter of the model. All these notions are instrumental in calculating the exact values of the model factors that were presented previously. For this particular model, the values of the model factors are given in Table I.

Having discussed the independent variables, we need to address now how a process model's understanding is captured. There are various dimensions in how far comprehension can be measured, for an overview see [65]. For our research, we focus on a **SCORE** variable. **SCORE** is a quantification of a respondent's accurate understanding of a process model. This ratio is determined by the set of correct answers to a set of seven closed questions and one open question. The closed questions confront the respondent with execution order, exclusiveness, concurrency, and repeatability issues (see Section II-A) which are linked to closed questions (yes/no/I don't

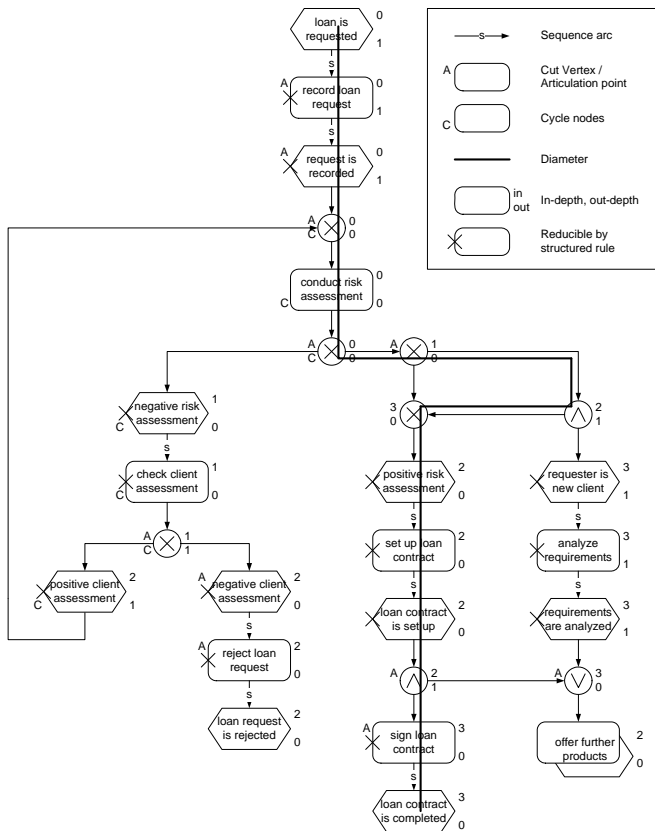


Fig. 2. Sample process model to illustrate the model factors

 TABLE I  
 METRICS' VALUES FOR THE SAMPLE PROCESS MODEL IN FIGURE 2

#NODES	27	AV. CONN. DEGREE	3
#ARCS	29	MAX. CONN. DEGREE	3
#TASKS	8	CONTROL FLOW COMPLEXITY	8
#CONNECTORS	8	MISMATCH	8
#AND-SPLITS	0	DEPTH	1
#AND-JOINS	2	CONNECTIVITY	1.074
#XOR-SPLITS	3	DENSITY	0.040
#XOR-JOINS	2	CROSS-CONNECTIVITY	0.065
#OR-SPLITS	0	SEQUENTIALITY	0.345
#OR-JOINS	1	SEPARABILITY	0.440
DIAMETER	14	STRUCTUREDNESS	0.556
TOKEN SPLITS	2	CONNECTOR HETEROGENEITY	0.819

know). The remaining question is open; a respondent is asked to identify and describe any model problem, if the respondent feels that any such problem exists. While the closed question can add 0 point (wrong answer) or 1 point (correct answer) to SCORE, the open question can add 0 points (wrong answer), 1 point (partially correct answer), or 2 points (completely correct answer). As such, the SCORE value for any model evaluation by a respondent may range between 0 and 9.

Finally, our expectations on how the various independent variables (personal and model factors) affect the dependent variable (process model understanding), can now be described as follows. For the personal factors, *more* theoretical knowledge or practical experience with respect to process modeling is likely to positively affect a person's understanding of a process model; less of these factors have the opposite effect. Furthermore, because all involved respondents received a

process modeling education at an academic level and students were not expected to have any extensive practical experience with process modeling, we did not expect that the exact educational background would have any affect on a person's understanding of a process model. This set of expectations can be summarized as hypothesis **H1**: The more the person can be regarded to be an expert, the better will be his or her understanding of a model.

Model factors have been hypothesized to have notable effects on their understanding, see [14], [18] for the related discussions. In short, the *higher* a process model's sequentiality, separability, or structuredness the easier it is to understand such a model; lower values have the opposite effect. Similarly, understandability of a process model will also increase by a *lower* number of nodes, arcs, tasks, and connectors – regardless of its kind – on the one hand, or *lower* values for its diameter, connectivity, density, token splits, average connector degree, maximum connector degree, mismatch, depth, control flow complexity, connector heterogeneity, and cross-connectivity on the other. Higher values of these model factors will have the opposite effect. This set of expectations can be summarized as hypothesis **H2**: The more complex the model is, the less it will be understood.

## B. Planning and operation

The survey was conducted in three phases. First, we collected a set of process models from practice used for documentation purposes. From this set, we originally selected eight that had an equivalent number of tasks (25), applied the uniform layout to each of them, and then constructed two additional variants for each of these. We then developed closed questions related to execution order, exclusiveness, concurrency and repeatability issues for each of the process models. We also included one open question that was the same for each model, i.e. "If there is any problem with this process (e.g. proper completion, deadlocks, etc.), please describe it.". The correct answers for all these questions were determined with the EPC analysis tools introduced in [66]. This tool was also used to calculate the set of process model metrics that we have described in Section III. For the first version of the questionnaire, we conducted a pre-test at Eindhoven University of Technology, involving 5 staff members and 7 Ph.D. students of the Information Systems group. The pre-test led to a reduction of the model set to 12 process models, i.e. four process model families, and a reformulation of some questions. We dropped the more simple models to prevent fatigue.

