

Empirical Validation of a Software Requirements Specification Checklist

Martin de Laat, Maya Daneva

University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
martindelaat90@gmail.com, m.daneva@utwente.nl

Abstract. [Context/Motivation] For areas such as Government IT Procurement, the Software Requirements Specification (SRS) often forms the basis for a public procurement. In these cases, having domain knowledge is often mutually exclusive to knowing RE. Domain experts lacking the necessary RE experience face issues assessing the quality and correctness of the SRS. This especially forms a problem for situations where the SRS acts as the base for a proposal and resulting contract, which is why often third party RE experts are consulted for evaluating the SRS beforehand. These experts are highly motivated to improve their process and offer a more uniform and better service. [Question/problem] Is our developed checklist a valid instrument to support the RE practitioner in the SRS validation process? [Principal ideas/results] We propose to empirically evaluate the checklist in a live study. Participants of our Evaluation Study are asked to simulate the validation of a sample SRS, guided by our checklist. We will analyze the data from a post-use questionnaire at the end of the session using a mixed method approach to assess the quality and usability of our checklist and its expected impact on the validation process. This assessment will be part of the overall validation of our checklist. [Expected Contribution] We expect to gain knowledge regarding the quality of our instrument. Secondly, this live study contributes to the validation, and thus the realization of a practical tool to be used by RE practitioners worldwide.

Keywords: Requirements engineering practice, RE, Software Requirements Specification, validation, checklist, empirical study

1 Research design and objectives

1.1 Research problem

For areas such as Government IT Procurement, the Software Requirements Specification (SRS) often forms the basis for a public procurement, e.g. through a request-for-proposals process. In many situations, having domain knowledge is mutually exclusive to knowing RE. This introduces two challenges, namely (1) the persons with domain knowledge lacking the tools to properly create and/or assess the SRS and (2) the RE experts potentially not having sufficient domain knowledge to cover. The use of an instrument supporting the SRS validation

process could be of help in both scenarios. Whenever a call for bid is involved, the requirements specification will form a basis for the contract. Not only does having a good SRS prevent ambiguity between the procurer and the supplier, it also acts as a safeguard against any legal problems further down the road. Because of this, validation is an integral part of the overall creation of an SRS. It is of importance that a representation of the stakeholders c.q. domain experts within the procurer's company understand and agree with their respective parts of the SRS and can reasonably assess its quality. Often, third party experts are contracted to validate the SRS before initiating the IT procurement process. These experts are highly motivated to improve their process and offer a more uniform and better service.

1.2 Motivation and research goal

Our motivation stems from improving the validation process in order to contribute to better software being developed. The goal of this live study is to evaluate our instrument (the checklist) and its effect on the validation process of a given SRS. The live study will most notably help identify strengths and weaknesses, identify possible missing elements and give insights to the applicability of the checklist by a variety of users [11].

1.3 Positioning of this live study within our research project

The project resulted as a collaboration of the first author and a large consulting company in the Netherlands acting as a third party expert in the validation of SRS's in the Government IT Procurement area. This research project started within the University of Twente's course on Advanced Requirements Engineering taught by the second author. The instrument is now being developed as part of the Master's Project for Computer Science at the University of Twente. Following [10], an instrument is first designed and its design is justified, and then evaluated gradually through sequenced applications in specific contexts from which lessons can be learned about the instrument's use and its improvement. The checklist itself is developed following the steps outlined by Stufflebeam [9]. The checks defined in our instrument are based on the results of a literature study and by combining elements of industry standards such as [3] (superseding [2]) and [8]. Interview sessions with field experts from the company mentioned above where the instrument was showcased have already taken place and have yielded positive results. Our checklist is now in a stage in which it should be empirically evaluated for its quality and desired effect. This proposal is focused on the design of the live study at REFSQ 2018 where it will be the first time it's applied in a test setting. Future plans consist of testing the use of the instrument in near-to-real-life settings.

1.4 Research objectives and research questions

The main goal of this live study is to get a hands-on impression of the quality of the instrument by having it applied by a variety of participants on a given

SRS. A side benefit of the live study is the announcement of our research to the scientific community and to gain traction amongst it. Specifically, we want to know what the strong and weak points of the checklist are and whether there are any missing elements (checks). To this end, we ask the following research questions:

RQ1 What is the quality of the checklist based on the selected criteria?

