

# The Influence of Syntactic Quality on Pragmatic Quality of Enterprise Process Models

Merethe Heggset<sup>1</sup>, John Krogstie<sup>1</sup> and Harald Wesenberg<sup>2</sup>

<sup>1</sup>Norwegian University of Science & Technology - NTNU, Norway

<sup>2</sup>Statoil ASA, Norway

[merethhe@gmail.com](mailto:merethhe@gmail.com), [krogstie@idi.ntnu.no](mailto:krogstie@idi.ntnu.no), [hwes@statoil.com](mailto:hwes@statoil.com)

**Abstract.** As approaches and tools for process and enterprise modelling are maturing, these techniques are being taken into use on a large scale in an increasing number of organizations. In this paper we report on the use of process modelling in connection to the quality system of Statoil, a large Norwegian oil company, in particular, on the aspects found necessary to be emphasized to achieve the appropriate quality of the models in this organization. Based on the investigation of usage statistics and user feedback on models, we have identified that there are problems in comprehending some of the models. Some of these models has poorer syntactic quality than the average syntactic quality of models of the same size. An experiment with improving syntactic quality on some of these models has given mixed results, and it appears that certain syntactic errors hinder comprehension more than others.

**Keywords:** Enterprise process models, experiment, model quality.

## 1 Introduction

Statoil is a Norwegian oil company with more than 23000 employees and around the same number of external contractors. Statoil operates in 36 different countries all over the world and have in the last decade been using enterprise modelling in order to structure their vast amounts of organizational knowledge and information. They report to have achieved a fair amount of success with enterprise modelling in its corporate management system [1] where workflow models are used extensively to communicate requirements and best practices throughout the enterprise. The enterprise model functions as a common point of reference for the entire organization, ensuring the quality of a large number of work processes and communicating requirements and best practices throughout the company. The models are used daily in large parts of the organization, and are a significant contributor in reducing operational, environmental and safety risks. To illustrate, the for Statoil important SIF-index (Serious Injury Frequency) which counts the number of incidents per million work hours has been reduced from 6 to around 0.8 in the period since the models were introduced. Every week Statoil employees and contractors perform approximately 2 million work hours. That said, the process models are only one approach to risk mitigation. One also experiences that the process models could be utilized even better.

A lot of research has been done in the field of enterprise process modelling, as well as on the subject of how to evaluate model quality [2], [3], [4], [5]. Much work is done regarding the use and creation of models on a theoretical level, but in order to better understand the mechanisms at work in the application of enterprise models, real-life cases provide interesting insights. How

enterprise models are actually used within an organization will vary from case to case, so collecting as much information as possible about this from several sources seems appropriate and useful.

As referenced above, many frameworks and methods for model quality assessment has been developed over the years. However, as stated by Moody [3], many of these methods suffer from a lack of adoption in practice. While the main goal of applying such frameworks in practice normally is providing a detailed evaluation of model quality in a specific case, it can also give indications of the usefulness of the framework and, based on the results, possibly enforce its position in the field which again may lead to a wider adoption in practice.

From the start of the current modeling initiative, Statoil has been aware of the need to balance different levels of quality of the models. According to [1], Statoil have found that it is useful to differentiate between at least three dimensions of model quality [6]: Syntactic quality (how well the model uses the modelling language), semantic quality (how well the model reflects the part of the world it represents) and pragmatic quality (how well the model is understood by the target audience), building upon distinctions first described in [7], which is a predecessor to the current SEQUAL framework on quality of models and modelling languages [2]. In enterprise models the balance between these dimensions becomes very important based on the goal of modelling; else the model will not be used by its intended target audience in the right way. In our analysis, we have applied SEQUAL.

This paper presents some of the results from an ongoing case study on the use of enterprise process models in Statoil, in particular, looking upon model usage, quality issues of existing models, and how better syntactic quality can influence the pragmatic quality (comprehension) of models. The main research question we have investigated in connection to this paper is” How do different levels of quality influence each other?” In particular, we look upon the influence of syntactic quality on pragmatic quality, including what is useful to state as guidelines on the syntactic level.

General background on SEQUAL is provided in Section 2. In Section 3 we describe the process models of the Statoil quality system in more detail, before we in Section 4 describe experiences from evaluations of the current models and an experiment on improving the syntactic quality of existing operational models. This extends the work presented in [8] that only reported results from investigating one model. Discussion of results, concluding remarks and ideas on further work on understanding the trade-off on different quality aspects are reflected in Section 5.