Second, we created six versions of the questionnaire, each with a different randomized order of the base models and its variants. The purpose was to eliminate learning effects throughout the answering. The questionnaire was filled out in class settings at the Eindhoven University of Technology, the University of Madeira, and the Vienna University of Economics and Business Administration by 73 students in total (see Table II). This led to a total of 847 complete model evaluations out of a theoretic maximum of 876 (= 73 students x 12 models). At that point in time, students were

TABLE II  
PARTICIPANTS IN THE SURVEY

TU Eindhoven	Uni Madeira	WU Vienna
30 students	23 students	20 students
graduate level	under-graduate level	under-graduate level
Petri nets	Petri nets and EPCs	EPCs

following or completing courses on process modeling at these institutions. Participation was voluntarily. The motivation for the students was the fact that they felt to be in a competitive situation with students from other universities, and that the questionnaire could be used as a good exam preparation. We captured the institution with the categorical variable EDUCATION as the study program differed. Eindhoven students had been taught about ‘soundness’ [67] (a general correctness criterion for workflow nets), reachability graphs, and related concepts. These concepts can be expected to help answering the questions of the survey. Moreover, Eindhoven students were at the graduate level while the students from Madeira and Vienna were still in their third year of undergraduate studies. The Eindhoven students were trained on Petri nets, the Vienna students in EPCs, and the Madeira students had knowledge of both the Petri nets and EPCs (see Table II).

The answers were coded and analyzed using the statistics software packages SPSS and STATGRAPHICS. While the correct answers to the closed questions could be counted automatically, all answers to the open questions were evaluated by both authors on the basis of consensus. To determine the reliability of our measuring instrument of understandability, namely SCORE, we determined Cronbach’s alpha for each of the process model families’ question sets, leading to values ranging between 0.675 and 0.817. These values are considered as acceptable, and comparable to other questionnaires used in the context of business process oriented research, see e.g. [68].

Third, to validate our findings we repeated the survey with a group of 8 professionals working within one of the world’s largest mobile telecommunications companies. At the time of the validation, all these people were working at the Dutch headquarters of this company in a unit concerned with supporting operational improvement projects; process modeling is one of their daily activities. Their participation was voluntary and was part of an in-house master class on Business Process Management, as provided by the authors. While 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 [65], [69], this validation is of importance considering the insecure external validity of student experiments and surveys in information systems research (see [70]).

#### IV. DATA ANALYSIS AND INTERPRETATION

This section presents the data analysis and interpretation starting with discussing aspects of understanding, continues with personal and model factors, and closes with an assessment of their relative importance.

#### A. Data Exploration

First of all, we tested for a normal distribution of the SCORE variable. The Kolmogorov-Smirnov test showed that the normality assumption does not hold ( $P$ -value = 0.000). Therefore, standard ANOVA techniques are not applicable for the hypothesis tests. Where possible, we will rely in the remainder on the non-parametric Kruskal-Wallis test, which is an analysis of variance by ranks. It is accepted as an alternative to ANOVA in case the considered variable is not normally distributed [71].

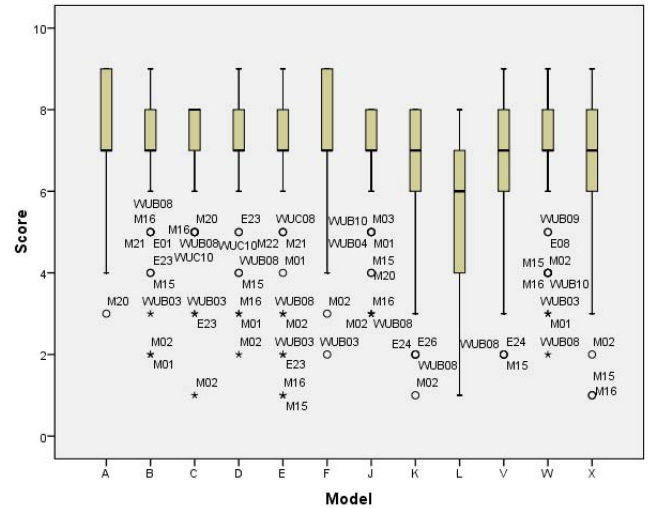


Fig. 3. Boxplot of SCORE per model

A more detailed understanding of the distribution of SCORE can be gained from the boxplots in Figure 3. They represent its distribution for each student per model. Some similar patterns emerge for all the models – with the exception of model L: (1) the median SCORE equals either 7 or 8, (2) 50% of the SCORE values always lie between 6 and 9, and (3) various outliers, some of them extreme, occur towards the lower end of the scale (these are tagged with the respondents’ identifiers). To test for any statistical differences in SCORE across the models we applied the Kruskal-Wallis test at a 95% confidence level [71], [72]. When all models are compared with this test *excluding* model L, no significant differences between the models can be observed with respect to SCORE ( $P$ -value = 0.257). This confirms that the respondents’ understanding is comparable across all models, with the exception of model L.

While models J, K, and L stem from the same family, there is a variation as displayed in Figure 4<sup>1</sup>. The second logical routing element from the top distinguishes the three models from each other. For model L this is an XOR-split routing element, for models J and K an AND-split and OR-split respectively. The variation in SCORE essentially stems from two questions relating to this part, which got few correct answers for model L ( $\ll 20$ ):

- “If T is executed for a case, can U be executed for the same case?”, and

<sup>1</sup>Note that the complete model L can be found in Appendix A.



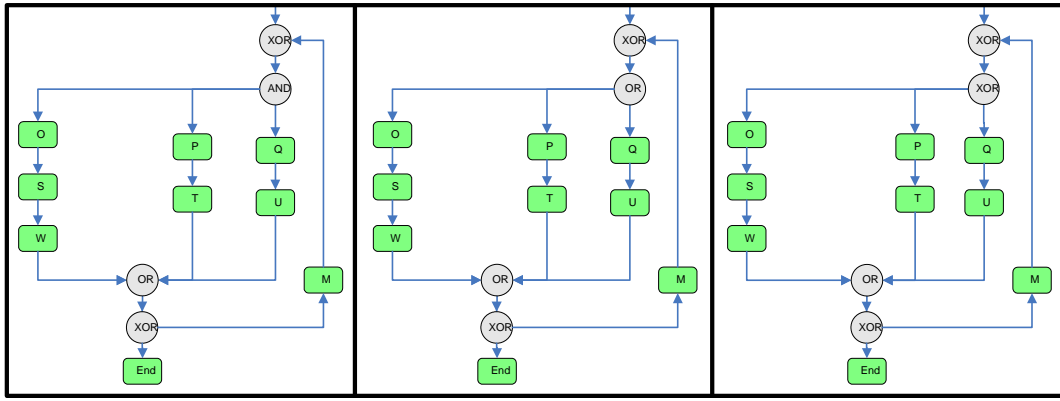


Fig. 4. Fragments of process models J, K, and L (from left to right)

- “Can T, M, and O all be executed for the same case?”

