2010

An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users

Dominik Q. Birkmeier
University of Augsburg, dominik.birkmeier@wiwi.uni-augsburg.de

Sebastian Klöckner
University of Augsburg, sebastian.kloeckner@wiwi.uni-augsburg.de

Sven Overhage
University of Augsburg, sven.overhage@wiwi.uni-augsburg.de

An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users

Journal:	*18th European Conference on Information Systems*
Manuscript ID:	ECIS2010-0228.R2
Submission Type:	Research Paper
Keyword:	Business process management, Conceptual modeling, Unified modeling language, Systems analysis and design/development

AN EMPIRICAL COMPARISON OF THE USABILITY OF BPMN AND UML ACTIVITY DIAGRAMS FOR BUSINESS USERS

Birkmeier, Dominik, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany, dominik.birkmeier@wiwi.uni-augsburg.de

Klöckner, Sebastian, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany, sebastian.kloeckner@wiwi.uni-augsburg.de

Overhage, Sven, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany, sven.overhage@wiwi.uni-augsburg.de

Abstract

The widespread implementation of Business Process Management (BPM) strategies has increased the demand for an integral approach to business process modeling, in which all stakeholders can effectively participate and together shape a company's business processes. Amongst others, this demand was a basis for the development of the Business Process Modeling Notation (BPMN) as a proposed industry standard. It does not only provide technical advantages such as a support for service-oriented computing, but also claims to be readily usable for business users. Following this presumption, BPMN is even used by the Object Management Group (OMG). It adopted BPMN instead of the Activity Diagram (UML AD) as the core standard to create a business modeling framework. For companies, however, changing to a new process modeling language is a significant expense factor. Furthermore, consolidated findings on whether BPMN is indeed more usable for business users than UML AD are missing. In this paper, we present results from a comprehensive empirical comparison of both languages, in which we examined the application by business users during a model creation task. Results indicate that the UML AD is at least as usable as BPMN, since BPMN did neither differ significantly in effectiveness, efficiency, nor user satisfaction.

Keywords: BPMN, UML, Activity Diagram, Usability, Business users, Empirical study.

1 INTRODUCTION

Successfully implementing a Business Process Management (BPM) strategy considerably depends on establishing an integral approach to business process modeling, which allows diverse parties, such as managers, analysts, business users, and information system designers to participate and together optimize a company's business processes (Weske 2007). In such an approach, stakeholders require a process modeling language that can easily be used and understood by business and IT parties in order to communicate relevant process semantics. This demand, amongst others, led to the development of the Business Process Modeling Notation (BPMN). BPMN supporters claim that the language is not only well suited for system development purposes, but is also usable and understandable for "all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally to the business people who will manage and monitor those processes" (OMG 2006).

Following this argument, the Object Management Group's (OMG) Business Modeling & Integration Domain Task Force recently adopted BPMN as the core standard to develop a new business modeling framework around. This activity comprises, amongst others, the specification of a BPMN meta-model as well as the standardization of means to model business rules, organizational structures, business goals, etc. (OMG 2007). With its turn to BPMN, the OMG deliberately decided not to make use of the Unified Modeling Language (UML) and its process modeling notation, the UML Activity Diagram (UML AD). The UML AD was deemed as being too technically oriented (White 2004). However, as BPMN also has technical roots and the adoption of a newly proposed modeling language is a significant expense factor for companies, the promised advantages for business users have to be backed with solid arguments. While BPMN's technical features (e.g. the integration into the service-oriented computing technology) are unquestioned, the claimed advantages for business users remain to be proven, however. Why should BPMN be better usable for business users? Where did this opinion originate from? And is it justified?

To the best of our knowledge, the presumed superiority of BPMN over UML AD has not been substantiated with sound theoretical arguments or consolidated empirical findings. Instead, several authors who conducted analytical comparisons have highlighted considerable similarities between the languages (White 2004, Wohed & van der Aalst & Dumas & ter Hofstede & Russell 2006) or found BPMN to be more complex (Recker & zur Muehlen & Siau & Erickson & Indulska 2009). Therefore, it ought to be evaluated thoroughly whether BPMN is really more usable for business users and for which reasons this might be the case. In this paper, we examine and compare the usability of BPMN 1.1 and UML AD (UML 2.x) for business users on the basis of an empirical study. Thereby, we test the conservative hypothesis that *UML AD is at least as usable as BPMN* during a model creation task. We try to falsify the proposition and to confirm that BPMN is indeed more usable than UML AD. The empirical comparison is based upon a set of process models that has been created in a laboratory experiment. Building upon such a *confirmatory, quantitative-positivistic approach* (Creswell 2008, Popper 1980), we will proceed as follows: after presenting related work in the next section, we introduce relevant theories and concepts for our study in order to derive and refine our proposition. We then present our study in which we compared BPMN and UML AD to test our hypothesis. Finally, we present results from the conducted study, discuss them, and introduce implications for practice and academia.

2 RELATED WORK

The evaluation and comparison of conceptual modeling languages in general and process modeling languages in particular has frequently been addressed in literature. To get a complete picture, recommendations have been made to combine analytical with empirical approaches (Gemino & Wand

2003). So far, BPMN and its claim to be more usable for business users have mainly been analyzed from an analytical perspective, though. Some authors used a semiotic quality framework with linguistic evaluation categories to analyze BPMN (Nysetvold & Krogstie 2005, Wahl & Sindre 2005). On that basis, Wahl and Sindre (2005) concluded that BPMN "is easily learned for simple use", although especially its advanced modeling concepts (e.g. the variety of event types) are likely to compromise the usability for business users. Nysetvold and Krogstie (2005) did not only analyze BPMN but also compared it to UML AD. They found BPMN to be superior with respect to learnability, precision, and its language patterns. However, BPMN and UML AD were judged to be equally suited "to improve communication between the IT-department and the business departments". Their findings are limited in significance though, since both languages were ranked against a very simplistic weighting scheme.

