de la Vara J.L., Marín B., Ayora C., Giachetti G. (2017) An Experimental Evaluation of the Understanding of Safety Compliance Needs with Models. In: Mayr H., Guizzardi G., Ma H., Pastor O. (eds) Conceptual Modeling. ER 2017. Lecture Notes in Computer Science, vol 10650. Springer, Cham
DOI: https://doi.org/10.1007/978-3-319-69904-2_20

An Experimental Evaluation of the Understanding of Safety Compliance Needs with Models

Jose Luis de la Vara[1], Beatriz Marín[2], Clara Ayora[3], and Giovanni Giachetti[4]

[1] Departamento de Informática, Universidad Carlos III de Madrid, Spain
[2] Facultad de Ingeniería, Universidad Diego Portales, Chile
[3] R&D Department, Treelogic, Spain
[4] Universidad Tecnológica de Chile INACAP, Chile
jvara@inf.uc3m.es, beatriz.marin@mail.udp.cl,
claraayora@gmail.com, ggiachetti@inacap.cl

Abstract. Context: Most safety-critical systems have to fulfil compliance needs specified in safety standards. These needs can be difficult to understand from the text of the standards, and the use of conceptual models has been proposed as a solution. Goal: We aim to evaluate the understanding of safety compliance needs with models. Method: We have conducted an experiment to study the effectiveness, efficiency, and perceived benefits in understanding these needs, with text of safety standards and with UML object diagrams. Results: Sixteen Bachelor students participated in the experiment. Their average effectiveness in understanding compliance needs and their average efficiency were higher with models (17% and 15%, respectively). However, the difference is not statistically significant. The students found benefits in using models, but on average they are undecided about their ease of understanding. Conclusions: Although the results are not conclusive enough, they suggest that the use of models could improve the understanding of safety compliance needs.

Keywords: safety-critical system, safety standard, safety compliance needs, model, understanding, comprehension, experiment.

1 Introduction

Safety-critical systems are those whose failure cam harm people, property, or the environment [12]. These systems must comply with safety standards, e.g., IEC 61508 for a wide range of industries, DO-178C in avionics, EN 50128 in railway, and ISO 26262 in automotive, as a way of assuring that they do not pose undue risks [13]. Safety standards specify safety compliance needs that must be satisfied [7], such as requirements to fulfil, data to manage, and activities to execute. System suppliers must understand and follow these needs, but this can be difficult. The standards are typically large textual documents that consist of hundreds of pages and define thousands of criteria for compliance. Ambiguity and inconsistencies are also usual in their text [12]. Practitioners have indeed acknowledged issues in understanding the standards [5][13].

As a solution, several authors have argued that conceptual models of safety compliance needs can help practitioners understand these needs, e.g. [14]. However, there exists little evidence of the extent to which the use of models improves this

understanding. Prior analyses are either based on experts' perceptions [7][14], not on actual model usage, or have only provided preliminary insights from pilot studies [6]. There is also a general lack of experiments related to safety certification [12].

We aim to fill the gaps regarding the analysis of the understanding of safety compliance needs with models. To this end, we have conducted an experiment to study the effectiveness, efficiency, and perceived benefits of understanding the needs with models. Sixteen Bachelor students answered questions about safety compliance needs in DO-178C and in EN 50128, using their text and models (UML object diagrams). The students also indicated their opinion about the use of models.

The paper is organised as follows. Section 2 presents the background, and Section 3 the experiment process. Section 4 reports the results and Section 5 our conclusions.

2 Background

Model-based approaches for the specification of safety compliance needs have been proposed for specific standards or parts of them (e.g. IEC 61508 [14]), and for specific compliance needs (e.g. related to processes [3]). Modelling standards for system assurance and certification have also been published [8]. Some studies have reported that models are used in industry for safety certification purposes [5][13].

For the experiment, we have used a holistic generic metamodel for the specification of safety compliance needs [7]. This metamodel supports the specification of different types of these needs: information about requirements, artefacts, and processes, and about their applicability. The metamodel can be used for different standards from several domains and has been validated with practitioners and data from real projects.

Regarding **related work**, we run a pilot experiment [6] to validate the experiment design, adjust it for the experiment reported in this paper, and derive hypotheses. We found both evidence and counterevidence of the improvement in the understanding of safety compliance needs with the use of models.

In other studies, experts have agreed that models of safety standards are easy to understand [14][7]. There are also some experiments related to safety certification (e.g. [1][4]), including on model-based approaches. Experiments that have evaluated the comprehension of model-based artefacts (e.g. [2][9]) have shown benefits in their use. Others have compared textual and graphical representations (e.g. [15][17]). The results of understanding tasks with models were better in some cases, and with text in others.