Figure 4 shows that the distinguishing connectors (AND-split and OR-split) directly allow for the interleaved execution of T and U. But even for L – the rightmost model in Figure 4 – it is possible that T and U will be executed for the same case. However, this can only happen after a cycle through M. It seems plausible that this is overlooked by many respondents. Similarly, with respect to the second question, many respondents may have failed to see that T, M, and O can be executed in the rightmost model (which is clearly also possible in the other two models).

This initial analysis provides us with two important insights. In the first place, the lack of significant differences in SCORE across most models potentially points to the fact that model size is indeed a primary factor that impacts on model understandability. The number of tasks for all the models is, by design, exactly the same, and so are the levels of understanding of these models. Furthermore, our detailed analysis of the exceptional model shows that the change of a single element *can* have a significant impact on a model’s understandability. So, despite the potentially dominant impact of size, the search for the additional impact factors seems indeed relevant, which is in line with the expectations of this research (see Section II-B).

### B. Personal factors

In this section we operationalize **H1** as follows: There is no effect of each predictor variable THEORY, PRACTICE, and EDUCATION on the expected predicted variable SCORE. The alternative hypothesis states: Some predictor variables do have an effect on the expected predicted variable SCORE.

Before we undertook our experiment, we had no reason to expect differences in SCORE between respondents with different academic backgrounds. All respondents had received at least a basic training in the use of process modeling techniques at the time they took the questionnaire. Also, the exposure to process modeling in practice would be negligible for all involved respondents. To test the absence of such a difference, we computed the average SCORE over the 12 models. In Figure 5, the distribution of the average SCORE as gained by the respondents from the different universities is shown as boxplots.

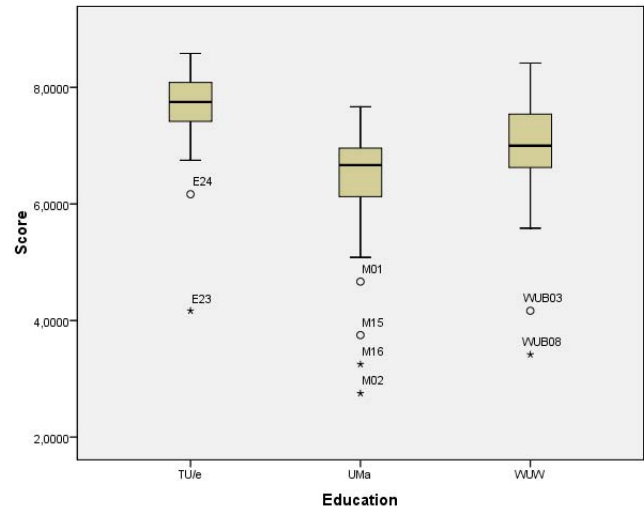


Fig. 5. Boxplot of average SCORE for different values of EDUCATION

If no difference would exist between the three distributions of total SCORE, students can be assumed to perform similarly across the three universities. To test this, we again applied the non-parametric Kruskal-Wallis test, because application of the Kolmogorov-Smirnov test indicates that with a 95% confidence average SCORE is not normally distributed for any university.

Contrary to our expectations, the application of the Kruskal-Wallis test does indicate that there is a statistically significant difference among the distributions at a 95% confidence level for the different types of education (P-value = 0.000). In other words, differences exist in the ability of respondents to answer questions correctly *across* the three universities. Additional pairwise Mann-Whitney tests [71] were conducted, taking into account the appropriate Bonferroni adjustment to control for Type 1 errors due to the three additional tests, i.e. using a stricter alpha level of 0.017 (= 0.05/3). These tests indicate that respondents from TU/e (Eindhoven) perform significantly better than respondents from the universities of Madeira (P-value = 0.000) and Vienna (P-value = 0.001), although the difference between the respondents from the universities of Madeira (UMa) and Vienna (WUW) is not significant (P-value



TABLE III  
PERSONAL FACTORS: KRUSKAL-WALLIS ANALYSIS OF DIFFERENCES IN UNDERSTANDING (SCORE)

personal factor	P-value
THEORY	0.293
PRACTICE	0.121
EDUCATION	0.000***

\*significant at 90% confidence level, \*\* at 95%, \*\*\* at 99%

= 0.027).

A retrospective analysis of the courses offered at the various universities revealed that the hours spent on actual modeling is the highest in Eindhoven, which may explain the noted difference. In particular, Eindhoven students have been explicitly and thoroughly taught about ‘soundness’ [67], a general correctness criterion for workflow nets, reachability graphs, and related concepts. An alternative explanation is that Eindhoven students are graduate students while the students from Madeira and Vienna are still in their third year of undergraduate studies. But note that this too indicates that a difference in theoretical knowledge is important. Interestingly, across the different universities different modeling techniques are taught. The Eindhoven students were trained in workflow nets (based on the Petri net formalism), the Vienna students in EPCs, and the Madeira students had knowledge of both the Petri net formalism and EPCs. So, interestingly, the choice of our EPC-like notation does *not* obviously favor students who are familiar with EPCs.

The analysis with the Kruskal-Wallis test of the other investigated personal factors, THEORY and PRACTICE, does not identify any statistically significant differences with respect to SCORE. The outcome of the Kruskal-Wallis tests for all personal factors considered is summarized in Table III. It should be noted that THEORY and PRACTICE both rely on self-assessments where EDUCATION can be determined objectively. With respect to the latter factor, we can conclude that it is potentially an important factor of influence in the understanding of process models, perhaps hinting at the effect of differences in theoretical background in understanding process modeling. The particular notation on which one receives training does not seem to be of any importance, but rather we would suggest that the knowledge of abstract process modeling notions does. Altogether, our analysis provides support for hypothesis **H1** that the more the person can be regarded to be an expert, the better will be his or her understanding of a model.