RQ2 What is the effect of the instrument on the validation process?

1.5 Research Methodology

For this specific live study, we consulted the empirical evaluation guidelines provided by Wieringa[10], which are general to the evaluation of any software engineering artifact or technology in context. In our evaluation study, participants will be provided with the materials mentioned in 2 and asked to validate a sample SRS with the help of our proposed instrument. Participants are invited to share their perceptions and experiences by completing a questionnaire. The quality of the instrument will be assessed based on criteria mentioned in section 2.2, rated by the participants in a questionnaire. This data is analyzed with an extended version of the mixed method approach by Martz[4]. The checklist consists of 112 checks. To cover all checks within the timespan of the session they are split into a 'core'- and four 'non-core' segments. The core segment consists of 30 checks and the remaining checks are evenly distributed into four groups. Each participant will cover the 'core' segment and one of the 'non-core'-segments. During the live session the four variations of materials will be distributed evenly.

1.6 About the authors

The first author, de Laat, is a Master's student in the Computer Science program of the University of Twente. He earned a bachelor's degree in Business & IT from the University of Twente. He has a broad professional experience in developing business solutions for Microsoft SharePoint and is currently a part-time, Product Designer and Ruby On Rails developer at Nedap Healthcare.

The second author, Daneva, has been using empirical evaluation techniques in her research since 2001. In 2012, she co-authored an online study that was featured in the REFSQ program, and was about evaluation of a checklist for reporting empirical research in RE.

2 Live study design and logistics

The study participants (50-60 expected) will be drawn from the pool of attendees at REFSQ 2018. We require no particular participant profile. For the research to produce informative results, we need at least 20 participants so that we have at least 5 participants working on each non-core segment of the checklist. All participants will attend jointly the introduction to our live study. After this, the participants will be divided into the available rooms to make sure everybody

has a seating spot and sufficient space. Participants will receive one of four sets of materials (see section 2) and asked to apply their respective set of checks during a 60 minute session. Afterwards, during a 10 minute period, participants are asked to fill in a questionnaire in order to provide perceptions, opinions and their personal observations from using the checklist. The final 10 minutes are reserved for the conclusion, thanking and gathering of the materials.

2.1 Equipment and materials involved in the study

We need the help of the local organization to book sufficient rooms where the participants can work in isolation from each other. Participants will be provided with (1) A sample (short) SRS, (2) their selected set of 'core' and 'non-core' checks including relevant information from the reference manual, (3) an answer-sheet with check-boxes to mark the results of the application of the check, (4) the post-use questionnaire. Those that brought a mobile device will be requested to fill in the questionnaire online using SurveyMonkey to speed up analysis. The checks given to the participant are described in full, checks that are not will be referenced by name only to allow for the participant to be able to say something about the completeness. Everybody will be given the option to leave their contact details if they desire to be kept informed about the research.

2.2 Post-Use Questionnaire

During the live study, all participants will be asked to fill in a questionnaire consisting of

1. A set of background questions assessing: (a) the participant's experience with SRS's, (b) which version of the non-core sets the participant received and (c) an indication of how much time they spent on the core v.s. the non-core part.
2. A critical feedback survey consisting of: (a) a set of closed questions and (b) a set of open questions

In the closed questions section the participants are requested to rate aspects of the checklist based on a 9-point Likert scale where 1 means "strongly disagree" and 9 means "strongly agree". The first set of closed questions are derived from [9] and test the instrument on its: applicability to the full range of intended use, clarity, comprehensiveness, concreteness, ease of use, fairness, parsimony and item pertinence to the content area.

The second set of questions, put together by the authors, expands on this list and asks the participants whether they find the instrument: to help prevent task saturation [6], to fit the work flow [1], can be completed in a sufficient period of time [1], contains sufficient break points, will contribute to the *process* of validating an SRS and will improve the quality of the validation output in a real life scenario.

The set of open questions consists of four open questions[4] where the participant is asked to identify possible strengths, identify possible weaknesses, identify

items missing from the checklist and give recommendations for improving the checklist.

Finally, the participant is given the option to write down any final thoughts regarding the instrument or the session. On a separate document the participants will be able to leave their contact details in case they wish to receive further notifications regarding the development of the instrument.

