Assessing the Impact of Hierarchy on Model Understandability—A Cognitive Perspective

Stefan Zugal[1], Jakob Pinggera[1], Barbara Weber[1], Jan Mendling[2], and Hajo A. Reijers[3]

[1] University of Innsbruck, Austria
{stefan.zugal|jakob.pinggera|barbara.weber}@uibk.ac.at
[2] Humboldt-Universität zu Berlin, Germany
jan.mendling@wiwi.hu-berlin.de
[3] Eindhoven University of Technology, The Netherlands
h.a.reijers@tue.nl

Abstract. Modularity is a widely advocated strategy for handling complexity in conceptual models. Nevertheless, a systematic literature review revealed that it is not yet entirely clear under which circumstances modularity is most beneficial. Quite the contrary, empirical findings are contradictory, some authors even show that modularity can lead to decreased model understandability. In this work, we draw on insights from cognitive psychology to develop a framework for assessing the impact of hierarchy on model understandability. In particular, we identify abstraction and the split-attention effect as two opposing forces that presumably mediate the influence of modularity. Based on our framework, we describe an approach to estimate the impact of modularization on understandability and discuss implications for experiments investigating the impact of modularization on conceptual models.

1 Introduction

The use of modularization to hierarchically structure information has for decades been identified as a viable approach to deal with complexity [1]. Not surprisingly, many conceptual modeling languages provide support for hierarchical structures, such as sub-processes in business process modeling languages like BPMN and YAWL [2] or composite states in UML statecharts. While hierarchical structures have been recognized as an important factor influencing model understandability [3, 4], there are no definitive guidelines on their use yet. For instance, for business process models, recommendations for the size of a sub-process, i.e., sub-model, range from 5–7 model elements [5] over 5–15 model elements [6] to up to 50 model elements [7]. Also in empirical research into conceptual models (e.g., ER diagrams or UML statecharts) the question of *whether and when* hierarchical structures are beneficial for model understandability seems not to be entirely clear. While it is common belief that hierarchy has a positive influence on the understandability of a model, reported data seems often inconclusive or even contradictory, cf. [8, 9].

As suggested by existing empirical evidence, hierarchy is not beneficial by default [10] and can even lead to performance decrease [8]. The goal of this paper is to have a detailed look at which factors cause such discrepancies between the common belief in positive effects of hierarchy and reported data. In particular, we draw on concepts from cognitive psychology to develop a framework that describes how the impact of hierarchy on model understandability can be assessed. The contribution of this theoretical discussion is a perspective to disentangle the diverse findings from prior experiments.

The remainder of this paper is structured as follows. In Sect. 2 a systematic literature review about empirical investigations into hierarchical structuring is described. Afterwards, concepts from cognitive psychology are introduced and put in the context of conceptual models. Then, in Sect. 3 the introduced concepts are used as basis for our framework for assessing the impact of hierarchy on understandability, before Sect. 4 concludes with a summary and an outlook.

2 The Impact of Hierarchy on Model Understandability

In this section we revisit results from prior experiments on the influence of hierarchy on model understandability, and analyze them from a cognitive perspective. Sect. 2.1 summarizes literature reporting experimental results. Sect. 2.2 describes cognitive foundations of working with hierarchical models.

2.1 Existing Empirical Research into Hierarchical Models

The concept of hierarchical structuring is not only applied to various domains, but also known under several synonyms. In particular, we identified synonyms *hierarchy, hierarchical, modularity, decomposition, refinement, sub-model, subprocess, fragment and module*. Similarly, model understandability is referred to as *understandability* or *comprehensibility*. To systematically identify existing *empirical investigations* into the *impact of hierarchy on understandability* within the domain of conceptual modeling, we conducted a systematic literature review [11]. More specifically, we derived the following key-word pattern for our search: (synonym modularity) X (synonym understandability) X experiment X model. Subsequently, we utilized the cross-product of all key-words for a full-text search in the online portals of Springer[1], Elsevier[2], ACM[3] and IEEE [4] to cover the most important publishers in computer science, leading to 9,778 hits. We did not use any restriction with respect to publication date, still we are aware that online portals might provide only publications of a certain time period. In the next step, we removed all publications that were not related, i.e., did not consider the impact of hierarchy on model understandability or did not report