### C. Model factors

In this section we operationalize **H2** as follows: There is no effect of a predictor variable capturing process model complexity on the expected predicted variable SCORE. The alternative hypothesis states: Some of these predictor variables do have an effect on the expected predicted variable SCORE.

To determine whether model factors, as described in Section III-A, are helpful to explain variations in model understandability, we used the models’ average SCORE (see Figure 3) and determined Pearson correlation coefficients with all potential factors. Recall that the ANOVA test is

TABLE IV  
MODEL FACTORS: CORRELATION ANALYSIS WITH UNDERSTANDING (SCORE)

model factor	corr.coeff.	P-value
#OR JOINS	-0.330	0.295
DENSITY	-0.618	0.032**
AV. CONNECTOR DEGREE	-0.674	0.016**
MISMATCH	-0.438	0.154
CONNECTOR HETEROGENEITY	-0.312	0.323
CROSS-CONNECTIVITY	-0.549	0.065*

\*significant at 90% confidence level, \*\* at 95%, \*\*\* at 99%

not applicable because of the non-normal distribution of the SCORE variable. Also, the Kruskal-Wallis test could not be used as an alternative because of the continuous scale on which many of the considered potential factors are measured (e.g. STRUCTUREDNESS and SEQUENTIALITY).

Of all the correlation coefficients that were established, six of them displayed the direction of the influence that we hypothesized upon with respect to the understandability of process models, i.e. #OR JOINS, DENSITY, AVERAGE CONNECTOR DEGREE, MISMATCH, CONNECTOR HETEROGENEITY, and CROSS-CONNECTIVITY. For example, the correlation coefficient between CONNECTOR HETEROGENEITY and average SCORE equals  $-0.312$ , which corresponds with the intuition that the use of a wider variety of connectors in a model will decrease its understandability. However, as can be seen in Table IV, only the correlation coefficients of DENSITY and AVERAGE CONNECTOR DEGREE are significant at a 95% confidence level.

To examine the value of the distinguished factors in explaining differences in SCORE more thoroughly, we developed various linear regression models – even though it should be noted that the number of 12 different model observations is quite low for this approach. We compared all 63 ( $= 2^6 - 1$ ) linear regression models that take a non-empty subset into account of the factors shown in Table IV. To differentiate between the regression models, we used the adjusted  $R^2$  statistic that measures how the variability in the SCORE is explained by each model. The best adjusted  $R^2$  statistic equals 79%, which is quite a satisfactory result. It belongs to the regression model that uses #OR JOINS, DENSITY, AVERAGE CONNECTOR DEGREE, MISMATCH, and CROSS-CONNECTIVITY. Use of the Durbin-Watson (DW) statistic test indicates no serial autocorrelation in the residuals at the 95% confidence level. A visualization of this regression model can be seen in Figure 6. Note that the outlying model L can be clearly identified at the bottom left corner.

As stated before, the number of models is too small to make strong claims. Under this proviso, it is interesting to see that the two factors which most convincingly relate to model understandability *both* relate to the number of connections in a process model, rather than, for example, the generated state space. The AVERAGE CONNECTOR DEGREE measures the model’s average of incoming/outcoming arcs per routing element, while DENSITY gives the ratio of existing arcs to the maximal number of arcs between the nodes in the model (i.e. when it would be completely connected). Both factors point to

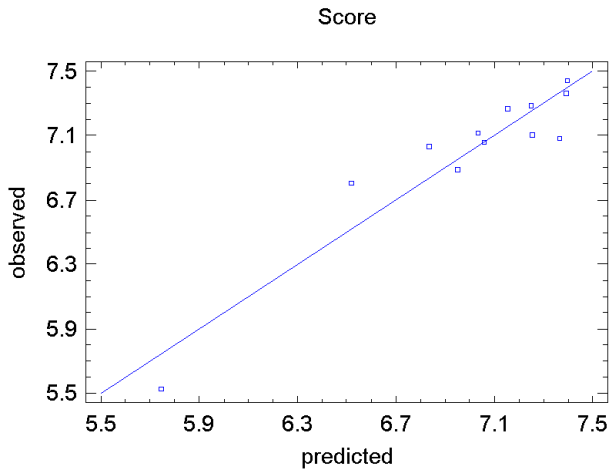


Fig. 6. Multivariate linear regression model explaining the average SCORE

the negative effect of a relatively high number of dependencies in a model on a model’s understandability. Apparently, if size is kept constant, factors related to complexity seems to be the most significant ones. Altogether, the findings tend to partially support **H2** that the more complex the model is, the worse will be the understandability of it. Given the small set of models, future research needs to further investigate those metrics that were significant in this survey.

#### D. Personal versus model factors

At this point, we have seen relations between both personal factors and model factors on the one hand and the objective understanding of a process model as measured through our SCORE variable on the other. To investigate which of these domains has the bigger influence, we developed various linear regression models to explain the variation in the average SCORE for each model. For this purpose, we used all 847 complete model evaluations at our disposal. Using the stepwise method as available in SPSS for automatically selecting the significant factors, we developed different regression models using (1) personal factors only, (2) model factors only, and (3) all factors. The idea behind this approach is to see which of the regression models has the better explanatory power. The results are summarized in Table V.

The adjusted  $R^2$  statistic displays rather low values, but this is not surprising. After all, personal factors will not differ for any evaluation that the same respondent is involved in. Similarly, model factors will not vary either for evaluations of the same model. Moreover, the understanding of the models is quite similar as we have achieved by keeping the size of the models different. Nonetheless, these results give an important insight into the relative importance of personal factors versus model factors in explaining model understanding. As can be seen from Table V, the adjusted  $R^2$  value for the personal factors is more than twice as high as that for the model factors. This is a clear indication that even though both types of factors may influence process model understanding, the influence of personal factors is bigger. In addition, the combination of both types of factors gives the highest explanatory power.