3 Experiment Process

We used the guidelines by Wohlin et al. [19] to design the experiment. The goal is to analyse the use of models to specify 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 and Engineering.

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 to understand safety compliance needs?

The subjects of the experiment are 16 students of a 3rd-year course on "Software development projects management" of a Bachelor's Degree in Computer Science and Engineering at Carlos III University of Madrid, Spain. In this course the students have to plan the development and validation of an application and to design it according to the ESA PSS-05-0 software engineering standard [10]. In the experiment the subjects have to identify safety compliance needs from excerpts of the text of safety standards and from models of these excerpts, and indicate their opinion about the models.

Based on the results of the pilot experiment [6], we formulate two null hypotheses that we aim to reject:

- $H_{1,0}$: There is no significant difference in the effectiveness of understanding safety compliance needs with the text of safety standards and with models.
- $H_{2,0}$: There is no significant difference in the efficiency of understanding safety compliance needs with the text of safety standards and with models.

The independent variables are: (1) the means used to represent safety compliance needs (model or text), and; (2) the standard considered (DO-178C requirements process or EN 50128 integration process, which are different to the standard used in the course). To represent the instances of the holistic metamodel, we use UML object diagrams.

Two dependent variables are the effectiveness and efficiency. In line with related work, e.g. [2][4], we use the F-measure (F_s) to quantify the effectiveness. It is based on the precision and recall in identifying safety compliance needs. We use the formulas for cases in which it is possible that a subject does not answer a question [9]. We use the effectiveness and the time (in minutes) to quantify efficiency ($Effy_s$) [1][15].

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

$$F_s = 2 \times \frac{precision_s \times recall_s}{precision_s + recall_s} \qquad Effy_s = 100 \times \frac{F_s}{minutes}$$

The third dependent variable is the perceived benefits in understanding safety compliance needs. It is evaluated with a questionnaire and a 5-point Likert scale (see Section 4.3) about the use of models to specify and to understand the needs [7].

The subjects are randomly divided into four groups in a within-subject 2x2 factorial design [18]: (1) DO-178C model (for the first task) and EN 50128 text (for the second task); (2) EN 50128 model and DO-178C text; (3) DO-178C text and EN 50128 model, and; (4) EN 50128 text and DO-178C model. The execution of the experiment is planned for 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 material preparation and the rest of authors validated it.

The subjects work offline and with the material[1] of each task printed: an introductory page, a two-page excerpt of a standard or models of the excerpts, and seven free-text questions. The subjects have to identify 11 safety compliance needs to correctly complete the questionnaire, the same in the text and in the model. The subjects need to

[1] https://sites.google.com/site/jldelavara/material/msac2016

record the time when they start and finish each task, and complete an opinion questionnaire.

Despite our effort to ensure experiment **validity**, some threats could impact it. For internal validity, we mitigated fatigue effects by running the experiment in the morning and having a break between the training and the tasks. Learning effects were mitigated by using different experimental objects, with similar size and complexity, in the two tasks. Regarding external validity, the use of students as subjects might concern the generalization of results. Nonetheless, recent studies argue that there are minor differences when students or practitioners are used [16]. Students can be regarded as novice practitioners [2], and it cannot be claimed that experience greatly helps practitioners better understand safety compliance needs [5]. We are also aware that the sample size is limited, but the number of students of the course was a constraint. The creation of the experimental material might be threatened by the interpretation of the standards (construct validity). To mitigate this threat, we used parts of standards for which we had access to models validated by practitioners. For conclusion validity, we use dependent variables that are widely used in experiments with a similar purpose, e.g. [2][4]. To analyse the statistical significance of the results, we use parametric tests when normality of data was confirmed and non-parametric tests otherwise, and a 0.05 level for the p-value. Finally, the selection of a given graphical notation (UML object diagram) affects conclusion validity.

4 Results and Interpretation

This section presents the results of the experiment and how we interpret them. No subject had knowledge about the standards used in the experiment or the parts of them. Their experience with UML class or object diagrams was homogeneous and similar to our expectations for 3rd-year Bachelor students in Computer Science and Engineering.

4.1 Effectiveness of Understanding (RQ1)

Table 1 shows the effectiveness of understanding safety compliance needs with models and with the text of standards. In addition to the value of the F-measure for each subject (F), the table shows the precision (P) and recall (R). Their mean values are similar to or higher than those in other experiments related to safety certification, e.g. [1][4], thus we regard subjects' overall effectiveness as acceptable and valid.