[1] http://www.springerlink.com
[2] http://www.sciencedirect.com
[3] http://portal.acm.org
[4] http://ieeexplore.ieee.org

empirical data. All in all, 10 relevant publications passed the manual check, resulting in the list summarized in Table 1. Having collected the data, all papers were systematically checked for the influence of hierarchy. As Table 1 shows, reported data ranges from negative influence [12] over no influence [12–14] to mostly positive influence [15]. These experiments have been conducted with a wide spectrum of modeling languages. It is interesting to note though that diverse effects have been observed for a specific notation such as statecharts or ER-models. In general, most experiments are able to show an effect of hierarchy either in a positive *or* a negative direction. However, it remains unclear *under which circumstances* positive or negative influences can be expected. To approach this issue, in the following, we will employ concepts from cognitive psychology to provide a systematic view on which factors influence understandability.

Work	**Findings**
Moody [15] Domain: ER-Models	Positive influence on accuracy, no influence / negative influence on time
Reijers et al. [16, 17] Domain: Business Process Models	Positive influence on understandability for one out of two models
Cruz-Lemus et al. [9, 18] Domain: UML Statecharts	Series of experiments, positive influence on understandability in last experiment
Cruz-Lemus et al. [13] Domain: UML Statecharts	Hierarchy depth of statecharts has no influence
Shoval et al. [14] Domain: ER-Models	Hierarchy has no influence
Cruz-Lemus et al. [8] Domain: UML Statecharts	Positive influence on understandability for first experiment, negative influence in replication
Cruz-Lemus et al. [12, 19] Domain: UML Statecharts	Hierarchy depth has a negative influence

Table 1. Empirical studies into hierarchical structuring

2.2 Inference: A General-Purpose Problem Solving Process

As discussed in Sect. 2.1, the impact of hierarchy on understandability can range from negative over neutral to positive. To provide explanations for these diverse findings, we turn to insights from cognitive psychology. In experiments, the understandability of a conceptual model is usually estimated by the difficulty of answering questions about the model. From the viewpoint of cognitive psychology, answering a question refers to a *problem solving task*. Thereby, three different problem-solving *"programs"* or *"processes"* are known: search, recognition and inference [20]. Search and recognition allow for the identification of information of low complexity, i.e., locating an object or the recognition of patterns. Most conceptual models, however, go well beyond complexity that can be handled by search and recognition. Here, the human brain as a *"truly generic problem solver"* [21] comes into play. Any task that can not be solved by search or recognition, has to be solved by deliberate thinking, i.e., inference, making *inference* the *most important cognitive process* for understanding conceptual models. Thereby, it is widely acknowledged that the human mind is limited by the capacity of its working memory, usually quantified to as 7±2 slots [22]. As soon as a mental task, e.g., answering a question about a model, overstrains this capacity, errors are likely to occur [23]. Consequently, mental tasks should always be designed such that they can be processed within this limit; the amount of working memory a certain task thereby utilizes is referred to as *mental effort* [24].

In the context of this work and similar to [25], we take the view that the impact of modularization on understandability, i.e., the influence on inference, ranges from negative over neutral to positive. Seen from the viewpoint of cognitive psychology, we can identify two *opposing forces* influencing the understandability of a hierarchically structured model. Positively, hierarchical structuring can help to reduce the mental effort through *abstraction* by reducing the number of model elements to be considered at the same time [15]. Negatively, the introduction of sub-models may force the reader to *switch her attention* between the sub-models, leading to the so-called *split-attention effect* [26]. Subsequently, we will discuss how these two forces presumably influence understandability.