As part of their survey, Recker et al. (2009) analyzed BPMN against the Bunge-Wand-and-Weber ontology. Overall, they confirmed BPMN to be a mature language, which is well suited for modeling business processes. Identified weaknesses referred to ambiguous language elements and ontological shortcomings, which, however, were classified to be not of immediate practical relevance. White (2004) and Wohed et al. (2006) used the workflow patterns as introduced by van der Aalst et al. (2003) to examine the expressive power of BPMN. They have shown the expressive power of BPMN to be comparable to those of established modeling languages and furthermore agree that there is a notable similarity between BMPN and UML AD constructs. White considered BPMN constructs to be more usable for business users since "although the UML 2.0 development included a more focused effort to upgrade the Activity Diagram in terms of its use for business people, it is still more technically oriented" (White 2004). However, he did not elaborate on why this might be the case.

Empirical evaluations of BPMN and especially on its usability for business users are still rare. While Recker and Dreiling (2007) have evaluated BPMN versus Event-Driven Process Chains (EPC, Dumas & Aalst & Hofstede 2005), they had a specific focus: to test teaching effects. They trained participants of the EPC group and tested their performance against untrained BPMN modelers, which makes it difficult to generalize their results. To the best of our knowledge, there is no empirical evaluation that focuses on confirming the claim that the usability of BPMN for business users is (a) higher than that of other process modeling notations in general and (b) higher than that of UML AD in particular.

3 THEORY AND PROPOSITIONS

A switch to a new process modeling language always results in significant investments into new tools, training of employees, translation of existing process models, etc. When taking into account that BPMN borrowed many concepts from existing languages as, e.g., UML AD, Event-Driven Process Chains, and Petri Nets (Weske 2007), it has to be questioned where the claimed better usability (Weske 2007, White 2004) comes from and if it really exists.

To identify the underlying reasons for a better usability of BPMN, the understanding of the term *usability* has to be clarified. In some cases usability is defined as a broad concept comparable to quality in use (Bevan 1995), while in other interpretations it is understood quite narrowly and distinguished from, for example, utility (Nielsen 1994). We adopted the definition of usability from the International Organization for Standardization (ISO), which explains the benefits in terms of user performance and satisfaction (ISO 1998). Furthermore, ISO defines usability as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Those aspects are further defined as follows (ISO 1998):

Effectiveness is the accuracy and completeness with which the users achieve the specified goals. Accuracy can be measured by the extent to which the quality of the output corresponds with the specified criteria, while completeness can be measured as the proportion of the target quantity, which has been achieved.

Efficiency is the level of effectiveness in relation to the expenditure of resources. These resources can include mental or physical effort, time, materials or financial cost.

Satisfaction is defined as the extent to which users are free from discomfort and their attitudes towards the use of the products. Amongst others, it can be assessed by asking the user to give a number corresponding to the strength of their feeling at any particular moment, or by asking users to rank products in order of preference.

Therefore, if a process modeling language is characterized as having a better usability, it has to be shown that the language is significantly better in at least one of those aspects. Efficiency, as relation between effectiveness and used resources, and satisfaction can be evaluated straightforwardly. In order to assess the effectiveness of a model creation task using a certain modeling language, the meaning of effectiveness and reasons for variations of effectiveness in the context of conceptual modeling have to be clarified and refined. As stated by the ISO, effectiveness in the sense of accuracy can be measured as quality of the outputs corresponding with specified criteria and therefore has to be interpreted as quality of the model in the context of conceptual modeling. Completeness, on the other side, does not seem to be directly applicable, as it is one criterion of model quality and the proportion of the target quantity cannot be directly determined for conceptual models.

According to Hadar and Soffer (Hadar & Soffer 2006, Soffer & Hadar 2003) and based on a model of Topi and Ramesh (2002), the quality of the model and variations of this quality have several determinants. The influencing factors and their interactions are shown in Figure 1 and briefly recapitulated in the following section as they can bias the results of empirical studies.

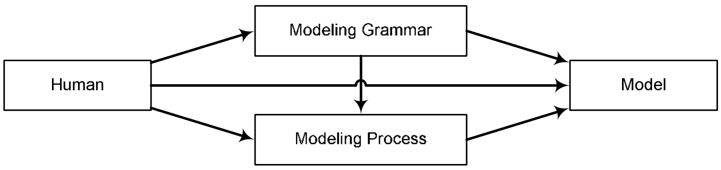

Figure 1. Factors that affect a conceptual model (Hadar & Soffer 2006).

The human factor that could influence the model results from the individual's perception and interpretation of reality, professional experience and the perception of model quality. But as long as differences in the human factor can be minimized or statistically balanced, e.g. by sample size, the human factor of individuals should not have an impact on the results of a study comparing the effectiveness of two given modeling languages.