The mean effectiveness with models is 17% higher than with the text of standards, and the median is 30% higher. This initial overall result suggests that the use of models improves the effectiveness of understanding safety compliance needs. According to the Shapiro-Wilk test, the sample for effectiveness with models is non-normal (p-value = 0.049 < 0.05), thus we selected the Wilcoxon test for $H_{1,0}$. The test result determines that the difference in the effectiveness when using models is not statistically significant (p-value = 0.096 > 0.05). Therefore, $H_{1,0}$ cannot be rejected and the results are not conclusive enough to confirm that the use of models improves the effectiveness of understanding safety compliance needs.

Despite the lack of statistical significance, we argue that most of the evidence from the results suggests that the use of models could improve the effectiveness of

understanding compliance needs. In addition to the differences of the means and the medians, the effectiveness with models is higher for 12 out of the 16 subjects (75%). The highest effectiveness (0.87) is with models, as a result of the highest precision (0.83) and recall (0.91). The effectiveness is above 0.7 for six subjects with models and for only two with text. We conjecture that the lack of statistical significance is due to sample size. This could be addressed in follow-up experiments. We have not observed any potentially relevant correlation between subject's experience and effectiveness.

When comparing the results with those from the pilot experiment [6], we consider that the results are coherent. The initial average gain in effectiveness from using models in the pilot was a 2%, but it raised up to 15% when an issue with a question about applicability information was taken into account. This is close to the 17% average gain in the experiment.

Table 1. Effectiveness and efficiency of understanding safety compliance needs

		Effectiveness						**Efficiency**			
		Models			**Text**			**Models**		**Text**	
Group	**Subj.**	P	R	**F**	P	R	**F**	T	**Effy**	T	**Effy**
1	**1**	0.67	0.91	**0.77**	0.55	0.55	**0.55**	18.88	**4.07**	13.75	**3.97**
	2	0.18	0.27	**0.21**	0.82	0.82	**0.82**	19.5	**1.1**	16	**5.11**
	3	0.83	0.91	**0.87**	0.5	0.45	**0.48**	18.65	**4.66**	25.73	**1.86**
	4	0.5	0.73	**0.59**	0.64	0.64	**0.64**	26.03	**2.28**	16.57	**3.84**
2	**5**	0.67	0.73	**0.7**	0.38	0.45	**0.42**	14	**4.97**	11.63	**3.58**
	6	0.75	0.82	**0.78**	0.47	0.73	**0.57**	18.08	**4.33**	17.58	**3.25**
	7	0.69	0.82	**0.75**	0.2	0.36	**0.26**	19.23	**3.9**	18.42	**1.40**
	8	0.62	0.73	**0.67**	0.36	0.36	**0.36**	26.12	**2.55**	14.38	**2.53**
3	**9**	0.62	0.73	**0.67**	0.4	0.56	**0.46**	22.92	**2.91**	17.5	**2.64**
	10	0.33	0.36	**0.35**	0.24	0.56	**0.33**	16.12	**2.16**	21.33	**1.56**
	11	0.87	0.64	**0.74**	0.62	0.73	**0.67**	15.77	**4.67**	21.93	**3.04**
	12	0.58	0.64	**0.61**	0.41	0.64	**0.5**	21.05	**2.89**	26.42	**1.89**
4	**13**	0.31	0.36	**0.33**	0.67	0.73	**0.7**	15.28	**2.18**	22.98	**3.03**
	14	0.29	0.45	**0.36**	0.8	0.73	**0.76**	21.32	**1.68**	25.32	**3.01**
	15	0.64	0.64	**0.64**	0.5	0.55	**0.52**	19.5	**3.26**	29.93	**1.74**
	16	0.56	0.82	**0.67**	0.33	0.18	**0.24**	29.83	**2.23**	32.58	**0.72**
	Mean	0.57	0.66	0.61	0.49	0.56	0.52	20.14	3.12	20.75	2.7
	Median	0.62	0.73	0.67	0.49	0.55	0.51	19.37	2.9	19.87	2.8
	Std. deviation	0.2	0.2	0.19	0.18	0.17	0.17	4.33	1.19	6.042	1.14

4.2 Efficiency of Understanding (RQ2)

Table 1 shows the results regarding efficiency of understanding safety compliance needs with models and with text. The table includes the data of the time spent in the tasks (T; in minutes) and the efficiency outcome (Effy). The mean effectiveness with models is 15% higher and the median 3%. The results from the Shapiro-Wilk test for normality shows that both the sample for efficiency with models and the sample for efficiency with text of safety standards are normal (p-value > 0.05). Therefore, we selected the paired t-test for $H_{2,0}$. The test result determines that the difference in efficiency is not statistically significant (p-value = 0.173 > 0.05). Thus, $H_{2,0}$ cannot be

rejected and the results are not conclusive enough to confirm that the use of models improves the efficiency of understanding safety compliance needs.