Abstraction. Through the introduction of hierarchy it is possible to group a part of a model into a sub-model. When referring to such a sub-model, its content is hidden by providing an abstract description, such as a complex activity in a business process model or a composite state in an UML statechart. The concept of abstraction is far from new and known since the 1970s as *"information hiding"* [1]. In the context of our work, it is of interest in how far abstraction influences model understandability. From a theoretical point of view, abstraction should show a positive influence, as abstraction reduces the amount of elements that have to be considered simultaneously, i.e., abstraction can hide irrelevant information, cf. [15]. However, if positive effects depend on whether information can be hidden, the way how hierarchy is displayed apparently plays an important role. Here, we assume, similar to [15, 17], that each sub-model is presented separately. In other words, each sub-model is displayed in a separate window if viewed on a computer, or printed on a single sheet of paper. The reader may arrange the sub-models according to her preferences and may close a window or put away a paper to hide information. To illustrate the impact of abstraction, consider the BPMN model shown in Fig. 1. Assume the reader wants to determine whether the model allows for the execution of sequence A, B, C. Through the abstraction introduced by sub-processes A and C, the reader can answer this question by looking at the top-level process only (i.e., activities A, B and C); the model allows to hide the content of sub-processes A and C for answering this specific question, hence reducing the number of elements to be considered.

Split-Attention Effect. So far we have illustrated that abstraction through hierarchical structuring can help to reduce mental effort. However, the introduction of sub-models also has its downsides. When extracting information from the model, the reader has to take into account several sub-models, thereby switching attention between sub-models. The resulting *split-attention effect* [26] then leads to increased mental effort, nullifying beneficial effects from abstraction. In fact, too many sub-models impede understandability, as pointed out in [4].

Again, as for abstraction, we assume that sub-models are viewed separately. To illustrate this, consider the BPMN model shown in Fig. 1. To assess whether activity J can be executed after activity E, the reader has to switch between the top-process as well as sub-processes A and C, causing her attention to split between these models, thus increasing mental effort.

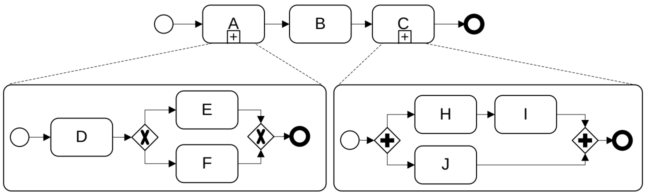

Fig. 1. Example of hierarchical structuring

While the example is certainly artificial and small, it illustrates that it is not always obvious in how far hierarchical structuring impacts a model's understandability.[5]

3 Assessing the Impact of Hierarchy

Up to now we discussed how the cognitive process of inferencing is influenced by different degrees of hierarchical structuring. In Sect. 3.1, we define a theoretical framework that draws on cognitive psychology to explain and integrate these observations. We also discuss the measurement of the impact of hierarchy on understanding in Sect. 3.2 along with its sensitivity to model size in Sect. 3.3 and experience in Sect. 3.4. Furthermore, we discuss the implications of this framework in Sect. 3.5 and potential limitations in Sect. 3.6.

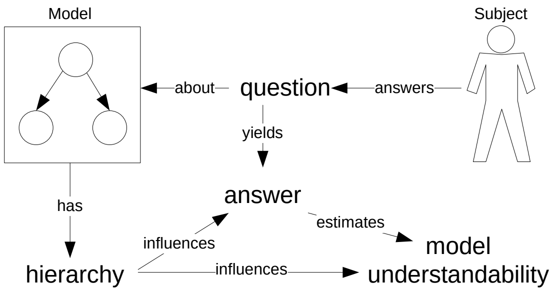

Fig. 2. Research model

3.1 Towards a Cognitive Framework

The typical research setup of experiments investigating the impact of hierarchy, e.g., as used in [8, 9, 15, 17, 18], is shown in Fig. 2. The posed research question

[5] At this point we would like to remark that we do not take into account class diagrams hierarchy metrics, e.g. [27], since such hierarchies *do not* provide abstraction in the sense we define it. Hence, they fall outside our framework.

thereby is how the *hierarchy* of a model influences *understandability*. In order to operationalize and measure model understandability, a common approach is to use the performance of answering questions about a model, e.g., accuracy or time, to *estimate* model understandability [9, 17, 18]. In this sense, a *subject* is asked to *answer questions* about a model; whether the model is hierarchically structured or not serves as treatment.