The modeling grammar itself can cause different variations due to differences in its set of constructs involved and its expressive power. Expressive power describes completeness (i.e., including all the constructs required for representing the domain) and clarity (i.e., without problems of construct redundancy, excess and overload) (Wand & Weber 1993). When comparing the modeling elements (set of constructs) of UML AD and BPMN, many similarities become obvious. There are, however, also differences, particularly regarding the modeling of data objects, events, or the data flow between process steps. For example, if data elements are included in a BPMN model, which is essential for most business processes, they have to be separated from the control flow (OMG 2006). Moreover, BPMN generally contains fewer graphical constructs than UML AD and instead uses variations of them to support similar process patterns. E.g., BPMN uses similar elements to model events of different types or to depict parallel, exclusive, and inclusive gateways. Especially this reduced set of graphical constructs in combination with the clear separation of control and data flow in BPMN are often emphasized in literature and used as a rationale to claim its better usability for business users (Weske 2007, White 2004). Yet, it has to be questioned, if these similar constructs might have a negative impact on the expressive power of BPMN. As some constructs of BPMN can be regarded as overloaded, the clarity of the language might be reduced (Wand & Weber 1993). Therefore, it has to be validated if the claim of better usability, due to the reduced set of graphical constructs, outweighs the reduction of expressive power. For a comprehensive comparison of both languages, the interested reader is referred to Wohed et al. (2006) and White (2004).

The modeling process can be divided into two main phases: the perception of reality and the representation of the perceived reality in the model. The perception of reality can be influenced by human factors, as discussed above. The representation of the mental model of the application domain then depends on the mapping of reality into modeling constructs. Imprecise semantics of modeling constructs, like BPMN's intermediate events or the variety of gateway semantics, and vague rules defining how to map real world phenomena into the modeling constructs are likely to have an impact on model quality. In addition, when taking the limited cognitive capacity of humans (Gemino & Wand 2005) and the problems of apparent complexity (Gemino & Wand 2003) into account, the separation of control and data flow in BPMN could have a negative influence on resulting models quality, as information objects can easily be forgotten by the modeler.

As there are reasonable doubts deduced from theoretical concepts, the claim of superior usability of BPMN (Nysetvold & Krogstie 2005, Weske 2007, White 2004) might be contested. In order to substantiate this claim, an intergrammar comparison (Gemino & Wand 2004) would have to result in a falsification of the proposition:

P: For business users, UML Activity Diagrams are at least as usable as BPMN models.

Based on the discussed differentiation of usability, the stated proposition *P* can be further refined into:

P1: Business users will be at least as effective in modeling with UML AD as with BPMN,
P2: Business users will be at least as efficient in modeling with UML AD as with BPMN,
P3: Business users will be at least as satisfied with modeling with UML AD as with BPMN.

If a test of the stated propositions leads to a falsification of any single one of them and hence BPMN proofs to be better than UML AD in at least one category, it would justify a shift to BPMN.

4 RESEARCH METHOD

The experiment conducted to examine the stated propositions followed the design used by Batra et al. (1990). In a related research topic they evaluated representations with different data modeling techniques in an empirical examination. The main factor examined in the experiment is the notation used for model creation by the participants. Besides the modeling language, we additionally introduced three different training levels to simulate an environment of different experiences as it is usually found in practice (Recker 2008). The analysis, however, concentrates on *one source of variation*, namely the modeling method, within a *completely randomized design* (Cobb 1998, Dean & Voss 1999).

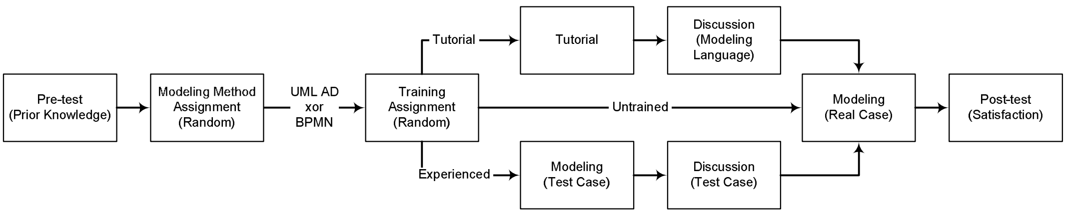

Figure 2. Design of experiment.

Figure 2 depicts the design of the experiment. It starts with a pre-test on prior domain and modeling knowledge, whose results were used to identify and exclude possible outliers afterwards. Two equally large groups were selected at random and allocated to different modeling techniques. Next, the participants in each group were randomly assigned to undergo different trainings. Depending on the assignments, short tutorials in BPMN or UML AD were provided for one third of the participants. Each of them took 45 minutes, included one small test case, was held by the same instructor and both were congruent with regard to their content and explanations. To simulate a second level of training, another third of the participants had to create a model of a complex test case without a detailed tutorial and were urged to discuss their individual experiences afterwards. Finally, the remaining participants did not receive any special training. The actual case was modeled by all participants at the same time. There was no time restriction; however, time was recorded for the following analysis. A survey on user satisfaction completed the experiment. Parts of the experiment were repeated with the same participants on a control case, which has marginal differences from the main case in its representational complexity (Bodart & Patel & Sim & Weber 2001).

Participants in the experiment were 30 graduate students in their final year with major or minor in business administration. They were randomly chosen out of over 80 volunteers for the study and split into two groups with 15 subjects each. Ex-ante interviews revealed that all of them had similar, only slightly differing backgrounds in conceptual modeling, mainly based on an undergraduate course that included modeling with Event-Driven Process Chains. Following Gemino and Wand (2005), we agree that the use of students is appropriate for such a type of study and can in fact be beneficial, since "prior knowledge on the problem solving (domain understanding) […] might have confounded the results". Furthermore, the incentive scheme used in the study aligns with Batra et al. (1990).