## V. THREATS TO VALIDITY

One particular aspect of the external validity of the presented research relates to the extent to which the used models are representative for real-world models. As explained, we countered this threat by our choice of real process models. While this has increased the level of external validity, this choice made it more difficult to manipulate the range of metric value ranges. In that sense, improved external validity may have actually lowered the internal validity.

The other important aspect which refers to a potentially limited external validity, relates to the involvement of students. As stated, we involved a number of experienced process modelers in a replication of the survey (see Section III-B). The average SCORE that the professionals attained in this replication ranged from 3.25 to 7.17, with an average value of 5.83. To compare, the SCORE values for all students combined ranged from 2.75 to 8.58, with an average value of 6.92. To test for any significant differences in process model understanding between the professional modelers and the students, we applied the Kruskal-Wallis test. Recall that we had already established that the average SCORE is not normally distributed for the student population, which justifies the choice for this test.

The application of the Kruskal-Wallis test indicates a statistically significant difference among the various populations at a 95% confidence level (P-value = 0.000). Additional pairwise Mann-Whitney tests [71] were conducted, taking into account the appropriate Bonferroni adjustment to control for Type 1 errors. In this way, the cut-off equals 0.008, which is determined as the alpha level of 0.05 divided by the number of pairwise tests (6). Interestingly, the professional models perform similarly as the students from Madeira (UMa) (P-value = 0.684) and Vienna (WUW) (P-value = 0.075) but *less* than the Eindhoven (TU/e) students (P-value = 0.000). While the median average SCORE for the professionals was 6.375, it was 7.75 for the Eindhoven students.

In other words, based on our replications it does not seem that students perform less with respect to cognitive tasks than professionals with their greater experience, which is a common worry for this kind of research. In comparison with one particular subgroup of the students, a contrary effect could be observed. In the context of this study, this may be an indication that knowledge of abstract modeling notions may be key in explaining process model understanding. In the wider context of empirical research, the outcome of the replication is encouraging with respect to the involvement of students instead of professional modelers.

## VI. DISCUSSION AND CONCLUSION

### A. Summary, results and research contributions

In this paper we have motivated and investigated the impact of personal and model related factors on understandability of process models. Our main hypotheses were that expert modelers will perform significantly better and that the complexity of the model affects understanding. We explored different operationalizations of these hypotheses and found supportive evidence. Furthermore, we calculated a combined regression

TABLE V  
REGRESSION MODELS FOR PERSONAL AND MODEL FACTORS

type	selected factors	adjusted $R^2$
PERSONAL FACTORS ONLY	THEORY, PRACTICE, EDUCATION	0.158
MODEL FACTORS ONLY	AVERAGE CONNECTOR DEGREE, MISMATCH, CONNECTOR HETEROGENEITY, and CROSS-CONNECTIVITY	0.061
ALL FACTORS	THEORY, PRACTICE, EDUCATION, #OR JOINS, DENSITY, AVERAGE CONNECTOR DEGREE, MISMATCH, and CROSS-CONNECTIVITY	0.231

model that permits first and preliminary conclusions on the relative importance of both groups of factors.

### B. Implications and recommendations

Our research has implications both for research and practice. We found that personal factors have a stronger explanatory power in terms of adjusted  $R^2$  than model related factors. Such a comparative conjecture is novel and requires further research efforts. Clearly, this result supports early research on visual programming which found that experts are capable of understanding even complex and poorly laid out diagrams [32]. It is highly relevant to further investigate this relationship, for instance using experiments, for its practical implications. From an industrial point of view the question of relative importance impacts investment decisions on process modeling. If personal competence is more critical to understanding, it might be more beneficial for a company to invest in training instead of putting efforts in restructuring complex models. While it is too early to draw such general conclusions from our research, it is clearly in line with expert and novice research on diagram understanding [32].

For this survey we have intentionally selected models of equivalent size. The reason for this choice was motivated by established research that points to the importance of size, and because little is known about the significance of other structural factors beyond size. We found that two model factors showed a significant correlation with understanding, namely average connector degree and density. An increase of these factors implies a negative effect on a model's understandability. Apparently, if size is kept constant, factors related to complexity seem to be the most significant ones. This is an interesting finding that calls for future research with a larger set of models of similar size and varying complexity.

Another important implication for research stems from our validation with process modeling experts from the telecommunications sector. While research often raises issues on external validity of student surveys for potential lack of commitment and motivation, we have found that the professionals could not be distinguished from the Vienna and Madeira students from their performance. Interestingly, professionals performed not as well as the Eindhoven students. Our validation seems to point to the value of theoretical knowledge for this particular problem domain rather than practical experience. It would be extremely valuable to try and replicate these findings for the effect this may have on empirical research in the field of process modeling and design.

### C. Limitations and future work

In this research, we aimed at investigating factors of process model understandability beyond size, which has been well established in prior research as an important factor. We focused on model complexity and personal modeling expertise while keeping other factors including model purpose, problem domain, notation, and layout constant. The exclusion of size as model related factor implies certain limitations regarding the interpretation of the relative importance of the investigated factors. While we found that expertise appears to matter more for understandability than complexity, it must be kept in mind that size remained constant. Future research is needed for analyzing the relative importance of model size in comparison to personal expertise, and should explicitly consider potential interaction effects. We also plan to investigate the significance of those factors for understanding that we neutralized in this research. In particular, we are currently working on experiments that investigate the importance of secondary notation (visual layout, color highlighting) on understanding. Finally, the case of model L points to research opportunities on the difficulty of understanding particular process model components. While only one element was modified, the behavior of the component changed significantly. Decomposition techniques such as reduction rules play an important role for the verification of process models. As certain components can be reduced because they are always correct, it might be interesting to investigate whether certain components can be understood with the same ease, even if they are moved to a different position in the process model.