## 2 Background on Modelling and Quality of Models

SEQUAL is a quality framework used for assessing the quality of models and modeling languages. The choice of using SEQUAL as an analytical lens for studying the Statoil enterprise model is mainly based on the fact that the company has addressed aspects of the enterprise model in the context of the three core quality levels of SEQUAL (syntactic, semantic and pragmatic) as also reported in the earlier work [1]. Krogstie and Arnesen [9] used a specialization of SEQUAL to evaluate various process modelling languages for use in Statoil. SEQUAL builds on early work on quality of models, but has been extended based on theoretical results [3], [4], [5] and practical experiences [2], [10], [11] with various extensions of the framework. It has earlier been used for evaluation of modelling and modelling languages of a large number of perspectives, including data, ontologies, process, enterprise, topological and goal-oriented modelling [2]. Quality has been defined referring to the correspondence between statements belonging to the following sets:

- **G**, the set of goals of the modelling task. The goals of modelling are many, and may vary greatly. Nysetvold and Krogstie outlines five main usage areas of enterprise models [12] (partly inspired by the Statoil PAKT taxonomy [13] and general model theory [14]):

- Human sense-making and communication: Actors can use the enterprise model to make sense of various aspects of the enterprise, and best practices and requirements can be communicated throughout the organization to create a common understanding.
  - Computer-assisted analysis: Models can be used, e.g., for simulation of improvements of process changes.
  - Business process management and quality assurance: Models can be used for quality assurance of work processes (e.g., ensuring compliance to regulations).
  - Model deployment and activation: The model can be deployed directly to be used for controlling, supporting and performing work. The activation can be either manual, automatic or interactive.
  - To give context for other tasks such as supporting system development projects.
- **D**, the domain, i.e., the set of all statements that can be stated about the situation. The goal of modelling (**G**) typically restricts the domain to only those things relevant to achieve this/these goal(s).
  - **L**, the language extension, i.e., what can be expressed by the modelling language chosen.
  - **M**, the externalized model itself.
  - **A**, what the social and technical actors involved in modelling have access to of the model **M**.
  - **K**, the explicit knowledge that the audience (both modelers and model interpreters) have of the domain.
  - **I**, the social actor (human) interpretation of the model.
  - **T**, the technical actor (tool) interpretation of the model.

The main quality types are:

- Physical quality: The basic quality goal is that **A** includes the relevant parts of the externalized model **M**, i.e., that the model is available to the relevant actors (and not others) for interpretation (**I** and **T**).
- Empirical quality deals with comprehensibility of the model **M**.
- Syntactic quality is the correspondence between the model **M** and the language extension **L**. Is the language used correctly in the model?
- Semantic quality is the correspondence between the model **M** and the domain **D**.
- Perceived semantic quality is the similar correspondence between the social actor interpretation **I** of a model **M** and his or her current knowledge **K** of domain **D**.
- Pragmatic quality is the correspondence between the model **M** and the actor interpretation (**I** and **T**) of it. Thus, whereas empirical quality focuses on if the model is understandable according to some objective measure that has been discovered empirically in, e.g., cognitive science, we at this level look on to what extent the model has actually been understood.
- The goal defined for social quality is agreement among social actor interpretations of the models.
- The deontic quality of the model relates to that all statements in the model **M** contribute to fulfilling the goals of modelling **G**, and that all the goals of modelling **G** are addressed through the model **M**.

When we structure different quality aspects according to these levels, one will find that there might be conflicts between the levels (e.g., what is good for semantic quality might be bad for pragmatic quality), thus, it is important to make a trade-off between achieving the different quality levels for achieving the main goals of modelling. In this paper, we focus on the relationship between syntactic quality and pragmatic quality.

### 3 Case Environment - Statoil Quality Management System

The enterprise model is realized through the Statoil management system. The Statoil Book [15], which is the foundation the management system is built upon, describes the management system as "the set of principles, policies, processes and requirements which support Statoil in fulfilling the tasks required to achieve their goals". It defines how work is done within the company, and all employees are required to act according to the relevant governing documentation.

The Management System consists of three main parts:

- Process models in ARIS<sup>1</sup>, the modelling solution from which all governing documentation is accessed by the end users.
- Docmap, used for handling and publishing textual governing documentation with more detailed requirements not represented in the process models directly.
- Disp, a tool which supports the process of handling applications for deviation permits in cases where compliance with a requirement is difficult or impossible to achieve.

