Vara González, José Luis de la; Marín, Beatriz; Giachetti, Giovanni; Ayora, Clara (2016). "Do Models Improve the Understanding of Safety Compliance Needs?: Insights from a Pilot Experiment". *ESEM 2016. Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement.* ACM, article 32.
http://dx.doi.org/10.1145/2961111.296262

Do Models Improve the Understanding of Safety Compliance Needs? Insights from a Pilot Experiment

Jose Luis de la Vara
Computer Science Dept.
Carlos III Univ. of Madrid
Leganes, Spain
jvara@inf.uc3m.es

Beatriz Marín
Facultad de Ingeniería
Universidad Diego Portales
Santiago, Chile
beatriz.marin@mail.udp.cl

Giovanni Giachetti
Facultad de Ingeniería
Universidad Andres Bello
Santiago, Chile
giovanni.giachetti@unab.cl

Clara Ayora
Independent researcher
Valencia, Spain
claraayora@gmail.com

ABSTRACT

Context. Many critical systems must meet safety compliance needs from safety standards. These standards are usually large textual documents whose compliance needs can be hard to understand. As a solution, the use of models has been proposed. *Goal.* We aim to provide evidence of the extent to which models improve the understanding of safety compliance needs. *Method.* We designed an experiment and ran a pilot to study the effectiveness, efficiency, and perceived benefits of understanding these needs, with the text of standards and with models in the form of UML object diagrams. *Results.* The overall results from 15 Bachelor students show that the effectiveness of understanding safety compliance needs increases very little with models (2%), and the efficiency even decreases (24%). Nonetheless, the results improve when the potential complexity in navigating the models is taken into account (15% effectiveness increase). The students find benefits in using the models but most consider that the models are hard to understand. *Conclusions.* The extent to which models improve the understanding of safety compliance needs seems to be lower than what the research community expects. New studies are necessary to confirm our initial insights.

CCS Concepts

• Software and its engineering → Software safety • Hardware → Safety critical systems • General and reference → Computing standards, RFCs and guidelines • Software and its engineering → Model-driven software engineering.

Keywords

Safety-critical system; safety standard; safety compliance needs; model; understanding; pilot experiment.

1. INTRODUCTION

Critical software-intensive systems in domains such as aerospace, railway, and automotive are subject to some form of safety assessment by a third party (e.g. a certification authority) as a way of assuring that the systems do not pose undue risks to people, property, or the environment. A common type of assessment is compliance to safety standards, usually referred to as safety certification [15]. Examples of safety standards used in industry [17] include IEC 61508 in a wide range of industries, DO-178C in avionics, EN 50128 in railway, and ISO 26262 in automotive.

ESEM'16, September 8–9, 2016, Ciudad Real, Spain.

DOI: http://dx.doi.org/10.1145/12345.67890

Safety standards are typically large textual documents. They can consist of hundreds of pages and define thousands of criteria for compliance. These criteria can be referred to as safety compliance needs [7], which include requirements to fulfil, data to manage, activities to execute, relationships between these elements, and information about when and how the elements should be addressed for a given critical system.

Safety compliance needs can be difficult to understand due to the size of the standards and to ambiguity and inconsistencies in their text [7], [17]. This can lead to certification risks, as a system supplier might miss or misinterpret some compliance needs and thus not able to develop a compliant system. As a solution, several authors have proposed the use of models and argue that model-based representations of safety compliance needs can help practitioners understand these needs, e.g. [16], [20]. However, there exists little evidence of the extent to which the use of models improves the understanding of safety compliance needs. The available studies [7], [19] are based on experts' perceptions, not on the actual usage of the models. There is also a general lack of experiments on approaches for safety certification [15].

We are currently working towards filling these gaps. In this paper, we present the results of a pilot experiment with Bachelor students to study the effectiveness, efficiency, and perceived benefits of understanding safety compliance needs with models. The students answered questions about safety compliance needs in the text and in models (UML object diagrams) of DO-178C and of EN 50128, and indicated their opinion about the use of models.