When taking into account the interplay of abstraction and split-attention effect, as discussed in Sect. 2.2, it becomes apparent that the impact of hierarchy on the performance of answering a question might not be uniform. Rather, each individual question may benefit from or be impaired by hierarchy. As the estimate of understandability is the average answering performance, it is essential to understand how *a single question* is influenced by hierarchy. To approach this influence, we propose a framework that is centered around the concept of mental effort, i.e., the load imposed on the working memory [24], as shown in Fig. 3. In contrast to most existing works, where hierarchy is considered as a dichotomous variable, i.e., hierarchy is present or not, we propose to view the impact of hierarchy as the result of two opposing forces. In particular, every *question* induces a certain *mental effort* on the reader caused by the question's complexity, also referred to as *intrinsic cognitive load* [23]. This value *depends* on model-specific factors, e.g., model size, question type or layout, and person-specific factors, e.g., experience, but is *independent* of the model's hierarchical structure. If hierarchy is present, the resulting mental effort is decreased by *abstraction*, but increased by the *split-attention effect*. Based on the resulting mental effort, a certain answering performance, e.g., accuracy or time, can be expected. In the following, we discuss the implications of this framework. In particular, we discuss how to measure the impact of hierarchy, then we use our framework to explain why model size is important and why experience affects reliable measurements.

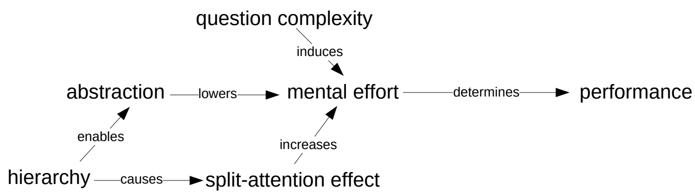

Fig. 3. Theoretical framework for assessing understandability

3.2 Measuring the Impact on Model Understandability.

As indicated [9, 8, 15, 17, 18] it is unclear whether and under which circumstances hierarchy is beneficial. As argued in Sect. 2.2, hierarchical structuring can affect answering performance positively by abstraction and negatively by the split-attention effect. To make this trade-off measurable *for a single question*, we provide an operationalization in the following. We propose to estimate the gains of abstraction by counting the number of model elements that can be "hidden" for answering a specific question. Contrariwise, the loss through the split-attention effect can be estimated by the number of context switches, i.e., switches between sub-models, that are required to answer a specific question. To illustrate the suggested operationalization, consider the UML statechart in Fig. 4. When answering the question whether sequence A, B is possible, the reader presumably benefits from the abstraction of state C, i.e., states D, E and F are hidden—leading to a gain of three (hidden model elements). On the contrary, when answering the question, whether the sequence A, D, E, F is possible, the reader does not benefit from abstraction, but has to switch between the top-level state and composite state C. In terms of our operationalisation, no gains are to be expected, since no model element is hidden. However, two context switches when following sequence A, D, E, F, namely from the top-level state to C and back, are required. Overall, it can be expected hierarchy *compromises* this question.

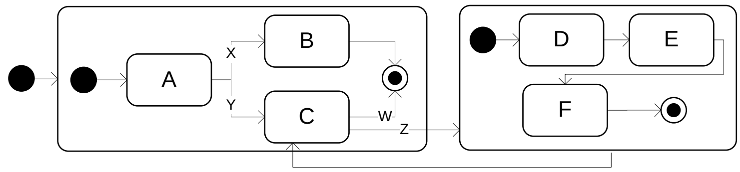

Fig. 4. Abstraction versus split-attention effect

Regarding the use of this operationalization we have two primary purposes in mind. First, it shall help experimenters to design experiments that are not biased toward/against hierarchy by selecting appropriate questions. Second, on the long run, the operationalization could help to estimate the impact of hierarchy on a conceptual model. Please note that these applications are to be viewed under some limitations as discussed in Sect. 3.6.

3.3 Model Size

Our framework defines two major forces that influence the impact of hierarchy on understandability: abstraction (positively) and the split-attention effect (negatively). In order that hierarchy is able to provide benefits, the model must be large enough to benefit from abstraction. Empirical evidence for this theory can be found in [9]. The authors conducted a series of experiments to assess the understandability of UML statecharts with composite states. For the first four experiments no significant differences between flattened models and hierarchical ones could be found. Finally, the last experiment showed significantly better results for the hierarchical model—the authors identified increased complexity, i.e., model size, as one of the main factors for this result. While it seems very likely that there is a certain complexity threshold that must be exceeded, so that desired effects can be observed, it is not yet clear where exactly this threshold lies. To illustrate how difficult it is to define this threshold, we would like to provide an example from the domain of business process modeling, where estimations range from 5–7 model elements [5] over 5–15 elements [6] to 50 elements [7]. In order to investigate whether such a threshold indeed exists and how it can be computed, we envision a series of controlled experiments. Therein, we will systematically combine different model sizes with degrees of abstraction and measure the impact on the subject's answering performance.