The three main objectives of the Management System are:

- Contributing to safe, reliable and efficient operations and enabling compliance with external and internal requirements.
- Helping the company incorporating their values, people and leadership principles into everything they do.
- Supporting business performance through high-quality decision-making, fast and precise execution and continuous learning.

Governing documentation (GD) describes what is to be achieved, how to execute activities, and ensures standardization. Each process area has governing documentation in the form of documents and/or process models, accessible from the ARIS start page. A three-level process model structure is developed. The bottom level, the so-called workflow diagrams, contains BPMN models [16] on the descriptive level. The quality system contains around 2000 BPMN models at this level, qualifying the case to be an example of BPMN in the large [17].

The management system function is responsible for creating and improving the management system based on business needs and ensuring that the governing documentation is understood and used, as well as monitoring compliance with work requirements. The work of the function follows a five-step cycle; Assess and plan, design, implement, use, and monitor and control. This is done in close collaboration with line management and owners of the governing documentation.

The enterprise process model is created according to a set of rules [18] for structuring and use of the process modelling notation, and can be used for a variety of purposes, such as compliance management, competence management, portfolio management, decision making and performance analysis. There are three levels of abstraction in the enterprise model: The contextual level, the conceptual level and the logical level, including the following interrelated diagrams:

- The top-level diagram is a mandatory navigational diagram visualizing core value chain processes, management processes, and support processes, capturing what they in Statoil term the contextual level. This is similar to what others have termed a process map [19], depicting the core, support and management processes at the highest level.
- The navigation diagram(s) are optional diagrams to support more tailored access to the processes than the top-level diagram.
- Model diagram: Is a mandatory diagram that visualizes the model of one process area in the organization.

---

<sup>1</sup> <http://www.softwareag.com>

- Process navigation diagram is an optional model for navigational support on the conceptual level.
- Workflow diagram - Contains BPMN models [16] on the descriptive level.

Using the Splunk tool<sup>2</sup> one can capture how often a certain page or model is accessed and how users navigate through the enterprise model. Table 1 lists the ten most frequently used workflow models (query as of 20/10 2014). 12 out of the 20 most used models represent safety critical processes, i.e., they are either classified as Safe work (a sub-category of Operation and Maintenance) or belong to the Safety process area. The high number of distinct users gives an indication of the high level of use of the models, which is partly due to that the use is mandatory in many operational areas.

**Table 1.** The 10 most used process models

Workflow model	Click-count	Distinct users	Hits per user
Prepare isolation plan	34580	4054	8.5
Apply for and evaluate work permit	24471	4145	5.9
Initiate modification	22975	2342	9.8
Perform work at high	20041	3953	5.1
Commissioning and handover of systems	18285	2308	7.9
Checklist for safe work	16349	3572	4.6
Safety incident	15649	1628	9.6
Prepare for activity that weakens safety system	15340	3438	4.5
Execute mechanical completion	13560	1993	4.5
Perform bolt tightening	13013	2076	6.3

When designing diagrams in the enterprise model, requirements in TR0002 - Enterprise structure and standard notation [18] shall be met. We found, many places [20] that Statoil has developed their own variant of BPMN, including only some part of the overall language, and adding some special notation. In [21] we provided a mapping of the Statoil modeling requirements from TR0002 to SEQUAL. In the next section we will, in particular, look upon the current syntactic quality issues of models relative to these guidelines (including lacking conformance to naming and labelling guidelines which in [21] was listed under empirical quality).

#### 4 Influence of Syntactic on Pragmatic Quality

During the end of 2013 and the beginning of 2014, a large user survey [22] was conducted in Statoil in order to better understand users' experiences and opinions related to the management system and governing documentation including the process models. 4828 employees took part in the survey, which equals to about half of those invited. Many challenges were identified from the survey, related to the management system itself, learning processes and work practice, all of which contribute in some way to the management system goals of safety, reliability and

<sup>2</sup> <http://www.splunk.com/view/splunk/SP-CAAAG57>

efficiency. The survey is seen as very useful, due to the large amount of quantitative data as well as the amount of descriptive feedback given by the participants.

The survey results are positioned relative to the SEQUAL categories in [23] and, among other things uncovered challenges regarding understanding of some of the models (pragmatic quality). Although a large proportion of users feel that the governing documentation is easy to understand, others report issues of vagueness and ambiguity. For instance, many of the survey respondents do not understand all the abbreviations used in the text and models, although the official requirements [18] mention particularly that abbreviations should not to be used.