The overall effectiveness (F-measure) of understanding safety compliance needs with models and with the text of the standards was similar, and the efficiency (F-measure/time) was higher with the text. The students found benefits in using the models, e.g. to determine how to comply with safety standards, but most considered that the models were hard to understand.

Although the main purpose of the pilot was to validate the design of the experiment, its results provide initial evidence of the effectiveness, efficiency, and perceived benefits of understanding safety compliance needs with the text of standards and with models. To the best of our knowledge, this paper is the first publication that presents such evidence from the results of an experiment. Nonetheless, further, stronger empirical evidence is necessary to draw definite conclusions. The pilot will also help us to adjust the experiment design for future executions, and to derive hypotheses for a follow-up experiment.

The rest of the paper is organised as follows. Section 2 introduces the background of the paper. Section 3 presents the experiment process, and Section 4 reports on the results. Finally, Section 5 summarises our main conclusions and future work.

2. BACKGROUND

The background of the paper is divided into model-based specification of safety compliance needs and related work.

2.1 Model-based specification of safety compliance needs

Several authors advocate for the use of models to understand safety standards, determine how to comply with them, properly follow them, and demonstrate compliance. Model-based approaches for the specification of safety compliance needs can be found for specific safety standards (e.g. IEC 61508 [20]), parts of them (e.g. testing with DO-178B [23]), and compliance needs (e.g. artefact information [16]). Modelling standards have also been published [18] and models are used in industry for safety certification purposes [6], [17].

For the pilot experiment, we use a holistic generic metamodel for the specification of safety compliance needs [7]. An excerpt is shown in Figure 1. This metamodel supports the specification of the different types of safety compliance needs, i.e. information about requirements, artefacts, and processes, and about their applicability. The metamodel can be used for a wide range of standards from different domains, and has been validated with practitioners and with data from several real projects.

With the metamodel, safety compliance needs are specified by means of: (1) reference requirements, conditions to fulfil (e.g. software modularity); (2) reference activities, units of behaviour to execute (e.g. software development processes); (3) reference roles, types of agents to be involved (e.g. designer); (4) reference artefacts, units of data to manage (e.g. safety plan); (5) reference techniques, specific ways to execute a reference activity or create a reference artefact (e.g. formal methods); (6) reference artefact relationships, relationships to record between two reference artefacts (e.g. satisfies; design satisfies requirements); and, (7) reference artefact attributes, characteristic of a reference artefact (e.g. test case result). All these classes specialise reference element. Relationships between them and applicability information can also be specified, e.g. whether a given reference technique should be used according to a system's criticality. Reference activity, artefact, role, and technique also specialise constrained reference assurable element.

Further information of the metamodel can be found in [7].

2.2 Related Work

To the best of our knowledge, the available evidence of the extent to which models might improve the understanding of safety compliance needs is based on experts' opinion. Therefore, this paper is the first study that analyses the potential improvement in a different way, with the results of a pilot experiment.

Twelve practitioners provided feedback on an IEC 61508 model [20]. Most of them agreed or strongly agreed that the model was easy to understand, and very probably or definitely would use the model to help in understanding the standard. Four practitioners that used the holistic generic metamodel at a training session provided feedback on the use of the models [7]. Overall, these practitioners found benefits in understanding safety compliance needs with models, especially the concepts of the standards and the relationships between them. Although the results from these publications are valuable towards showing whether models can improve the understanding of safety compliance needs, further studies are necessary. These studies should be based on actual model usage and compare it with the use of the text of standards.

There are indeed very few experiments on approaches for safety certification [15], including on model-based approaches. Among the experiments related to safety certification, Briand et al. [3] analysed a SysML-based traceability approach for safety inspections and showed that it increases decisions' correctness. Abdulkhaleq and Wagner [1], Jung et al. [12], Mouaffo et al. [14] compared safety analysis techniques, and Cyra and Gorski [5] validated an approach for argument assessment.