3.4 Experience

Besides the size of the model, the reader's experience is an important subject-related factor that should be taken into account [28]. To systematically answer why this is the case, we would like to refer to Cognitive Load Theory [23]. As introduced, it is known that the human working memory has a certain capacity, if it is overstrained by some mental task, errors are likely. As learning causes additional load on the working memory, novices are more likely to make mistakes, as their working memory is more likely to be overloaded by the complexity of the problem solving task in combination with learning. Similarly, less capacity is free for carrying out the problem solving task, i.e, answering the question, hence lower performance with respect to time is to be expected. Hence, experimental settings should ensure that most mental effort is used for problem solving instead of learning. In other words, subjects are not required to be experts, but *must be* familiar with hierarchical structures. Otherwise, it is very likely that results are influenced by the effort needed for learning. To strengthen this case, we would like to refer to [8], where the authors investigated composite states in UML statecharts. The first experiment showed *significant benefits* for composite states, i.e., hierarchy, whereas the *replication* showed *significant disadvantages* for composite states. The authors state that the *"skill of the subjects using UML for modeling, especially UML statechart diagrams, was much lower in this replication"*, indicating that experience plays an important role.

3.5 Discussion

The implications of our work are threefold. First, hierarchy presumably does not impact answering performance uniformly. Hence, when estimating model understandability, results depend on *which* questions are asked. For instance, when only questions are asked that do not benefit from abstraction, but suffer from the split-attention effect, a bias adversely affecting hierarchy can be expected. None of the experiments presented in Sect. 2.1 describes a procedure for defining questions, hence inconclusive results may be attributed to unbalanced questions. Second, for positive effects of hierarchy to appear, presumably a certain model size is required [9]. Third, a certain level of expertise is required that the impact of hierarchy instead of learning is measured, as to be observed in [8].

3.6 Limitations

While the proposed framework is based on established concepts from cognitive psychology and our findings coincide with existing empirical research, there are some limitations. First, our proposed framework is currently based on theory only, an empirical evaluation is yet missing. To counteract this problem, we are currently planning a thorough empirical validation, cf. Sect. 4. In this vein, also the operationalization of abstraction and split-attention effect needs to be investigated. For instance, we do not know yet whether a linear increase in context switches also results in a linearly decreased understandability, or the correlation can be described by, e.g., a quadratic or logarithmic behavior. Second, our proposal focuses on the effects on a single question, i.e., we can not yet assess the impact on the understandability of the entire model. Still, we think that the proposed framework is a first step towards assessing the impact on model understandability, as it is assumed that the overall understandability can be computed by averaging the understandability of all possible individual questions [29].

4 Summary and Outlook

We first had a look at studies on the understandability of hierarchically structured conceptual models. Hierarchy is widely recognized as viable approach to handle complexity—still, reported empirical data seems contradictory. We draw from cognitive psychology to define a framework for assessing the impact of hierarchy on model understandability. In particular, we identify abstraction and the split-attention effect as opposing forces that can be used to estimate the impact of hierarchy with respect to the performance of answering a question about a model. In addition, we use our framework to explain why model size is a prerequisite for a positive influence of modularization and why insufficient experience can bias measurement in experiments. We acknowledge that this work is just the first step towards assessing the impact of hierarchy on model understandability. Hence, future work clearly focuses on empirical investigation. First, the proposed framework is based on well-established theory, still, a thorough empirical validation is needed. We are currently preparing an experiment for verifying that the interplay of abstraction and split-attention effect can actually be observed in hierarchies. In this vein, we also pursue the validation and further refinement of the operationalization for abstraction and split-attention effect.