Several materials, in the interest of brevity not depicted in the paper, were provided to the participants of the study[1]. For the pre-test, a survey, aiming on identifying prior knowledge, had to be filled out. During the model creation task, participants were supplied with one sheet of detailed working instructions, a complete description of the process in natural language and several large empty pages to prepare the models. In addition, each participant received four pages of information on the application of their assigned modeling method, containing the available modeling primitives and common patterns, as well as one end-to-end example. These had been independently checked, compared and validated by several faculty members with experience in both modeling techniques.

Prior knowledge on the utilized business domain by some participants "might create substantial difficulties in an experimental study" (Gemino & Wand 2004). Thus, we decided to choose a process from a domain that was equally well-known to all participants. As shown in Table 1, the case contains all basic control flow patterns as introduced by van der Aalst et al. (2003), and one of their structural patterns. Its complexity is comparable to most of the reference business processes found in the German standard reference on business information systems for e-commerce (Becker & Schütte 2004).

Primitive		**Basic control flow pattern**		**Structural pattern**	
18	Flow element	3	Sequence (regular)	1	Arbitrary cycle
12	Data element	1	Sequence (conditional)		
		1	Parallel split		
		4	Synchronization		
		3	Exclusive choice		
		3	Simple merge		

Table 1. Case characteristics (after van der Aalst et al. (2003)).

To gather information on the user satisfaction with their respective modeling notation, a post-test survey was conducted. Figure 3 depicts the respective four questions of the post-test survey. On the first question, a high value indicates a highly satisfied user, whereas on the last three questions small values are favorable.

Q1 - Do you think you have understood the modeling language thoroughly?	(not at all) 1 … 2 … 3 … 4 … 5 … 6 (completely)
Q2 - Do you think the modeling language is challenging for you?	(not at all) 1 … 2 … 3 … 4 … 5 … 6 (highly)
Q3 - Do you think the concept of the modeling language is difficult?	(not at all) 1 … 2 … 3 … 4 … 5 … 6 (highly)
Q4 - Do you think the application of the modeling language is difficult?	(not at all) 1 … 2 … 3 … 4 … 5 … 6 (highly)

Figure 3. Post-test survey on user satisfaction.

[1] All Materials, including the sample solutions of the processes, can be requested from the authors.

During the statistical analysis we mainly applied the programming language and software environment *R* (Crawley 2007) for statistical computing and graphics. In addition, we also utilized *SPSS* (Norusis 2008) for various calculations and *GGobi* (Cook & Swayne & Buja 2007) for interactive graphics. Hypothesis testing was primarily performed via *Student's t tests*. Where necessary, *Kolmogorow-Smirnow tests* and *Bartlett's test* helped to check for normality and equal variance conditions.

5 RESULTS

5.1 Data scoring

Three raters graded the prepared models from each subject for correctness by comparing it with the adequate solution (Batra et al. 1990). Sample solutions were prepared by four experienced modelers independently and afterwards discussed and matched against each other. Since every model is prepared from a subjective view and hence there is no one solution (Hadar & Soffer 2006), tolerated variations from the sample solutions were defined. Only differentiations exceeding the defined tolerance are marked as failures. The inter-rater reliability was close to a hundred percent and the few remaining differences were discussed until a common grading was found.

In order to assess the quality of the resulting models the grading of the prepared models had to be operationalized and concrete failure types needed to be defined. Gemino and Wand (2004) propose accuracy, correctness, detail, completeness, quality and discrepancies, rated by experts, as possible measures for affected (outcome) variables. Yet, they do not describe how the criteria can be measured. The "Guidelines of Modeling (GoM)", proposed by Schuette and Rotthowe (1998), on the other side suggest construction adequacy, language adequacy, economic efficiency, clarity, systematic design and comparability. Both propositions cannot be exactly matched with each other and not all of the proposed criteria are applicable to the experiment. Based on our experience in evaluating exams of courses on Event-Driven Process Chains as well as an analysis of errors found in these solutions, we developed a grading scheme for evaluating the conceptual models (as Batra et al. (1990)). We decided to split our evaluation criteria into a language part, for all language-related errors, and an application part, for all errors due to a wrong application of the modeling language. This splitting follows the model of Hadar et al. (2006), who use modeling grammar and modeling process as factors influencing the model quality. In addition, most of the criteria presented by Gemino and Wand (2004) as well as Schuette and Rotthowe (1998) can be mapped onto our criteria. The composition of the criteria as used in the grading scheme is shown in Figure 4. The scheme has been evaluated by further faculty members, experienced in teaching process modeling (Batra et al. 1990). Any necessary changes were discussed and implemented.

As with all languages, violations of the notational correctness generally become manifest in syntax, semantic or pragmatic errors (Silverstein 1972) and were therefore attributed to the language part. For the *language* criteria *syntax* and *semantic*, we separated between single and repetitive errors. More than three consecutive, equal errors were counted as one repetitive error. An incorrect modeling of a flow pattern that is due to a wrong, while allowed, combination of language elements (e.g. a wrongly modeled loop) formed a *pragmatic* error.