One of the main purposes of the document TR0002 [18] is to ensure a high syntactic quality in the models made. The document provides an overview of the allowed symbols and naming conventions, both symbol specific and more generally.

According to [2] syntax errors are of two kinds:

- Syntactic invalidity, in which words or graphemes not part of the language are used.
- Syntactic incompleteness, in which the model lacks constructs or information to obey the language's grammar.

The degree of syntactic quality can be measured as one minus the rate of erroneous statements, i.e.,

$$\text{Syntactic quality} = 1 - (\# M_E \setminus L + M_{\text{missing}}) / \# M_E \quad (1)$$

where  $M_E$  designate the explicit model statements and  $M_{\text{missing}}$  is the number of statements that would be necessary to make the model syntactically complete.

As missing statements (syntactic incompleteness) are very rare in the case studied, so the formula is simplified to

$$\text{Syntactic quality} = 1 - (\# M_E \setminus L) / \# M_E \quad (2)$$

In the following evaluation, the degree of syntactical correctness was first measured on seven workflow models. In the user survey [22], respondents were asked to give examples of processes that were interpreted differently within their department/unit. This list of processes was used as a basis when selecting models for evaluation. Due to a high number of models listed, not all could be evaluated. The following criteria were applied when selecting models:

1. The process is directly mentioned by respondents in the user survey [22] as a cause for misunderstandings and different interpretations, and implicitly mentioned at least twice.
2. The total number of nodes and edges in the model is larger than 20.
3. The model is one of the 100 most used workflow models.

Implicit mentions could for instance be references to a process chain that the workflow is part of, or the process or parts of the process being described by a sentence without naming the process or its identifier.

In Table 2 the selected seven models are listed including the syntactic errors. The rules are annotated according to the symbol or aspect they are related to, i.e.:

- N: Naming conventions
- T: Task
- OT: Optional Task
- G: Gateways
- SP: Collapsed Sub Process
- CA: Collaboration Activity
- SF: Sequence Flow
- W: Wrongly used concept

**Table 2.** Syntactic quality measurements

Model	Size	Breaches	SYN	AVG
Apply for and evaluate work permit	21	7xN2, 2xG2, 2xG3, 2xN2, CA3	0.55	0.87
Prepare isolation plan	23	CA3, G2, N2	0.89	0.89
Project control	24	12xN4, N2,E1, CA2, CA4, G2, G3,SF1	0.48	0.82
Execute mechanical completion	30	2xN4, 4xN7, 4xE1, 3xW, 4x N2, 2xSF1, G2, G3	0.37	0.80
Set, verify and approve isolation	30	2xN2, 2x SF1, CA4	0.87	0.80
Safety incident (SF103)	39	E1, 7xN2, 3xN2, 3xG2,2xG3, SP2, 4xW	0.58	0.78
Commissioning and handover of systems	46	2xE1, SP1, SP2, 2xSF1, 16x N4, 2xN2, 5xG2, 5xG3, 3xT1	0.39	0.78

The size of the model is equal to the total number of nodes (symbols) and edges (arrows). After measuring the syntactic quality (SYN) of these seven selected workflow models, they were compared to the average quality of other models of a similar size. The criteria used when choosing models for comparison were the same as the criteria listed above, except for criterion 1 which was inverted - only models without direct mentions were found appropriate. For each of the "troublesome" models, the three models closest in size from the top 100 list that also fit the set criteria were evaluated. The results are summarized in Table 2, indicating errors of the types found in the bullet list below. We can recognize some error types also identified in more general frameworks such as 7PMG [24], but as we see, many more error types than the 7 in 7PMG are described:

- N1: Names on symbols and expressions shall be formulated in singular form.
- N2: Avoid names with more than four words if possible.
- N3: A name shall not be a detailed description.
- N4: The first letter of a symbol name shall be in upper case. All other letters should be lower case.
- N5: Proper names shall start with upper case letters.
- N6: The Statoil official name of a concept shall be used when alternatives exist.
- N7: Abbreviations should be avoided.
- T1: The title of a task shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset).
- OT1: The title of an optional task shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset).
- OT2: The use of an optional task is only allowed within a collaboration activity.
- OT3: It is not allowed to connect sequence flows to the optional task symbol.
- SP1: The title of a collapsed sub-process shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset).
- SP2: The collapsed sub-process symbol is drawn using a standard activity shape with a "+" attached.
- CA1: The tasks grouped by a collaboration activity symbol shall not be sequenced in time or contain dependencies.
- CA2: The title of a collaboration activity shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset).
- CA3: The name of a collaboration activity shall be unique and you shall not name the collaboration activity with names that have been used in the tasks that have been framed by the collaboration activity symbol.