Although there is no statistical significance, some aspects of the results make us believe that the use of models could improve the efficiency of understanding of safety compliance needs. We have argued above that the results suggest that effectiveness could increase with models, and efficiency is directly based on effectiveness. The efficiency is above 4.0 for four subjects when using models, and only for one when using text. The lack of statistical significance might be an effect of sample size.

As counter evidence of the increase in efficiency when using models, the average time to execute the tasks is only a 3% higher with the text of safety standards. With such a little decrease in time when using models, it is not likely that efficiency improvement is significant unless effectiveness improvement also is. The mean gain in efficiency with models is also lower (15%) than the mean gain in effectiveness (17%). Finally, the efficiency is above 3.0 for seven subjects when using models and for eight when using the text.

In the pilot experiment [6] the efficiency of understanding compliance needs with models was quite lower than with the text (24%). This might have been a result of issues in the experimental design that the adjustments for this experiment have mitigated.

4.3 Perceived Benefits in the Use of Models (RQ3)

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

The median of four statements is Agree, and at least three subjects strongly agreed on them. No subject disagreed that "The models help in understanding the relationships between the concepts", and the statements with the highest number of subjects that disagreed or strongly disagreed are "The models help in understanding the concepts" and "The models are easy to understand" (7 subjects; 44%). In addition to the latter statement, some subject strongly disagreed that "The models are easier to understand than the text of the safety standards I have dealt with". "The models are easy to understand" is also the only statement for which no subject strongly agreed.

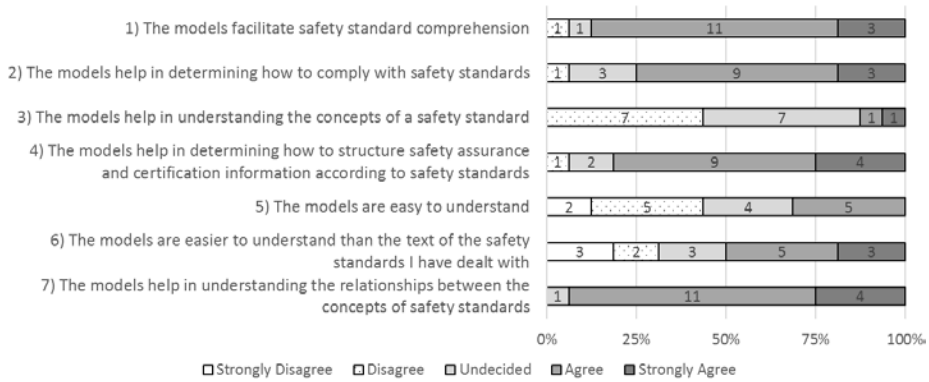

Fig. 2. Perceived benefits in the use of models to understand safety compliance needs

Despite the overall benefits found, the models do not seem to be regarded as easy to understand or easier to understand than the text. This could be due to the graphical notation used in the experiment. The experience with UML might also influence the perceived benefits. We plan to gain deeper insights into this aspect by running the experiment with students of courses on model-driven engineering.

In the pilot experiment [6], the widest agreement was on "The models help in understanding the relationships between the concepts" too, and the ratio of subjects that disagreed or strongly disagreed that "The models are easy to understand" was higher. The latter is also the only statement for which some practitioner disagreed in [7], and all the practitioners agreed or strongly agreed upon the former. Interestingly, the median in the study with practitioners, the pilot experiment, and the experiment for "The models are easier to understand than the text of the safety standards I have dealt with" is Undecided or Undecided-Agree. This supports the proposal of investigating notations that could be more suitable to represent compliance needs. Different graphical notations might help to increase the perception of the benefits.

Most of the practitioners that provided feedback on a model of IEC 61508 [14] regarded it as easy to understand. The model was presented as a class diagram, and these practitioners might have more experience with UML than our subjects. In experiments on security assessment (e.g. [11]), the number of positive aspects regarding perceived ease of use and perceived usefulness was higher for models than for text.