For the application part, we merged the proposed criteria of Gemino and Wand (2004), Schuette and Rotthowe (1998), as well as our own experience into adequacy, consistency, strictness, granularity and data handling as evaluation criteria for the successful application of the modeling language. A model can be defective with respect to its *adequacy* by comprising redundancies or a limited flexibility. The flexibility of a process is compromised, whenever independent activities are or have to be described as sequences. The *consistency* is compromised by building contradictions or differing denominations into the model and can be compared to naming conventions and clarity (Schuette & Rotthowe 1998), as well as discrepancies (Gemino & Wand 2004). *Data handling* was introduced to account for informational equivalency (Gemino & Wand 2004), as BPMN emphasizes the separation of data flows, while

UML does not make such a distinction. It is biased by including superfluous data elements, leaving out relevant data elements or omitting data flows. The *granularity* is impacted by either leaving out or unnecessarily splitting up individual flow elements. It accounts for errors concerning minimalism and degree of abstraction in the sense of Schuette and Rotthowe (1998), as well as detail and completeness in the view of Gemino and Wand (2004). *Strictness* collects the errors, which are due to wrong mappings of reality onto modeling constructs, e.g. by introducing wrong constraints for workflows or omitting workflows, and could be compared to correctness (Gemino & Wand 2004).

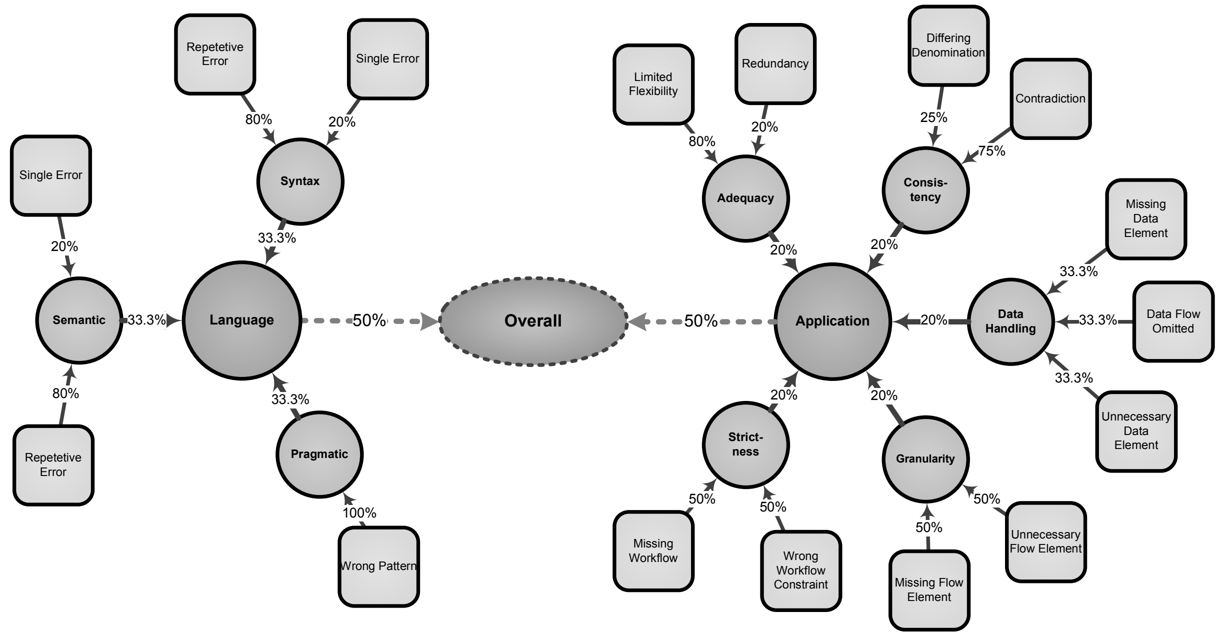

Figure 4. Criteria and aggregation.

To evaluate and compare usability, measures for effectiveness, efficiency and satisfaction were defined. As argued, the *effectiveness* is determined by the model quality, which can be measured in several predefined criteria. In each model, we identified and counted single failures. However, to evaluate the model quality through the main criteria, aggregated scores had to be built. First, the counted occurrences of errors were scaled to each form a score between 0 (worst) and 100 (best). In doing so, a score of 100 in one of the single failure categories is equal to zero errors. A score of 0 is assigned for the highest amount of single errors in one category over all models. Such a score is assigned to each model in all of the single failure categories. The approach uses a linear transformation and, thus, no information is changed or lost. Since the single failure categories are less capable to provide an overview on the different modeling techniques, they are aggregated to form the defined evaluation criteria depicted in Figure 4. Most of the applied weights are balanced, except where equal weights would be unreasonable. In a second step, e*fficiency*, as defined above, can be measured by the achieved score-points per time. Since the time needed to complete the model creation task was recorded (in minutes), the efficiency can be calculated for every criterion by dividing the effectiveness scores through time. The corresponding unit is “points per minute”. Furthermore, each question of the post-test survey was recorded on a scale from 1 to 6 and, hence, can be directly used to measure user *satisfaction*.

5.2 Data analysis

The three propositions, concerning effectiveness, efficiency and satisfaction, were tested on the collected data. The t test for two independent samples of equal size assumes Gaussian variables in both groups. Furthermore, it is applicable for equal and unequal variances in the groups, though the former situation is preferable. Therefore, Bartlett’s test can be performed to prejudge the equality of variances, but it strongly relies on the condition of both groups following a normal distribution. Hence, if the normality assumption is violated, equality of variances cannot be reliably evaluated (marked with *n/a* in the following analysis). In such cases, the more conservative version of the t test for unequal variances is applied. Although t tests are rather robust towards violations of its preconditions (Boneau 1960), test results have to be interpreted more carefully whenever an assumption is violated.