Many publications report experiments on model understanding. In the software engineering research community, experiments have been conducted for models such as ER and UML class diagrams [8], UML sequence diagrams [2], UML state charts [4], SysML requirements diagrams [21], and use cases and Tropos [10]. Textual and graphical representations have been compared, e.g. for requirements [22] and software architecture design [11]. Experiments on business process model understanding have also been conducted [13], including the comparison of the understanding of business processes with textual descriptions and with business process models, e.g. [19]. The results of these experiments show that the use of models can facilitate understanding, thus we conjecture that models might improve the understanding of safety compliance needs.

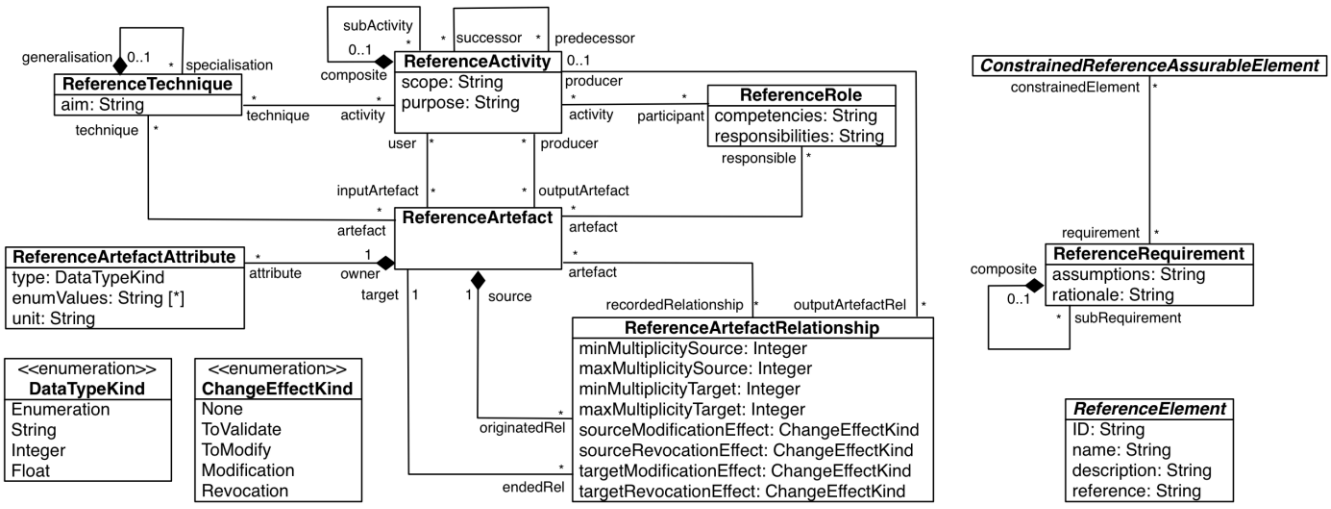

Figure 1. Excerpt of the holistic generic metamodel for model-based specification of safety compliance needs

3. EXPERIMENT PROCESS

We used the guidelines by Wohlin et al. [24] to design the experiment. The goal is to analyse the use of models for specifying safety compliance needs for the purpose of evaluation with respect to effectiveness, efficiency, and perceived benefits of understanding safety compliance needs from the point of view of the researcher in the context of Bachelor students in Computer Science. We formulated three research questions (RQs):

- RQ1. Does the use of models increase the effectiveness of understanding safety compliance needs?
- RQ2. Does the use of models increase the efficiency of understanding safety compliance needs?
- RQ3. Do users find benefits in the use of models for understanding safety compliance needs?

The following subsections summarise the planning and operation of the experiment. We also discuss the main validity aspects.