- CA4: Each of the tasks framed by the collaboration activity symbol must have a unique title, clarifying different type of activities performed by different roles.
- E1: You shall define the title of a start or end event as a noun (reflecting the asset) followed by a verb past participle (reflecting the activity performed to add value to the asset).
- G1: You shall not name parallel gateways.
- G2: The title of a diverging exclusive gateway shall consist of the term control (can be replaced with check, verify, evaluate or clarify) followed by a noun (reflecting the object submitted to control).
- G3: The exclusive flow shall be described through an adjective or a phrase describing the alternative flows. You shall not use yes or no when designing exclusive gateways.
- SF1: A sequence flow shall have only one source and one target.
- SF2: You should not use more than one sequence flow from an activity.
- W: Using the wrong symbol (or similar errors).

#### 4.1 Experiment Design and Results

In the experiment, two workflow models were selected, and changes were made to these models to increase their syntactic quality according to the guidelines developed as described above. Participants were asked to answer a range of questions related to the models in order to measure their understanding and, thus, the pragmatic quality of the models.

The original intention was to use only Statoil employees from different departments and locations as participants, but since it proved to be difficult to find enough volunteers in Statoil, a student experiment was carried out in parallel. In total, 18 students and 9 Statoil employees participated in the study. In order to avoid participants answering based on personal knowledge rather than by consulting the models, the participants from Statoil did not have first-hand experience with the modelled processes. The models selected for the experiment had a syntactic quality below average, and were found to be easily improvable by correcting mistakes, according to the rules found in TR0002 [18] listed above. Improvements were made to several models before selecting the two workflow models chosen here:

- SF103 - Safety incident
- OM05.07.01.03 - Reset isolation and pressurize

Key numbers for these workflow models are given in Table 3. SF103 was also part of the syntactic quality evaluation reported in Table 2 because it was highlighted in the user survey as a model subject to misinterpretations. OM05.07.01.03 was not directly mentioned, but has as many as 9 implicit mentions, mostly due to the "parent" process being listed.

**Table 3.** Characteristics of workflow models used in the experiment

Model	Hits	Size	Syntactic quality
SF103	16752	39	0.56
OM05.07.01.03	6662	29	0.72

Syntactic quality was here measured on the Norwegian versions of the models, as the experiment was to be conducted in Norwegian. This was decided in order to avoid language-related misunderstandings, as all of the respondents were native Norwegian speakers. With the conventions and the metric used, there might be slight differences in measured quality between versions in different languages, as some of the rules are related to naming. The Norwegian version of SF103 had a low original syntactic quality of 0,56, while OM05.07.01.03 had a moderate syntactic quality of 0,72 compared to average of models of this size (see Table 2).

When making the new versions, the models were adjusted to make the syntactic quality as close to 1 as possible. Quite major changes were made to SF103, as many of the errors were large, e.g., the wrong symbol was used in several cases. With OM05.07.01.03, the changes made were mostly corrections in the naming of symbols and splitting of arrows (The identified errors were 4xN2, 2xG2, G3, 2xE1, i.e., mostly to do with naming and gateways).

The participants were each given two models to interpret - one original and one modified. The participants were split into four groups, and each group was given a different combination of models, following a Latin square design, outlined in Table 4. As shown in the table, two groups were given the new SF103 and the old OM05.07.01.03. The other two were given the new OM05.07.01.03 and the old SF103. The order of presentation was also reversed for half of the groups, to avoid the order to affecting the results. In addition, they were given an overview of the language notation.

**Table 4.** Latin Square experimental design

Group	Model 1	Model 2
Group1	SF103 (new)	OM05.07.01.03(old)
Group2	OM05.07.01.03(old)	SF103 (new)
Group3	SF103 (old)	OM05.07.01.03(new)
Group4	OM05.07.01.03(new)	SF103 (old)

The participants were each given 15 questions connected to SF103, and 10 questions connected to OM05.07.01.03. When summarizing the results, each wrongly answered question was given -1 points, unanswered questions were given 0 and correct answers were given a score of 1. The total number of available points for each model is the result of (number of participants x number of questions), e.g.,  $9 \times 15 = 135$  for questions to the old SF103 in the student experiment. Results from the experiment with Statoil employees should be given some emphasis in the analysis, although any differences in the performance of students and employees would also be interesting given the different experience with the domain and the specific version of BPMN.