The effect of intervening variables was examined using an analysis of covariance (ANCOVA) technique (Kutner & Nachtsheim & Neter & Li 2005). Two possible covariate factors might have an influence on the analysis: prior domain knowledge and the assigned training. Domain knowledge information was recorded in the surveys. The analysis showed that domain knowledge was comparable between subjects and had no significant influence on the results. Therefore, it is safe to be left disregarded in the further analysis. The different experience levels, which were intentionally introduced and monitored through the assigned trainings, constitute the second possible covariate factor. The analysis revealed that the experience levels have neither an influence on the effectiveness nor on the satisfaction scores. On the other hand, in compliance with our expectations, the experience has a significant influence on the time needed for the model creation task. Consequently, the training has an impact on the efficiency scores (points per minute) as well. Nevertheless, as our analysis showed, the influence of the factor *modeling method* is independent from a possible inclusion of the experience level. Thus, it is safe to be removed from the further analysis as well.

		Effectiveness									Efficiency								
Variable		ND [1]	EV [2]	t test [3]		Summary Statistics [6]					ND [1]	EV [2]	t test [3]		Summary Statistics [7]				
				t	p-value	Min	Max	Mean	Med [4]	SD [5]			t	p-value	Min	Max	Mean	Med [3]	SD [4]
Overall	AD	✓	✓	-1,331	0,903	95	99	97,3	97	1,29	✓	✓	-0,825	0,792	0,97	2,14	1,39	1,25	0,373
	BPMN	✓				94	98	96,7	97	1,45	✓				0,84	1,92	1,28	1,27	0,305
Language	AD	✓	✓	-0,541	0,703	94	100	98,1	99	2,09	✓	✓	-0,806	0,786	0,97	2,16	1,40	1,26	0,377
	BPMN	✓				92	100	97,6	99	2,61	✓				0,82	1,94	1,29	1,27	0,309
Application	AD	✓	✓	-1,241	0,888	94	99	96,8	97	1,52	✓	✓	-0,856	0,800	0,98	2,12	1,38	1,25	0,371
	BPMN	✓				93	99	96,1	96	1,71	✓				0,87	1,89	1,27	1,27	0,299
Syntax	AD	×	n/a	0,318	0,376	94	100	98,5	99	1,60	✓	✓	-0,712	0,759	0,99	2,16	1,40	1,26	0,373
	BPMN	✓				94	100	98,7	99	1,84	✓				0,83	1,96	1,31	1,30	0,318
Semantics	AD	✓	n/a	0,741	0,233	94	100	98,8	100	2,11	✓	✓	-0,704	0,756	0,97	2,16	1,40	1,28	0,373
	BPMN	×				97	100	99,3	100	1,22	✓				0,86	1,96	1,32	1,30	0,309
Pragmatics	AD	✓	✓	-0,970	0,830	86	100	96,8	100	5,03	✓	✓	-0,985	0,833	0,93	2,17	1,38	1,27	0,386
	BPMN	✓				83	100	94,9	97	5,85	✓				0,76	1,89	1,25	1,21	0,306
Adequacy	AD	✓	✓	-2,131	0,979	91	100	97,4	98	2,44	✓	✓	-0,973	0,830	0,99	2,07	1,38	1,26	0,364
	BPMN	✓				94	100	95,7	95	1,99	✓				0,88	1,85	1,27	1,27	0,285
Consistency	AD	✓	n/a	1,125	0,136	91	100	97,9	100	3,08	✓	✓	-0,654	0,741	0,97	2,17	1,39	1,27	0,378
	BPMN	×				94	100	99,0	100	2,00	✓				0,88	1,96	1,31	1,27	0,311
Strictness	AD	×	n/a	-1,987	0,971	97	100	99,1	100	1,39	✓	✓	-0,877	0,806	1,00	2,10	1,41	1,27	0,379
	BPMN	✓				93	100	97,8	98	2,04	✓				0,87	1,96	1,30	1,30	0,318
Granularity	AD	✓	✓	0,603	0,276	81	100	94,0	95	4,47	✓	✓	-0,658	0,742	0,90	2,10	1,34	1,19	0,378
	BPMN	✓				88	98	94,9	95	3,31	✓				0,81	1,86	1,26	1,23	0,300
Data Handling	AD	✓	✓	-2,033	0,974	90	99	96,1	98	3,44	✓	✓	-1,119	0,864	0,98	2,15	1,36	1,27	0,365
	BPMN	✓				82	99	93,1	93	4,56	✓				0,86	1,83	1,23	1,23	0,290

[1] Normally Distributed, [2] Equal Variances, [3] One-tailed, [4] Median, [5] Standard Deviation, [6] Unit: points (0-100), [7] Unit: points per minute

Table 2. Effectiveness and Efficiency – Tests and summary statistics.

Proposition 1. The left side of Table 2 contains an overview of descriptive statistics and testing results for the effectiveness measures. Considering the p-values for the one-tailed t test, it becomes obvious that the results are unambiguous, as for none of the criteria the difference is close to being significant. Thus, the stated *proposition cannot be rejected* and it can be concluded that *business users are at least as effective in modeling with UML AD as with BPMN.* A closer look at the results reveals that the UML AD group has higher means in language and application scores, as well as in four of the lower aggregated criteria. Wherever differences in medians are present, UML AD scores are higher as well.

While the language scores are rather equal, Figure 5 reveals that especially the application of UML AD seems to be more successful. An examination of the contrary hypothesis, of BPMN being at least as effective for business users as UML AD, leads to deeper insights. Such a hypothesis will be significantly rejected for the criteria adequacy, strictness and data handling.