3.1 Experiment Planning

The context of the experiment is a 3rd-year course of a Bachelor's Degree in Computer Science and Engineering at Carlos III University of Madrid, Spain. The course is titled "Software development projects management" and it is obligatory. The subjects are students of the group of the course with teaching in English. The students of this group have to plan the development and validation of a software application and to design it following the ESA PSS-05-0 software engineering standard and its associated guides [9]. In the experiment, the students have to answer questions about safety compliance needs from excerpts of the text of safety standards and from models of the excerpts. They have to identify the needs in each representation. The students also indicate their opinion about the models.

The independent variables are (1) the means of representing safety compliance needs (model or text of a safety standard) and (2) the safety standard considered (DO-178C requirements process or EN 50128 integration process, which are different to the standard used in the course). The main reason for selecting these processes is that the students have to deal with similar activities during the course. We use UML object diagrams to represent the instances of the holistic generic metamodel. This diagram has been used in related work to model safety compliance needs, e.g. [20]. We use a requirements model, a process model, an artefact model, and an applicability model to cover all the types of compliance needs.

The dependent variables of the experiment are the effectiveness, efficiency, and the perceived benefits in understanding safety compliance needs. The F-measure is used for effectiveness. It is based on the precision and recall in identifying safety compliance needs, and it has been used in related experiments, e.g. [2], [3]. We used the formulas for cases in which it is possible that a question has no answer [8]:

$$precision_s = \frac{\sum_i |answer_{s,i} \cap correct_i|}{\sum_i |answer_{s,i}|}$$

$$recall_s = \frac{\sum_i |answer_{s,i} \cap correct_i|}{\sum_i |correct_i|}$$

$$F_s = 2 \times \frac{precision_s \times recall_s}{precision_s + recall_s}$$

Efficiency is based on the F-measure and the time in minutes:

$$Effy_s = 100 \times \frac{F - measure_s}{minutes}$$

The perceived benefits are evaluated with a previously used questionnaire that contains statements about the use of models for specifying and understanding safety compliance needs [7].

For the pilot, all the students have to complete a questionnaire (object) about safety compliance needs in a text excerpt and in a model (two different tasks). Four groups are planned in a between-subject design: EN 50128 model and text (1), and vice versa (4), and DO-178C model and text (2), and vice versa (3). Using the same standard for the two tasks is not a threat because we mostly aimed to validate the experiment design with the pilot. We also do not analyse the results on effectiveness and efficiency from the second task to avoid the threat of learning. The follow-up experiment will be different. The students receive the material for the second task after finishing the first and, after the second task, they complete the questionnaire about the perceived benefits of the models. We do not analyse the perceived benefits of the text of the standards because there might exist a strong threat that the answers are biased due to the subjects' experience with the models. The students can also provide comments and suggestions.

Running the pilot was planned to require a maximum of two hours, one for training and one for performing the tasks. The first author, as main expert in safety certification, was the main responsible for creating the material. The rest of authors validated the material, and two performed the tasks to check the completion time. As a result, we made some adjustments in the material. The material for the tasks consists of an introductory page, a two-page excerpt of a standard or the models of the excerpts, and six questions (e.g. "What information should the High-Level Requirements conform to?" for DO-178C). The subjects have to identify ten safety compliance needs to correctly complete the questionnaire, the same in the text and in the model (e.g. Software Requirements Standards for the above question). For training, we prepared a presentation on safety assurance for critical systems and on the holistic generic metamodel, with metamodel usage examples for ESA PSS-05-0, to ensure homogeneous knowledge. The final material of the experiment is available online[1].

3.2 Experiment Operation

The pilot experiment was performed in May 2015, the last week of the second semester. Twenty-one students participated. The training duration was close to our plan. At the end of the training, and before performing the tasks, we told the students that the tasks targeted research purposes and that the students' performance would not affect their course grade. We nonetheless explicitly asked the students to do their best and execute the task in an exam-like way, e.g. not asking other students.

Four groups were then randomly created from the order in which the students were sit in the classroom. The subjects recorded the time when they started and finished each task. They performed the tasks faster than expected (15 minutes on average for the first task). We will take this into account for the follow-up experiment.