## APPENDIX A QUESTIONNAIRE MATERIAL USED

### Part 1: Personal characteristics

- How do you assess your theoretical knowledge on business process modeling? (I have weak theoretical knowledge / I have rather weak theoretical knowledge / I have mediocre theoretical knowledge / I have rather strong theoretical knowledge / I have strong theoretical knowledge)
- How often do you use business process modeling in practice? (I never use business process modeling in practice / I sometimes use business process modeling in practice / I regularly use business process modeling in practice, but not every day / I use business process modeling in practice every day)

### Part 2: Model understanding (Model L as example)

Please have a look at process model L in Figure 7 and answer the questions below.

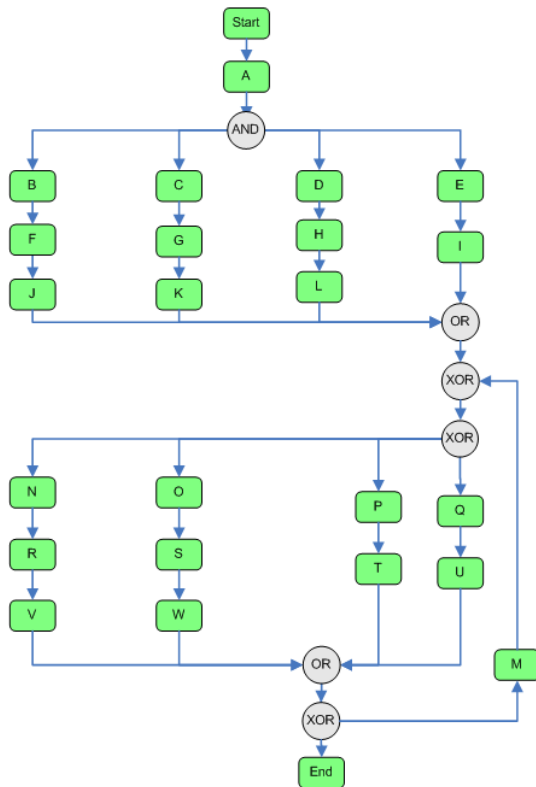


Fig. 7. Model L

- How do you evaluate process model L? (It is easy to understand the model / It is rather easy to understand the model / It is rather difficult to understand the model / It is difficult to understand the model)
- If N is executed for a case, can P be executed for the same case? (Yes/No/I don't know)
- Can U be executed more than once for the same case? (Yes/No/I don't know)
- If T is executed for a case, can U be executed for the same case? (Yes/No/I don't know)
- If B is executed for a case, can Q be executed for the same case? (Yes/No/I don't know)
- Can M be executed more than once for the same case? (Yes/No/I don't know)
- If T is executed for a case, is U then always executed for the same case? (Yes/No/I don't know)
- Can T, M, and N all be executed for the same case? (Yes/No/I don't know)
- If there is any problem with this process, e.g. proper completion, deadlocks, etc., please describe it. (free text)

## ACKNOWLEDGMENT

The authors would like to thank Jorge Cardoso and Ronny Mans for their help in carrying out the survey, and all the people that participated in it.

## REFERENCES

- [1] Y. Wand and R. Weber, "Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda," *Information Systems Research*, vol. 13, no. 4, p. 363, 2002.
- [2] I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Gallo, "How do practitioners use conceptual modeling in practice?" *Data & Knowledge Engineering*, vol. 58, no. 3, pp. 358–380, 2006.
- [3] M. Dumas, W. M. P. van der Aalst, and A. H. M. ter Hofstede, Eds., *Process Aware Information Systems: Bridging People and Software Through Process Technology*. Hoboken, New Jersey: John Wiley & Sons, 2005.
- [4] T. Erl, *Service-oriented Architecture: Concepts, Technology, and Design*. Upper Saddle River, New Jersey: Prentice Hall, 2005.
- [5] C. Ferris, "What are web services?" *Communications of the ACM*, vol. 46, no. 6, pp. 31–32, 2003.
- [6] J. Dehnert and W. Aalst, "Bridging The Gap Between Business Models And Workflow Specifications," *International J. Cooperative Inf. Syst.*, vol. 13, no. 3, pp. 289–332, 2004.
- [7] M. Rosemann, "Potential pitfalls of process modeling: part a," *Business Process Management Journal*, vol. 12, no. 2, pp. 249–254, 2006.
- [8] J. Mendling, "Empirical studies in process model verification," *LNCS Transactions on Petri Nets and Other Models of Concurrency II, Special Issue on Concurrency in Process-Aware Information Systems*, vol. 2, pp. 208–224, 2009.
- [9] J. Mendling, H. Verbeek, B. Dongen, W. Aalst, and G. Neumann, "Detection and Prediction of Errors in EPCs of the SAP Reference Model," *Data & Knowledge Engineering*, vol. 64, no. 1, pp. 312–329, January 2008.
- [10] H. Simon, *The Sciences of the Artificial*. MIT Press, 1996.
- [11] A. Maes and G. Poels, "Evaluating quality of conceptual modelling scripts based on user perceptions," *Data & Knowledge Engineering*, vol. 63, no. 3, pp. 701–724, December 2007.
- [12] M. Genero, G. Poels, and M. Piattini, "Defining and validating metrics for assessing the understandability of entity–relationship diagrams," *Data & Knowledge Engineering*, vol. 64, no. 3, pp. 534–557, 2008.
- [13] J. Mendling, H. Reijers, and J. Cardoso, "What makes process models understandable?" in *Proc. BPM'07*, 2007, pp. 48–63.
- [14] I. Vanderfeesten, H. Reijers, J. Mendling, W. Aalst, and J. Cardoso, "On a Quest for Good Process Models: The Cross-Connectivity Metric," *Lecture Notes in Computer Science*, vol. 5074, pp. 480–494, 2008.
- [15] E. Kavakli and P. Loucopoulos, "Experiences with goal-oriented modelling of organisational change," *IEEE Transactions on Systems, Man and Cybernetics - Part C*, vol. 36, no. 2, pp. 221–235, 2006.
- [16] J. M. Zaha, M. Dumas, A. H. M. ter Hofstede, A. P. Barros, and G. Decker, "Bridging global and local models of service-oriented systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 38, no. 3, pp. 302–318, 2008.
- [17] G. Keller, M. Nüttgens, and A.-W. Scheer, "Semantische Prozessmodellierung auf der Grundlage "Ereignisgesteuerter Prozessketten (EPK)"; Institut für Wirtschaftsinformatik, Saarbrücken, Germany, Heft 89, 1992.
- [18] J. Mendling, *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, ser. Lecture Notes in Business Information Processing. Springer, 2008, vol. 6.
- [19] Object Management Group, "Business Process Modeling Notation (BPMN) Specification," Object Management Group, Final Adopted Specification, dtc/06-02-01, February 2006.
- [20] E. Kindler, "On the semantics of EPCs: Resolving the vicious circle," *Data & Knowledge Engineering*, vol. 56, no. 1, pp. 23–40, 2006.
- [21] J. Esparza, "Decidability and complexity of petri net problems - an introduction," in *Lectures on Petri Nets I: Basic Models, Advances in Petri Nets, held in Dagstuhl, September 1996*, ser. Lecture Notes in Computer Science, W. Reisig and G. Rozenberg, Eds., vol. 1491. Springer, 1998, pp. 374–428.
- [22] A. Kovalyov and J. Esparza, "A polynomial algorithm to compute the concurrency relation of free-choice signal transition graphs," in *Prof. of the International Workshop on Discrete Event Systems, WODES'96*. Edinburgh: The Institution of Electrical Engineers, 1996, pp. 1–6.
- [23] M. Weidlich, J. Mendling, and M. Weske, "Efficient consistency measurement based on behavioural profiles of process models," *IEEE Transactions on Software Engineering*, 2010, accepted for publication.
- [24] T. Green and M. Petre, "Usability analysis of visual programming environments: A 'cognitive dimensions' framework," *J. Vis. Lang. Comput.*, vol. 7, no. 2, pp. 131–174, 1996.
- [25] T. Green, "Conditional program statements and their comprehensibility to professional programmers," *Journal of Occupational Psychology*, vol. 50, pp. 93–109, 1977.
- [26] I. Vessey, "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature\*," *Decision Sciences*, vol. 22, no. 2, pp. 219–240, 1991.