### **SF103 - Safety Incident**

The overall results for SF103 are summarized in Table 5. As shown, the modified version of SF103 scored much higher than the original version both in the Statoil experiment and the student experiment. Some specific questions are worth taking a closer look at, as they give insight into certain problem areas and normal misunderstandings. Question 2 stands out, as all of the Statoil participants answered wrongly when looking at the old version of the model, and half of those looking at the new:

#### *2. True or false: The process always starts with a safety incident occurring*

Looking at the student respondents the change is even bigger: as many as 7 out of 8 answered the question wrongly for the original version, and only two made the same mistake when the new version of the model was given. The question is related to events. In the process model, there are two possible triggers to the process. In the original version, many event-related symbols are used incorrectly, e.g., there are two cases of "end event" symbols with sequence flows pointing out from them, and event symbols are used instead of task symbols even though the process does not start or end at these points. It is therefore not surprising that the respondents have trouble distinguishing the actual process triggers.

**Table 5.** SF103 results

Experiment	Old version	New version
Statoil	33/60 p (55%)	52/60 p (87%)
Students	93/135 p (69%)	122/135 p (90%)

The next critical question is number 6 (the question had three alternatives):

*6. What is special about the activity "categorize, classify and decide causes"?*

2 of 4 respondents answered incorrectly when looking at the old model, while everyone managed to answer correctly when looking at the new. This might be due to that the sub-process symbol used in the original model does not correspond exactly to the one defined in the standard notation overview, as it lacks the "+" a collapsed sub-process is supposed to have attached to it, according to the text (this is, however not depicted in the legend overview). However, this mismatch is not reflected in the students' responses - all of them answered the question correctly.

Question 9 also got two wrong answers with the original version, and none with the new:

*9. The process ends when an accident investigation is carried out*

Here, some of the students are also confused: the old version lead to three wrong answers and one unsure (unanswered), whereas the new lead to only correct answers. This question is also event-related, so the reasoning is the same as for question 2.

### **OM05 .07.01.03 - Reset Isolation and Pressurize**

The results for OM05.07.01.03 are shown in Table 6. The syntactic quality of this model was higher than of the model discussed above. In this case, the new version actually got a lower score, but the difference is not very big. Among Statoil employee, the difference is also evenly spread among the questions, none of the questions differ by more than two points (corresponding to one mistake less or more) between the two model versions.

The question with the lowest score for both versions was question 3:

*3. Yes or no: Should the area technician always contribute to approving the execution?*

A similar result can be seen among both groups. The question is connected to an optional task. Even though it is specified in the legend that a task symbol with a stippled line is optional, many are not able to distinguish this from a regular task.

Question 6 also gave some interesting results:

*6. What should be investigated when arriving at the symbol "Safety valve?" (old version) /"Check safety valve" (new version)?*

All of the Statoil employees answered the question correctly for both versions, except for one who was "unsure" (old version), whereas in the student experiment, four of the respondents looking at the old version skipped the question and one gave the wrong answer. Everyone answered correctly when looking at the new model. The question pertains to a gateway symbol which in the old version is labelled merely "Safety valve?" ("Sikkerhetsventil?" in Norwegian) where exits are annotated with 'yes' and 'no'. The text is not very descriptive, so without any domain knowledge, it could be very difficult getting the meaning of this gateway symbol. This might explain why the Statoil employees got this one right while so many students were unsure.

Even though the Statoil respondents did not have first-hand knowledge about the process; they have probably picked up some knowledge about the general domain over the years of working in the oil industry.

**Table 6.** OM05.07.01.03 results

Experiment	Old version	New version
Statoil	31/40 p (78%)	27/40 p (68%)
Students	64/90 p (70%)	59/90 p (66%)

## 5 Discussion, Conclusion and Further Work

The quality system of Statoil is developed supporting, in particular, compliance to requirements to reduce risk, an area where large improvements have been observed over the last decade. Still, one finds challenges with, among other things, the comprehension of some of the models as described above. While the requirements given in TR0002 are quite detailed and structured, providing guidelines on most levels of SEQUAL, they are not always followed in practice. Measurements on syntactic quality show that syntax errors are quite common in the workflow models.