2.3 Data collection and analysis

Data resulting from both the paper and electronic questionnaires will be combined and analyzed using off-line tooling on a secure computer. The recording units (a single word or phrase describing a strength or weakness) resulting from the open questions are grouped into categories identical to those of the closed questions. The categories of the first set of closed questions will attribute to answering RQ1 and the second set to answering RQ2. Diverging stacked bar charts and a Grouped Bar Chart will help visualize the quantitative responses as described by Robbins, N. B., & Heiberger, R. M. [7]. In an effort to compare the responses from the open questions to those of the closed questions, we will plot the 'Mean Scores' against the 'Net Strength' as proposed by Martz [4]. The final facet of validity addressed in the investigation, will be consequential validity. That is, providing evidence and rationale for evaluating the intended and unintended consequences of interpretation and use [5].

2.4 Threats to validity

Internal threats for validity include the possibility that some participants have experience in using checklists. Also, it may happen that some participants spend more time on one section than the other (for instance, the core v.s. non-core parts). This all may affect our results. However, we plan to mitigate these threats by collecting information on the prior checklist-related experience of the participants and on how much time they spent on core v.s. non-core checks. If we get less than 20 participants we would only be able to assess the checklist based on its 'core' contents. An external threat to validity is that the provided sample SRS might not be representable to those used in a real scenario, affecting for instance the applicability of the instrument. We attempt to mitigate this by selecting a reasonably standard sample SRS. Outside the scope of this live study, we plan to mitigate this in future test sessions by having multiple experts test a diverse set of requirements specifications.

2.5 Promotion, incentives and the sharing of the data

We will work with the local organization so that our invitation to the study is received by each REFSQ attendee upon registration. During the conference we aim to share our findings but might not have sufficient time to analyze all results, meaning we would have to limit our analysis to data derived from the

online responses. Apart from contributing to our research, participants will have the option to voice their opinions and/or concerns regarding the development of the checklist. Participants who chose to leave their contact information will be kept informed about the progress of the instrument and will be sent a copy of the resultant paper(s). After the conference the results of the live study will be part of a Technical Report issued by the University of Twente. We are planning of writing a journal paper in which we will motivate our checklist proposal and resent its empirical evaluation. This live study results will be also included in the journal submission.

2.6 Ethics, confidentiality and consent

The filling in of the questionnaire will be done anonymously and participants will be asked to give consent on the analysis, manipulation and publication of the data they provide during the live study. Apart from assessing the participants' experience with requirements specifications, no personally identifiable information will be asked of the participants during the test session. Optional participants' contact details will be collected on a separate document to guarantee the results of the questionnaire cannot be traced back to the participant.

References

1. Gawande, A.: Checklist Manifesto, The (HB). Penguin Books India (2010)
2. IEEE: IEEE 830: Recommended Practice for Software Requirements Specifications. Tech. rep. (1998)
3. ISO, I.E.C.: IEEE. 29148: 2011-Systems and software engineering-Requirements engineering. Tech. rep. (2011)
4. Martz, W.: Validating an evaluation checklist using a mixed method design. *Evaluation and Program Planning* 33(3), 215–222 (2010)
5. Messick, S.: Validity of Psychological Assessment. *American Psychologist* 50(9), 741–749 (1995)
6. Murphy, J.D.: *Business is Combat: A Fighter Pilot's Guide to Winning in Modern Warfare*. Harper Collins (2010)
7. Robbins, N.B., Heiberger, R.M.: Plotting Likert and Other Rating Scales. *Joint Statistical Meetings* pp. 1058–1066 (2011)
8. Robertson, J., Robertson, S.: *Volere. Requirements Specification Templates* (2000)
9. Stufflebeam, D.L.: *Guidelines for developing evaluation checklists: the checklists development checklist (CDC)*. Kalamazoo, MI: The Evaluation Center. Retrieved on January 16, 2008 (2000)
10. Wieringa, R.: *Design Science Methodology for Information Systems and Software Engineering* (2014), <http://portal.acm.org/citation.cfm?doid=1810295.1810446>
11. Wieringa, R., Daneva, M.: Six strategies for generalizing software engineering theories. In: *Science of Computer Programming*. vol. 101, pp. 136–152 (2015)