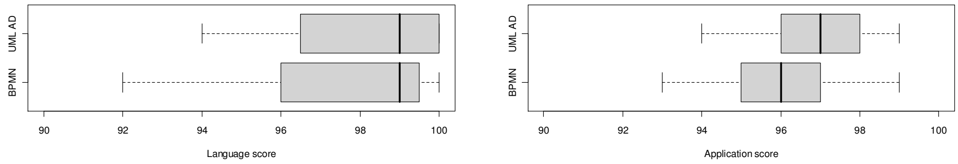

Figure 5. Effectiveness – Boxplots of language and application scores.

Proposition 2. As discussed, the basis of all efficiency indicators is the time needed to finish the model creation task. The amount of time needed is more spread out for UML AD modelers than for participants using BPMN. Means and medians of both groups are between 74 and 80 minutes. A t test showed no significant difference in model creation time between BPMN and UML AD (p-value: 0.492). The statistical results on efficiency shown in the right side of Table 2 are even more distinct than those for effectiveness. All reported p-values are larger than 74% and thus, they are far from rejecting the claimed proposition. For every single criterion, the means in the UML AD group are higher than those in the BPMN group. Overall, business users modeling with UML AD are at least equally efficient as with BPMN.

Proposition 3. Finally, the examination of the post-test survey results (detailed table not depicted due to reasons of brevity) gives information about the average satisfaction of business users with the different modeling notations. As discussed, a high value on question 1 and low values on the other three are favorable. For all questions the means and medians of BPMN and UML AD are relatively close to each other and thus no clear tendency can be seen. This observation is validated by the p-values of one-tailed t tests, which range between 0.39 and 0.79 and, hence, are far from indicating significant differences. However, since the reliability of the t test is reduced, due to the interval scaled data (levels 1 to 6), the insignificance of the results was additionally confirmed by ordinal logistic regression tests (Kutner et al. 2005). Consequently there is no reason to discard the stated proposition and it can be concluded that business users are at least as satisfied with UML AD as with BPMN.

As we failed to reject any of the refined propositions (*P1, P2* and *P3*), the main proposition *P* (*UML Activity Diagrams are at least as usable as BPMN models*) cannot be falsified as well. An examination of the control process supported these results. The study we performed can easily be replicated and, hence, an independent confirmation of the results is possible. However, as for any empirical study there are some limitations as to what extent the results can be generalized. First of all, the number of samples could be larger. We plan to increase the sample size and validate our results in further experiment settings in order to increase the external validity of our findings. Additionally, more complex cases might be different in usability for business users, although in our opinion this would only strengthen our observations and might even support the findings in categories, where the UML AD is already significantly better than BPMN.

6 CONCLUSIONS AND IMPLICATIONS

In this paper we evaluated the usability of BPMN and UML AD based on a comprehensive empirical comparison. Starting with a comparative discussion of both languages, we deduced several doubts concerning the claimed superior usability of BPMN compared to UML AD for business users. In order to favor BPMN, the stated proposition would have to be falsified in at least one of the deduced aspects of usability (being effectiveness, efficiency and satisfaction). After describing the study design, we provided several findings that are suited to judge the proposition. The discussed empirical results show that BPMN was not able to falsify any of the sub-propositions. In contrast, extended examinations revealed that UML AD was significantly more effective in the criteria *data handling* and *adequacy*.

With respect to the modeling of flexible processes, in which self-contained activities should preferably be allowed to run in parallel, UML AD turned out to be superior. The usage of BPMN instead promoted a rather sequential modeling style in which unrelated activities run one after the other. Such a modeling unnecessarily degrades the process flexibility. Another remarkable observation concerns the separation of control and data flow in BPMN, which apparently mislead participants to leave out parts of the data flow. Originally being introduced as a means to separate concerns and reduce the modeling complexity (Weske 2007), this concept turned out to be inferior to a combined flow modeling as present in UML AD. Admittedly, we could not confirm our theoretically deduced doubt that the decision of reducing the set of modeling constructs and instead introducing variants, would lead to a higher rate of mistakes, which stemmed from confusion due to a reduction of grammar clarity. This aspect remains to be examined more closely.

Our results have implications for both practice and academia. For practice, the results showed that for business users a higher usability of BPMN compared to UML AD cannot be empirically supported. Although, in literature BPMN is currently often claimed to be more useable (Nysetvold & Krogstie 2005, Weske 2007, White 2004) and even standardization organizations such as the OMG seem to have followed that conclusion, there are indications that BPMN still has shortcomings, which are likely to hinder its efficient adoption by business users in practice. And where business users are unable to use a modeling language adequately, the communication between the various stakeholders is compromised. Taking into account that our results also revealed that the use of BPMN implied a decrease of process flexibility, a misfit with the basic idea of optimizing business processes becomes obvious. Therefore, especially in environments where the technical advantages of BPMN are not needed, or not usable because of technical limitations, a possible switch to BPMN should be carefully deliberated.

Future research should concentrate on further examining the empirical indications identified in this paper. In order to definitely judge the usability of BPMN and UML AD for business users, a deeper understanding of their individual strengths and weaknesses has to be gained. Following the framework of Gemino and Wand (2004), additional empirical studies of model creation, as well as model interpretation tasks, should be conducted. The consolidated findings of such studies could provide a basis for merging BPMN and AD, which is eventually being planned in future to form a truly unified process modeling notation that combines the strengths of both languages (White 2004). Therefore, it seems reasonable that efforts towards a version 2.0 of BPMN should consider already gained experiences and proven findings to support the primary goal of BPMN: "to provide a notation that is readily understandable by all business users" (OMG 2006).