The user survey [22], interviews and conversations provided valuable insights into how users experience the management system. Some measures can be taken to achieve higher quality. Some users feel that governing documentation is hard to understand. Increased understanding is a necessity if 100% compliance is the goal. Measures that can contribute to this include applying the language guidelines and naming conventions more strictly and tailoring the complexity of models according to the needs of its target audience. Also providing relation to organizational ontologies (e.g., organizational structure and domain ontologies) might be helpful [25]. The experiment gave mixed results at first sight. Whereas improvement in labelling and syntax appeared to improve the comprehension in one of the cases, the other case, which had less severe syntactic errors initially, showed no difference, illustrating that good syntactic quality can be useful for comprehension, but that in some cases other aspects are more important if the syntactic quality is sufficiently good. It might also give insights to what kind of labelling issues and syntactic errors are important. Completely wrong use of symbols is here clearly more problematic than not following recommendations like N2: “Avoid terms with more than four words if possible”. Incidentally, it is interesting to note that in an earlier version of the guidelines [6], this guideline was even less strictly stated: ‘A name with more than four words should be carefully evaluated’.

The main threat to validity in the model quality experiment is that the number of participants was low. Hence, the data cannot be used for proving or disproving a hypothesis with statistical significance, and the trends discovered may be coincidental. Additionally, students are not part of the target group of the enterprise model, and the findings would have greater validity if all participants were Statoil employees, preferably employees who use the enterprise process model frequently in their everyday work.

Although it is not possible to make generalizations about the effect of the syntactic changes on understanding, it was still useful looking at some specific questions and seeing that many, even Statoil employees, do interpret the model wrongly - be they syntactically correct or not. From some of the answers, it was clear that not everyone know and understand the standard notation, so increasing awareness about models and a common modelling standard is important for efficient use of the models.

There are several possibilities for further work related to the Statoil enterprise process model. Based on the internal evaluation, new modeling standards and tool support is being developed. When the new functionality developed has been implemented in full-scale, the actual effect of

these changes on model quality in practice can be analyzed. A new user survey, similar to the one carried out here will be distributed by Statoil when these changes have been put into effect. Studying the results based on the new standards and tools and comparing them to the old ones may give important insight into the real value of such changes. In particular, following the implementation of the new TR0002 document in practice, and how it impacts model quality and use is an interesting possibility for future research.

Another possibility is to carry out a more quantitative study, in which an experiment similar to the model quality experiment reported here is carried out in a larger scale with enough respondents to get statistically significant results. Lastly, using SEQUAL to evaluate enterprise process modelling in other large organizations can, together with this work, provide the basis for making generalizations useful for the practice of enterprise process modelling.