5 Conclusion

The textual descriptions of compliance needs in safety standards can be difficult to understand. The use of conceptual models has been proposed as a solution, but there is a lack of empirical evidence that confirms the benefits of this usage. This paper has presented an experiment with 16 subjects, separated into four different groups, that interpreted models and textual specifications of safety compliance needs. The results show that the use of models can improve the effectiveness and efficiency of understanding safety compliance needs by 17% and 15%, respectively. However, this does not guarantee statistical significance of the advantage in using models to understand safety compliance. This makes it impossible to reject the hypotheses formulated. Further experiments are needed to obtain more conclusive results.

From a deeper analysis, we have observed that the representation of applicability information seems to be more effective in the text of safety standards than in models. We conjecture that the use of a hybrid specification, combining graphical modelling and tables, could be an alternative to study. Another aspect to consider is the use of specific notations to model safety compliance needs instead of existing notations such as the UML object diagrams used. Finally, although the use of models might not significantly improve the understanding of safety compliance needs, it can still be beneficial for safety certification, e.g. for automated compliance management [14].

As main future work, we plan to conduct new experiments to evaluate different modelling approaches to specify safety compliance needs (e.g. BPMN and goal models). We expect that, as a consequence, we will be able to draw stronger conclusions and to guide the selection of adequate specification style alternatives according to the safety compliance needs to be represented.

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). We also thank the subjects that participated in the experiment.

References

1. Abdulkhaleq, A, Wagner, S.: A controlled experiment for the empirical evaluation of safety analysis techniques for safety-critical software. In: EASE 2015, pp. 16:1-16:10.
2. Abrahão, S. et al.: Assessing the Effectiveness of Sequence Diagrams in the Comprehension of Functional Requirements. IEEE T. Softw. Eng. 39(3), 327-342 (2013)
3. Ayora, C., et al.: Variability management in process families through change patterns. Inform. Softw. Tech. 74, 86-104 (2016)
4. Briand, L., et al.: Traceability and SysML design slices to support safety inspections: A controlled experiment. ACM T. Softw. Eng. Meth. 23(1), 9:1-9:43 (2014)
5. de la Vara, J.L., et al.: An Industrial Survey on Safety Evidence Change Impact Analysis Practice. IEEE T. Softw. Eng. 42(12), 1095-1117 (2016)
6. de la Vara, J.L., et al.: Do Models Improve the Understanding of Safety Compliance Needs? Insights from a Pilot Experiment. In: ESEM 2016, pp- 32:1-32:6.
7. de la Vara, J.L., et al.: Model-based specification of safety compliance needs for critical systems: A holistic generic metamodel. Inform. Softw. Tech. 72, 16-30 (2016)
8. de la Vara, J.L., et al.: An analysis of safety evidence management with the Structured Assurance Case Metamodel. Comput. Stand. Interfaces 50, 179-198 (2017)
9. De Lucia, A., et al.: An experimental comparison of ER and UML class diagrams for data modelling. Empir. Softw. Eng. 15(5), 455-492 (2010)
10. ESA. Software engineering and standardisation (2006) http://www.esa.int/ TEC/Software_engineering_and_standardisation/TECBUCUXBQE_0 html
11. Labunets, K., et al.: An Experimental Comparison of Two Risk-Based Security Methods. In: ESEM 2013, pp 163-172.
12. Nair, S., et al.: An extended systematic literature review on provision of evidence for safety certification. Inform. Softw. Tech. 56(7), 689-717 (2014)
13. Nair, S., et al.: Evidence management for compliance of critical systems with safety standards: A survey on the state of practice. Inform. Softw. Tech. 60, 1-15 (2015)
14. Panesar-Walawege, R.K., et al. Supporting the verification of compliance to safety standards via model-driven engineering. Inform. Softw. Tech. 55(5), 836-864 (2013)
15. Razali, R., et al.: Experimental Comparison of the Comprehensibility of a UML-based Formal Specification versus a Textual One. In: EASE 2007
16. Salman, I., et al.: Are Students Representatives of Professionals in Software Engineering Experiments? In: ICSE 2015
17. Sharafi, Z., et al.: An empirical study on the efficiency of graphical vs. textual representations in requirements comprehension. In: ICPC 2013
18. Vegas, S., et al.: Crossover Designs in Software Engineering Experiments: Benefits and Perils. IEEE T. Softw. Eng. 42(2), 120-135 (2016)
19. Wohlin, C, et al. Experimentation in Software Engineering (2nd ed.). Springer (2012)