- [27] J. Esparza, "Reachability in live and safe free-choice petri nets is np-complete," *Theor. Comput. Sci.*, vol. 198, no. 1-2, pp. 211–224, 1998.
- [28] R. Hauser, M. Friess, J. M. Küster, and J. Vanhatalo, "An incremental approach to the analysis and transformation of workflows using region trees," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 38, no. 3, pp. 347–359, 2008.
- [29] E. Teruel, J. M. Colom, and M. Silva, "Choice-free petri nets: A model for deterministic concurrent systems with bulk services and arrivals," *IEEE Transactions on System, Man and Cybernetics. Part A*, vol. 27, no. 1, pp. 73–83, 1997.
- [30] J. Mendling, H. A. Reijers, and J. Recker, "Activity labeling in process modeling: Empirical insights and recommendations," *Inf. Syst.*, vol. 35, no. 4, pp. 467–482, 2010.
- [31] C. Ware, H. C. Purchase, L. Colpoys, and M. McGill, "Cognitive measurements of graph aesthetics," *Information Visualization*, vol. 1, no. 2, pp. 103–110, 2002.
- [32] M. Petre, "Why looking isn't always seeing: Readership skills and graphical programming," *Commun. ACM*, vol. 38, no. 6, pp. 33–44, 1995.
- [33] E. McCormick and M. Sanders, *Human factors in engineering and design*. McGraw-Hill Companies, 1982.
- [34] C. Wickens and A. Kramer, "Engineering psychology," *Annual Review of Psychology*, vol. 36, no. 1, pp. 307–348, 1985.
- [35] D. L. Moody, "The é;physicsé; of notations: Toward a scientific basis for constructing visual notations in software engineering," *IEEE Trans. Software Eng.*, vol. 35, no. 6, pp. 756–779, 2009.
- [36] T. Green, W. Ribarsky, and B. Fisher, "Building and applying a human cognition model for visual analytics," *Information Visualization*, vol. 8, no. 1, pp. 1–13, 2009.
- [37] P. Lindsay and D. Norman, *Human Information Processing: An introduction to psychology*, second edition ed. Academic Press, 1977.
- [38] K. Ericsson and A. Lehmann, "Expert and exceptional performance: Evidence of maximal adaptation to task constraints," *Annual Review of Psychology*, vol. 47, no. 1, pp. 273–305, 1996.
- [39] W. Chase and H. Simon, "The minds eye in chess. Visual Information Processing," 1973.
- [40] R. Riding and I. Cheema, "Cognitive styles: An overview and integration," *Educational Psychology*, vol. 11, no. 3-4, pp. 193–215, 1991.
- [41] J. Meyer, J. Thomas, S. Diehl, B. Fisher, D. Keim, D. Laidlaw, S. Miksch, K. Mueller, W. Ribarsky, B. Preim *et al.*, "From Visualization to Visually Enabled Reasoning."
- [42] T. McCabe, "A complexity measure," *IEEE Transactions on Software Engineering*, vol. 2, no. 4, pp. 308–320, 1976.
- [43] S. Chidamber and C. Kemerer, "A metrics suite for object oriented design," *IEEE Transaction on Software Engineering*, vol. 20, no. 6, pp. 476–493, 1994.
- [44] N. Fenton and S. Pfleeger, *Software Metrics. A Rigorous and Practical Approach*. PWS, Boston, 1997.
- [45] G. Lee and J.-M. Yoon, "An empirical study on the complexity metrics of petri nets," *Microelectronics and Reliability*, vol. 32, no. 3, pp. 323–329, 1992.
- [46] M. Nissen, "Redesigning reengineering through measurement-driven inference," *MIS Quarterly*, vol. 22, no. 4, pp. 509–534, 1998.
- [47] S. Morasca, "Measuring attributes of concurrent software specifications in petri nets," in *METRICS '99: Proceedings of the 6th International Symposium on Software Metrics*. Washington, DC, USA: IEEE Computer Society, 1999, pp. 100–110.
- [48] J. Cardoso, *Workflow Handbook 2005*. Lighthouse Point, FL, USA: Future Strategies, Inc., 2005, ch. Evaluating Workflows and Web Process Complexity, pp. 284–290.
- [49] —, "Process control-flow complexity metric: An empirical validation," in *Proceedings of IEEE International Conference on Services Computing (IEEE SCC 06)*, Chicago, USA, September 18-22. IEEE Computer Society, 2006, pp. 167–173.
- [50] G. Canfora, F. García, M. Piattini, F. Ruiz, and C. Visaggio, "A family of experiments to validate metrics for software process models," *Journal of Systems and Software*, vol. 77, no. 2, pp. 113–129, 2005.
- [51] E. R. Aguilar, F. García, F. Ruiz, and M. Piattini, "An exploratory experiment to validate measures for business process models," in *First International Conference on Research Challenges in Information Science (RCIS)*, 2007.
- [52] J. Vanhatalo, H. Völzer, and F. Leymann, "Faster and more focused control-flow analysis for business process models through sese decomposition," in *Service-Oriented Computing - IC3OC 2007, Fifth International Conference, Vienna, Austria, September 17-20, 2007, Proceedings*, ser. Lecture Notes in Computer Science, B. Krämer, K.-J. Lin, and P. Narasimhan, Eds., vol. 4749. Springer, 2007, pp. 43–55.