We checked the data after the experiment execution for validation and decided to discard the results from six students because we found indicators of careless response, e.g. several answers with information that was not in the models or the texts. Among the 15 subjects with valid results, no subject had previous knowledge about DO-178C or EN 50128, all the subjects had dealt with UML class or object diagrams in some course, and seven subjects had used UML class or object diagrams in some real project.

[1] https://sites.google.com/site/jldelavara/material/msac2015

3.3 Threats to Validity

Using more than one safety standard in the experiment avoids mono-operation bias and contributes to **construct validity**. Creating a model of safety compliance needs involves an interpretation of the corresponding standard that might not be shared, which is a threat. Nonetheless, the UML object diagrams used correspond to models of DO-178C requirements process and of EN 50128 integration process that we had created in the past and practitioners had validated. This contributes to model validity.

The specific representation of the models (e.g. their layout), their potential complexity, the expressiveness of UML object diagrams, subject fatigue, and having discarded the results from six subjects might have affected **internal validity**.

Conclusion validity is threatened by the selection of a given graphical notation for the models (UML object diagram). This might also be a confounding factor. Having only run a pilot experiment, mostly targeted at validating the experiment design, with a low number of subjects, and for which no statistical tests have been run, also negatively affects conclusion validity.

Regarding **external validity**, we refrain from widely generalising the results and conclusions. We have only run a pilot experiment, with excerpts of two standards and with a notation. Nonetheless, the results are still valuable for a general safety audience because DO-178C and standards similar to EN 50128 (IEC 61508 and derived standards) appear to be the most frequently used safety standards in industry [6], [17]. In relation to the use of students as subjects, it is common to consider that their performance can be close to novice practitioners', e.g. [2]. The available evidence further indicates that it cannot be claimed that experience greatly helps practitioners better understand safety compliance needs [6].

We discuss some further validity aspects in Section 4.

4. RESULTS AND INTERPRETATION

This section presents the results of the pilot experiment and how we interpret them. We use insights from related work to compare and discuss the results. The results related to each RQ are presented in different subsections.

Table 1 shows the results for the effectiveness and efficiency of understanding safety compliance needs, whereas Figure 2 shows the results for the perceived benefits in the use of models.

Table 1. Experiment results: effectiveness and efficiency of understanding safety compliance needs

			Individual results			Total results		
	Gr.	**Subj.**	**F**	**F'**	**Effy**	**F**	**F'**	**Effy**
Model	**1**	**1**	0.42	0.47	3.22	0.58	0.62	3.38
		2	1	1	6.01			
		3	0.7	0.67	3.12			
	2	**4**	0.42	0.45	2.22			
		5	0.43	0.48	2.89			
		6	0.47	0.5	3.11			
		7	0.57	0.63	3.29			
		8	0.67	0.73	3.17			
Text	**3**	**9**	0.55	0.5	4.74	0.57	0.54	4.47
		10	0.46	0.48	3.86			
		11	0.57	0.54	3.18			
		12	0.75	0.73	5.29			
	4	**13**	0.45	0.4	5.95			
		14	0.7	0.67	5			
		15	0.53	0.47	3.23			

4.1 Effectiveness of understanding (RQ1)

The effectiveness of the subjects in understanding safety compliance needs is shown in column *F* of Table 1. The effectiveness with models and the effectiveness with the text of the standards is very similar. The use of models resulted in an increase of average effectiveness by only 2%. This further appears to be a consequence of the good results by subject 2. These results suggest that the effectiveness gain from using models for the understanding safety compliance needs is lower than what the research community expects. Nonetheless, someone can also claim that full effectiveness has only been reached with a model.