References

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

Becker, J. and R. Schütte (2004). Handelsinformationssysteme. 2. Ed. Verlag Moderne Industrie, Frankfurt.

Bevan, N. (1995). Measuring usability as quality of use. Software Quality Journal, 4, 115-150.

Bodart, F., A. Patel, M. Sim and R. Weber (2001). Should Optional Properties Be Used in Conceptual Modelling: A Theory and Three Empirical Tests. Information Systems Research, 12 (4), 384-405.

Boneau, C.A. (1960). The Effects of Violations of Assumptions Underlying the t Test. Psychological Bulletin, 57 (1), 49-64.

Cobb, G.W. (1998). Introduction to Design and Analysis of Experiments. Springer, New York, NY.

Cook, D., D.F. Swayne and A. Buja (2007). Interactive and Dynamic Graphics for Data Analysis: With R and GGobi. Springer, New York, NY.

Crawley, M.J. (2007). The R Book. Wiley, Hoboken, NJ.

Creswell, J.W. (2008). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 3. Ed. Sage Publications, Thousand Oaks, CA.

Dean, A. and D. Voss (1999). Design and Analysis of Experiments. Springer, New York, NY.

Dumas, M., W.M.P.v.d. Aalst and A.H.M.t. Hofstede (2005). Process-Aware Information Systems. Bridging People and Software Through Process Technology. Wiley, Hoboken, NJ.

Gemino, A. and Y. Wand (2003). Evaluating Modeling Techniques based on Models of Learning. Communications of the ACM, 46 (10), 79-84.
Gemino, A. and Y. Wand (2004). A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering, 9 (4), 248-260.
Gemino, A. and Y. Wand (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. Data & Knowledge Engineering, 55, 301–326.
Hadar, I. and P. Soffer (2006). Variations in Conceptual Modeling: Classification and Ontological Analysis. Journal of the Association for Information Systems, 7 (8), 569-593.
ISO (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11 : Guidance on usability. ISO Standard, ISO 9241-11:1998(E), International Standard Organization.
Kutner, M.H., C.J. Nachtsheim, J. Neter and W. Li (2005). Applied Linear Statistical Models. 5. Ed. McGraw-Hill/Irwin, Boston, MA.
Nielsen, J. (1994). Usability Engineering. Morgan Kaufmann, San Francisco.
Norusis, M.J. (2008). SPSS 16.0 Guide to Data Analysis. 2. Ed. Prentice Hall, Upper Saddle River.
Nysetvold, A.G. and J. Krogstie (2005). Assessing Business Processing Modeling Languages Using a Generic Quality Framework. In CAiSE'05 Workshops, pp. 545-556.
OMG (2006). Business Process Modeling Notation Specification. Final Adopted Specification, dtc/06-02-01, Object Management Group.
OMG (2007). Business Process Model and Notation (BPMN) 2.0. Request for Proposal, BMI/2007-06-05, Object Management Group.
Popper, K.R. (1980). The Logic of Scientific Discovery. Hutchinson, London.
Recker, J. (2008). BPMN Modeling - Who, Where, How and Why. BPTrends, 5 (3), 1-8.
Recker, J. and A. Dreiling (2007). Does It Matter Which Process Modelling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages without Formal Education. In Proceedings of the ACIS 2007.
Recker, J., M. Indulska, M. Rosemann and P. Green (2009). Business Process Modeling - A Comparative Analysis. Journal of the Association for Information Systems, 10 (4), 333-363.
Recker, J., M. zur Muehlen, K. Siau, J. Erickson and M. Indulska (2009). Measuring Method Complexity: UML versus BPMN. In Proceedings of the AMCIS 2009.
Schuette, R. and T. Rotthowe (1998). The Guidelines of Modeling - an approach to enhance the quality in information models. In ER '98, pp. 240-254, LNCS 1507, Springer-Verlag.
Silverstein, M. (1972). Linguistic Theory: Syntax, Semantics, Pragmatics. Annual Review of Anthropology, 1, 349-382.
Soffer, P. and I. Hadar (2003). Reusability of Conceptual Models: The Problem of Model Variations. In EMMSAD, Proceedings of the Eighth CAiSE/IFIP8.
Topi, H. and V. Ramesh (2002). Human factors research on data modeling: a review of prior research, an extended framework and future research directions. Advanced Topics in Database Research, 3, 188-217.
van der Aalst, W.M.P., A.H.M. ter Hofstede, B. Kiepuszewski and A.P. Barros (2003). Workflow Patterns. Distributed and Parallel Databases, 14 (1), 5-51.
Wahl, T. and G. Sindre (2005). An Analytical Evaluation of BPMN Using a Semiotic Quality Framework. In CAiSE'05 Workshops, pp. 533-544.
Wand, Y. and R. Weber (1993). On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. Journal of Information Systems, 3 (4), 217-237.
Weske, M. (2007). Business Process Management: Concepts, Languages, Architectures. Springer, Berlin, Heidelberg.
White, S.A. (2004). Process Modeling Notations and Workflow Patterns. In The Workflow Handbook 2004 (Fischer, L. Ed., pp. 265-294, Future Strategies Inc., Lighthouse Point, FL.
Wohed, P., W.M.P. van der Aalst, M. Dumas, A.H.M. ter Hofstede and N. Russell (2006). On the Suitability of BPMN for Business Process Modelling. In BPM 2006, pp. 161-176, LNCS 4102, Springer.