- [53] W. Aalst and K. Lassen, "Translating unstructured workflow processes to readable BPEL: Theory and implementation," *Information and Software Technology*, vol. 50, no. 3, pp. 131–159, 2008.
- [54] J. Recker and 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 *18th Australasian Conference on Information Systems*, M. Toleman, A. Cater-Steel, and D. Roberts, Eds. Toowoomba, Australia: The University of Southern Queensland, 2007, pp. 356–366.
- [55] J. Mendling and M. Strembeck, "Influence factors of understanding business process models," in *Proc. of the 11th International Conference on Business Information Systems (BIS 2008)*, ser. Lecture Notes in Business Information Processing, W. Abramowicz and D. Fensel, Eds., vol. 7. Springer-Verlag, 2008, p. 142153.
- [56] A. Lakhotia, "Understanding someone elses code: Analysis of experiences," *Journal of Systems and Software*, vol. 23, no. 3, pp. 269–275, 1993.
- [57] K. Sarshar and P. Loos, "Comparing the control-flow of epc and petri net from the end-user perspective," in *Business Process Management, 3rd International Conference, BPM 2005, Nancy, France, September 5-8, 2005, Proceedings*, ser. LNCS 3649, W. Aalst, B. Benatallah, F. Casati, and F. Curbera, Eds., 2005, pp. 434–439.
- [58] J. Hahn and J. Kim, "Why are some diagrams easier to work with? effects of diagrammatic representation on the cognitive integration process of systems analysis and design," *ACMTCHI: ACM Transactions on Computer-Human Interaction*, vol. 6, 1999.
- [59] R. Agarwal, P. De, and A. Sinha, "Comprehending object and process models: An empirical study," *IEEE Transactions on Software Engineering*, vol. 25, no. 4, pp. 541–556, 1999.
- [60] T. Moher, D. Mak, B. Blumenthal, and L. Leventhal, "Comparing the Comprehensibility of Textual and Graphical Programs: The Case of Petri Nets," in *Empirical Studies of Programmers: Fifth Workshop: Papers Presented at the Fifth Workshop on Empirical Studies of Programmers, December 3-5, 1993*, C. Cook, J. Scholtz, and J. Spohrer, Eds. Ablex Pub, 1993, pp. 137–161.
- [61] H. Purchase, "Which Aesthetic has the Greatest Effect on Human Understanding?" in *Graph Drawing: 5th International Symposium, Gd'97, Rome, Italy, September 18-20, 1997: Proceedings*. Springer, 1997, pp. 248–261.
- [62] H. A. Reijers and J. Mendling, "Modularity in process models: Review and effects," in *Business Process Management - BPM 2008*, ser. Lecture Notes in Computer Science, M. Dumas, M. Reichert, and M.-C. Shan, Eds. Milan, Italy: Springer, 2008, vol. 5240, pp. 20–35.
- [63] V. Basili, L. Briand, and W. Melo, "A validation of object-oriented design metrics as quality indicators," *IEEE Transactions on Software Engineering*, vol. 22, no. 10, pp. 751–761, 1996.
- [64] J. Mendling and W. Aalst, "Towards EPC Semantics based on State and Context," in *Proceedings of the 5th GI Workshop on Business Process Management with Event-Driven Process Chains (EPK 2006)*, M. Nüttgens and F.J. Rump and J. Mendling, Ed. Vienna, Austria: German Informatics Society, December 2006, pp. 25–48.
- [65] A. Gemino and Y. Wand, "A framework for empirical evaluation of conceptual modeling techniques," *Requir. Eng.*, vol. 9, no. 4, pp. 248–260, 2004.
- [66] J. Mendling and W. Aalst, "Formalization and Verification of EPCs with OR-Joins Based on State and Context," in *19th Conference on Advanced Information Systems Engineering (CAISE 2007)*, ser. Lecture Notes in Computer Science, J. Krogstie, A. Opdahl, and G. Sindre, Eds., vol. 4495, 2007, pp. 439–453.
- [67] W. Aalst, *Business Process Management*. Springer Verlag, 2000, vol. LNCS 1806, ch. Workflow Verification: Finding Control-Flow Errors Using Petri-Net-Based Techniques, pp. 161–183.
- [68] R. Hung, "Business Process Management as Competitive Advantage: a review and empirical study," *Total Quality Management & Business Excellence*, vol. 17, no. 1, pp. 21–40, 2006.
- [69] J. Parsons and L. Cole, "What do the pictures mean? guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques," *Data & Knowledge Engineering*, vol. 55, no. 3, p. 327342, 2005.
- [70] D. Moody, "Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions," *Data & Knowledge Engineering*, vol. 55, no. 3, pp. 243–276, 2005.
- [71] S. Siegel and N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. McGraw Hill, 1988.
- [72] L. Cohen, L. Manion, and K. Morrison, *Research methods in education*. Routledge, 2007.