We conducted a second analysis for effectiveness. When checking the data, we noticed that most of the subjects that had worked first with models, had answered the question about applicability information wrong. This information is presented in DO-178C and EN 50128 in tables, in a structured and compact way. The applicability models consisted of 14-17 nodes and of 16-22 links. The subjects had to navigate between the nodes to answer the question, which might have introduced additional difficulty and thus a threat to validity. Indeed, a subject indicated that this model had been the only one with which he had had problems.

We therefore considered that the performance of the subjects that had worked first with models might be worse because of issues with the applicability model, and decided to analyse effectiveness without taking the corresponding question into account. The results are shown in column *F'* of Table 1. In this case, there is a considerable effectiveness increase with the use of models (15% gain), which supports the claim regarding the improvement of the understanding of safety compliance needs with the use of models.

Someone could easily conjecture that using models can facilitate understanding based on previous experiments (see Section 2.2). Nonetheless, and in line with prior work that have compared understanding effectiveness with text and with models, the results of the pilot provide, in our opinion, inconclusive evidence. Although experiments with e.g. business process descriptions indicate that using models can increase understanding effectiveness [19], others with e.g. requirements [22] and architecture specifications [11] do not support this claim.

In summary, we cannot draw clear conclusions about RQ1. Some evidence suggests that the effectiveness of understanding safety compliance needs is similar with the text of standards and with models, and some evidence suggests that the effectiveness is higher with models. We will study this in more depth in the follow-up experiment. We will use, and try to reject, the hypothesis that the effectiveness is similar with the text of standards and with models. We will also address the issue with the question about the applicability model in this experiment.

4.2 Efficiency of understanding (RQ2)

The results of the pilot experiment regarding efficiency are shown in column *Effy* of Table 1. It appears that there is reasonable evidence to consider that using models to understand safety compliance needs is not more efficient than using the text of safety standards. First, the total results show that the efficiency with the models was quite lower (24%). Second, the efficiency gain with the text seems to be clear despite the similar results in effectiveness. Third, nine out of the ten subjects with the highest efficiency are in groups 3 and 4. The efficiency decrease when using models must be confirmed in the follow-up experiment, and with statistical tests. Related work has also reported on more efficient understanding with text than with models, e.g. [22].

Nonetheless, there are some aspects that might have affected the results on efficiency. First, the subjects had dealt with the text of ESA PSS-05-0 during the entire course, thus they might be more familiar with identifying safety compliance needs in text. Second, and as a follow-up reason, the subjects might have more experience in having to understand the text of safety standards than in having to understand models in general and UML object diagrams in particular. These aspects could be addressed in future experiment operations. For example, longer training sessions on the holistic generic metamodel could be held. In the scope of the course in which the pilot was executed, the subjects could be asked to create a model of ESA PSS-05-0 on their own before performing the experiment tasks.

Unlike for effectiveness, we did not conduct a second analysis for efficiency without considering the question about applicability information, because the subjects did not record when they started and finished answering each question. It is an open question to us the extent to which the question about the applicability models affected efficiency. Answering this question without having to e.g. navigate between several nodes of the models probably requires less time. For the follow-up experiment, and after adjusting the design, our hypothesis will be that the efficiency of understanding safety compliance needs is similar with the text of standards and with models. We will try to reject it.

4.3 Perceived benefits in the use of models (RQ3)

Figure 2 shows the results about the subject's perceived benefits in the use of models for understanding safety compliance needs. The numbers in the bars indicate the data points of each possible answer to the corresponding statement.

The median of five out of the seven statements is "Agree", and at least four subjects agreed or strongly agreed upon each statement. We thus conclude that there was an overall agreement upon the benefits in using models. However, the median of statement 6 is "Undecided", and the statement has more disagreement answers than agreement ones. Most subjects further disagreed or strongly disagreed that the models are easy to understand. Therefore, despite the benefits found, models do not seem to be regarded, in general, as better than the text of standards. We interpret this as evidence that the models could be improved. For example, a different graphical representation could be used, other than UML object diagrams. This could contribute to making the models easier to understand, contributing also to both effectiveness and efficiency of the understanding of safety compliance needs.