References

1. Parnas, D.L.: On the Criteria to be Used in Decomposing Systems into Modules. Communications of the ACM **15** (1972) 1053–1058
2. van der Aalst, W., ter Hofstede, A.H.M.: YAWL: Yet Another Workflow Language. Information Systems **30** (2005) 245–275
3. Davies, R.: Business Process Modelling With Aris: A Practical Guide. Springer (2001)
4. Damij, N.: Business process modelling using diagrammatic and tabular techniques. Business Process Management Journal **13** (2007) 70–90
5. Sharp, A., McDermott, P.: Workow Modeling: Tools for Process Improvement and Application Development. Artech House (2011)
6. Kock, N.F.: Product flow, breadth and complexity of business processes: An empirical study of 15 business processes in three organizations. Business Process Re-engineering & Management Journal **2** (1996) 8–22
7. Mendling, J., Reijers, H.A., van der Aalst, W.M.P.: Seven process modeling guidelines (7pmg). Information & Software Technology **52** (2010) 127–136
8. Cruz-Lemus, J.A., Genero, M., Manso, M.E., Piattini, M.: Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams. In: Proc. MODELS '05. (2005) 113–125

9. Cruz-Lemus, J.A., Genero, M., Manso, M.E., Morasca, S., Piattini, M.: Assessing the understandability of UML statechart diagrams with composite states—A family of empirical studies. Empir Software Eng **25** (2009) 685–719
10. Burton-Jones, A., Meso, P.N.: Conceptualizing systems for understanding: An empirical test of decomposition principles in object-oriented analysis. ISR **17** (2006) 38–60
11. Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. JSS **80** (2007) 571–583
12. Cruz-Lemus, J., Genero, M., Piattini, M.: Using controlled experiments for validating uml statechart diagrams measures. In: Software Process and Product Measurement. Volume 4895 of LNCS. Springer Berlin / Heidelberg (2008) 129–138
13. Cruz-Lemus, J., Genero, M., Piattini, M., Toval, A.: Investigating the nesting level of composite states in uml statechart diagrams. In: Proc. QAOOSE '05. (2005) 97–108
14. Shoval, P., Danoch, R., Balabam, M.: Hierarchical entity-relationship diagrams: the model, method of creation and experimental evaluation. Requirements Engineering **9** (2004) 217–228
15. Moody, D.L.: Cognitive Load Effects on End User Understanding of Conceptual Models: An Experimental Analysis. In: Proc. ADBIS '04. (2004) 129–143
16. Reijers, H., Mendling, J., Dijkman, R.: Human and automatic modularizations of process models to enhance their comprehension. Inf. Systems **36** (2011) 881–897
17. Reijers, H., Mendling, J.: Modularity in Process Models: Review and Effects. In: Proc. BPM '08. (2008) 20–35
18. Cruz-Lemus, J.A., Genero, M., Morasca, S., Piattini, M.: Using Practitioners for Assessing the Understandability of UML Statechart Diagrams with Composite States. In: Proc. ER Workshops '07. (2007) 213–222
19. Cruz-Lemus, J.A., Genero, M., Piattini, M., Toval, A.: An empirical study of the nesting level of composite states within uml statechart diagrams. In: Proc. ER Workshops. (2005) 12–22
20. Larkin, J.H., Simon, H.A.: Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive Science **11** (1987) 65–100
21. Tracz, W.J.: Computer programming and the human thought process. Software: Practice and Experience **9** (1979) 127–137
22. Miller, G.: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review **63** (1956) 81–97
23. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science **12** (1988) 257–285
24. Paas, F., Tuovinen, J.E., Tabbers, H., Gerven, P.W.M.V.: Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist **38** (2003) 63–71
25. Wand, Y., Weber, R.: An ontological model of an information system. IEEE TSE **16** (1990) 1282–1292
26. Sweller, J., Chandler, P.: Why Some Material Is Difficult to Learn. Cognition and Instruction **12** (1994) 185–233
27. Chidamber, S.R., Kemerer, C.F.: A metrics suite for object oriented design. IEEE Trans. Softw. Eng. **20** (1994) 476–493
28. Reijers, H.A., Mendling, J.: A Study into the Factors that Influence the Understandability of Business Process Models. SMCA **41** (2011) 449–462
29. Melcher, J., Mendling, J., Reijers, H.A., Seese, D.: On Measuring the Understandability of Process Models. In: Proc. BPM Workshops '09. (2009) 465–476