## References

- [1] H. Wesenberg, "Enterprise Modeling in an Agile World," PoEM 2011, in Proc. the 4th conference on Practice of Enterprise Modeling, Oslo, Norway, Nov. 2-3, 2011. Available: [http://dx.doi.org/10.1007/978-3-642-24849-8\\_10](http://dx.doi.org/10.1007/978-3-642-24849-8_10)
- [2] J. Krogstie, "Model-based development and evolution of information systems: A quality approach," Springer, London, 2012. Available: <http://dx.doi.org/10.1007/978-1-4471-2936-3>
- [3] D. L. Moody, "Theoretical and practical issues in evaluating the quality of conceptual models: Current state and future directions," *Data and Knowledge Engineering* 55, pp. 243-276, 2005. Available: <http://dx.doi.org/10.1016/j.datak.2004.12.005>
- [4] H. J. Nelson, G. Poels, M. Genero and M. Piattini, "A conceptual modeling quality framework," in *Software Quality Journal* 20, pp. 201-228, 2012. Available: <http://dx.doi.org/10.1007/s11219-011-9136-9>
- [5] R. Price and G. Shanks, "A semiotic information quality framework: Development and comparative analysis," *Journal of Information Technology*, 20 (2), pp. 88-102, 2005. Available: <http://dx.doi.org/10.1057/palgrave.jit.2000038>
- [6] Statoil, "TR0002 Enterprise Structure and Standard Notation. version 1," 2009.
- [7] O. I. Lindland, G. Sindre, and A. Sølvsberg, "Understanding Quality in Conceptual Modelling," *IEEE Software*, vol. 11, no. 2, pp. 42-49, 1994. Available: <http://dx.doi.org/10.1109/52.268955>
- [8] M. Heggset, J. Krogstie, and H. Wesenberg, "The Influence of Syntactic Quality of Enterprise Process Models on Model Comprehension," *CAiSE Forum, Stockholm CEUR*.
- [9] J. Krogstie and S. de Flon Arnesen, "Assessing enterprise modeling languages using a generic quality framework," in *Information modeling methods and methodologies*, pp. 1537-9299, 63(79), 2005.
- [10] J. Krogstie, V. Dalberg, and S. M. Jensen, "Process modeling value framework," in Y. Manolopoulos, J. Filipe, P. Constantopoulos, J. Cordeiro, Eds., *Selected papers from 8th International Conference, ICEIS 2006*, pp. 309-321, Springer, Paphos, 2008. Available: [http://dx.doi.org/10.1007/978-3-540-77581-2\\_21](http://dx.doi.org/10.1007/978-3-540-77581-2_21)
- [11] D. L. Moody, G. Sindre, T. Brasethvik, and A. Sølvsberg, "Evaluating the quality of process models: empirical testing of a quality framework," in *Proc. 21st International Conference on Conceptual Modeling Tampere (ER'2002)*, Finland, Oct. 7-11, 2002. Available: [http://dx.doi.org/10.1007/3-540-45816-6\\_36](http://dx.doi.org/10.1007/3-540-45816-6_36)
- [12] L. C. Christensen, B. W. Johansen, N. Midjo, J. Onarheim, T. Syvertsen, and T. Totland, "Enterprise modeling-practices and perspectives," in *Computers in Engineering*, pp. 1071-1084, 1995.
- [13] A. G. Nysetvold and J. Krogstie, "Assessing Business Process Modeling Languages Using a Generic Quality Framework," in K. Siau, Ed., *Advanced Topics in Database Research*, vol. 5, pp. 79-93. Idea Group, 2006.
- [14] H. Stachowiak, "Allgemeine Modelltheorie," Springer, Wien, 1973.
- [15] Statoil, "Statoilboken." [Online]. Available: <http://www.statoil.com/no/About/TheStatoilBook/Downloads/Statoil-Boken.pdf>, 2014.
- [16] B. Silver, "BPMN Method and Style," Cody-Cassidy Press, 2012.

- [17] C. Houy, P. Fettke, P. Loos, W. M. P. van der Aalst, and J. Krogstie, "Business Process Management in the Large," *Business & Information Systems Engineering* (6), 2011. Available: <http://dx.doi.org/10.1007/s12599-011-0181-5>
- [18] Statoil, "TR0002 Enterprise Structure and Standard Notation. version 3", 2013.
- [19] M. Malinova, H. Leopold, and J. Mendling, "A Meta-Model for Process Map Design," in *CAISE Forum 2014*, June 16-20, Thessaloniki, Greece, 2014.
- [20] G. Aagesen and J. Krogstie, "Analysis and design of business processes using BPMN," in Jan vom Brocke & Michael Rosemann, Eds., *Handbook on Business Process Management*, Springer, 2010. Available: [http://dx.doi.org/10.1007/978-3-642-00416-2\\_10](http://dx.doi.org/10.1007/978-3-642-00416-2_10)
- [21] M. Heggset, J. Krogstie, and H. Wesenberg, "Ensuring quality of large scale industrial process collections: Experiences from a case study," in *The Practice of Enterprise Modeling*, pp. 11-25. Springer, 2014. Available: [http://dx.doi.org/10.1007/978-3-662-45501-2\\_2](http://dx.doi.org/10.1007/978-3-662-45501-2_2)
- [22] Statoil, *Management system user survey u13*, 2014.
- [23] M. Heggset, J. Krogstie, and H. Wesenberg, "Understanding Model Quality Concerns when Using Process Models in an Industrial Company," in *Proc. EMMSAD 2015*, Springer, 2015. Available: [http://dx.doi.org/10.1007/978-3-319-19237-6\\_25](http://dx.doi.org/10.1007/978-3-319-19237-6_25)
- [24] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven Process Modeling Guidelines (7PMG)," *Information and Software Technology (IST)*, 52(2), pp. 127–136, 2010. Available: <http://dx.doi.org/10.1016/j.infsof.2009.08.004>
- [25] Y. Lin, D. Strasunskas, S. Hakkarainen, J. Krogstie, and A. Sølvyberg, "Semantic annotation framework to manage semantic heterogeneity of process models," Paper presented at the *CAiSE'2006*, Luxembourg, 2006. Available: [http://dx.doi.org/10.1007/11767138\\_29](http://dx.doi.org/10.1007/11767138_29)