Regarding the comments and suggestions by the subjects, two subjects explicitly indicated that the models were easier to understand than the text, whereas another indicated that the model was complex. A subject suggested that the best alternative was the combination of text and models. This is supported by the results from some prior studies on other models, e.g. [21].

When having a more detailed look at RQ3 results, we found that the three subjects that had attended a highest number of courses in which they had dealt with UML class or object diagrams (three or four courses) agreed that the models were easy to understand. Two of these subjects also had the largest industrial experience with these diagrams (three and four real projects, respectively). This suggests that knowledge and experience with the modelling language and notation are key so that users find the models of safety compliance needs easy to understand. This can be studied in future experiments with e.g. a longer training session.

We have identified similarities when comparing RQ3 results with those for RQ1 and RQ2. The found benefits in the models can be interpreted as an indicator of their effective support to understand safety compliance needs. In addition, the perceived ease of understanding of models and of the text of standards seem to be similar, thus someone could expect a similar effectiveness for them. The fact that most subjects did not find the models easy to understand could justify the lower efficiency of understanding safety compliance needs with models.

The feedback from practitioners on models created with the holistic generic metamodel [7] has the same median for statements 1, 2, 3, 4, and 7. The median with practitioners is higher for statements 5 and 6, but only slightly. Nonetheless, the graphical representations used by the practitioners were different. They used a graphical editor with a BPMN-like process representation and forms. When compared to the practitioners' feedback on an IEC 61508 model [20], there is a stark difference. Most of these practitioners found the IEC 61508 model easy or very easy to understand. A possible reason is that the IEC 61508 model was created as a class diagram, which might be easier to understand than an object diagram. On the other hand, the feedback from these practitioners was based on a presentation of the model, whereas the feedback of the experiment subjects is based on tasks identifying safety compliance needs. Therefore, the feedback from the subjects might be regarded as more valid.

Figure 2. Experiment results: perceived benefits in the use of models for understanding safety compliance needs

5. CONCLUSION

Understanding safety compliance needs for a critical system can be difficult in practice. Several authors have proposed the use of models as a solution, but there is very little evidence of the extent to which models improve the understanding of the needs.

We have reported a pilot experiment with 15 students conducted to compare the understanding of safety compliance needs with the text of safety standards and with models in the form of UML object diagrams. The overall effectiveness was similar (2% difference), but it increased with models when a question about applicability information was not taken into account (15% gain). The efficiency was higher with the text (24%), but this result might be threatened by the subject's recent experience with the text of a safety standard. All the subjects found some benefit in the use of models, and there was an overall agreement upon models' help in determining how to comply with safety standards and in understanding the relationships between their concepts. However, most subjects considered that the models were not easy to understand, and there is no a clear result about whether the models are easier to understand than the text of safety standards.

As main conclusions, the use of models does not seem to improve the understanding of safety compliance needs as much as expected, and how to make the models easier to understand should be investigated. Nonetheless, new studies are necessary to confirm the initial insights provided, taking into account the possible threats and justifications for the results discussed in Section 4. For the follow-up experiment, we hypothesise that the effectiveness and efficiency of understanding safety compliance needs with the text of standards and with models are similar.

Although models might not contribute to understanding safety compliance largely, we still argue that their use is beneficial for safety certification. The advantages of using models go beyond the understanding of compliance needs, e.g. a model-based representation of safety certification information enables the automated management of the compliance with a standard [20]. Further, there is neither conclusive evidence that using the text of the standards is better to understand safety compliance needs.

As future work, we plan to adjust the current design and run the follow-up experiment. We would also like to conduct a family of experiments on how different notations (e.g. BPMN and goal models) impact the understanding of safety compliance needs.

6. ACKNOWLEDGMENTS

The research leading to this paper has received funding from the AMASS project (H2020-ECSEL grant agreement no 692474; Spain's MINECO ref. PCIN-2015-262) and the AMoDDI project (Ref. 11130583). The authors also thank the subjects of the pilot.

7. REFERENCES

[1] Abdulkhaleq, A. and Wagner, S. 2015. A controlled experiment for the empirical evaluation of safety analysis techniques for safety-critical software. In *EASE 2015.*

[2] Abrahão, S. et al. 2013. Assessing the Effectiveness of Sequence Diagrams in the Comprehension of Functional Requirements. *IEEE T. Software Eng.* 39, 3, 327-342.

[3] Briand, L. et al. 2014. Traceability and SysML design slices to support safety inspections: A controlled experiment. *ACM T. Softw. Eng. Meth.* 23, 1, 9:1-9:43.

[4] Cruz-Lemus, J. A. et al. 2009. Assessing the understandability of UML statechart diagrams with composite states. *Empir. Softw. Eng.* 14, 6, 685-719.

[5] Cyra, L. and Górski, J. 2008. Expert Assessment of Arguments. In *SAFECOMP 2008.*

[6] de la Vara, J. L. et al. 2016. An Industrial Survey on Safety Evidence Change Impact Analysis Practice. *IEEE T. Software Eng.* (accepted paper; preprint available)

[7] de la Vara, J. L. et al. 2016. Model-based specification of safety compliance needs for critical systems: A holistic generic metamodel. *Inform. Software Tech.* 72, 16-30.

[8] De Lucia, A. et al. 2010. An experimental comparison of ER and UML class diagrams for data modelling. *Empir. Softw. Eng.*15, 5, 455-492.

[9] ESA. 2006. Software engineering and standardisation. http://www.esa.int/TEC/Software_engineering_and_standardisation/TECBUCUXBQE_0.html

[10] Hadar, I. et al. 2013. Comparing the comprehensibility of requirements models expressed in Use Case and Tropos. *Inform. Software Tech.* 55, 10, 1823-1843.

[11] Heijstek, W. et al. 2011. Experimental Analysis of Textual and Graphical Representations for Software Architecture Design. In *ESEM 2011.*

[12] Jung, J. et al. 2013. Experimental Comparison of Two Safety Analysis Methods and Its Replication. In *ESEM 2013.*

[13] Mendling, J. et al. 2012. Factors of process model comprehension. *Decis. Support Sys.* 53, 1, 195-206.

[14] Mouaffo, A. et al. 2014. Controlled experiments comparing fault-tree-based safety analysis techniques. In *EASE 2014.*

[15] Nair, S. et al. 2014. An extended systematic literature review on provision of evidence for safety certification. *Inform. Software Tech.* 56, 7, 689-717.

[16] Nair, S. et al. 2014. Safety Evidence Traceability: Problem Analysis and Model. In *REFSQ 2014.*

[17] Nair, S. et al. 2015. Evidence management for compliance of critical systems with safety standards: A survey on the state of practice. *Inform. Software Tech.* 60, 1-15.

[18] OMG. 2015. Structured Assurance Case Metamodel (SACM). http://www.omg.org/spec/SACM/

[19] Ottensooser, A. et al. 2012. Making sense of business process descriptions. *J. Syst. Softw.* 85, 3, 596-606.

[20] Panesar-Walawege, R. K. et al. 2013. Supporting the verification of compliance to safety standards via model-driven engineering. *Inform. Software Tech.*55, 5, 836-864.

[21] Scanniello, G. et al. 2014. On the effect of using SysML requirement diagrams to comprehend requirements: results from two controlled experiments. In *EASE 2014.*

[22] Sharafi, Z. et al. 2013. An empirical study on the efficiency of graphical vs. textual representations in requirements comprehension. In *ICPC 2013.*

[23] Stallbaum, H. and Rzepka M. 2010. Toward DO-178B-compliant Test Models. In *MoDeVVa 2010.*

[24] Wohlin, C. et al. 2012. *Experimentation in Software Engineering*. 2nd ed. Springer, Heidelberg.