Improving the measurement of the understandability of requirements

Joseph Kasser, William Scott, Xuan-Linh Tran, Sergey Nesterov
Systems Engineering and Evaluation Centre
University of South Australia
Mawson Lakes, SA 5095

Abstract

Recognizing that the protection of information and infrastructure that are used to create, store, process and communicate information is vital to the productivity and security of our nation, and to assure continuity and growth of our infrastructure-based society, defining current and emerging requirements for information security education is as, if not more, vital. This is because if the requirements are not correct, the results will not meet the needs. Yet poor requirements have been cited as a major cause of cost and schedule overruns in the acquisition of Defence and information systems for at least 15 years. Yet in all this time, academia has not been able to alleviate the problem through traditional teaching techniques. Moreover the commercially available requirements tools manage sets of requirements but do not provide the capability to determine if a requirement is good or bad. This paper summarizes the concept behind a prototype educational and research software tool that has helped students to understand the difficulty in writing good requirements, reports on some results in the classroom, discusses the difference between readability and understandability and proposes a research program to develop a metric for the understandability of requirements.

Background

Poor requirements have been cited as a major cause of cost and schedule overruns in the acquisition of Defence and information systems for at least 15 years, see (Carson, 2001;Hooks, 1993;Jacobs, 1999;Kasser and Schermerhorn, 1994) etc. Yet in all this time, academia has not been able to alleviate the problem through traditional teaching techniques. In an attempt to alleviate the problem through education a postgraduate Requirements Workshop for the writing of good requirements was first developed at the University of Maryland in 1999 and later incorporated into classes at the University of South Australia. The workshop entailed students reviewing and commenting on a sample requirements document. The workshop was limited by time so that only a few pages of the document were reviewed by the students. The focus of the discussion during the workshop was on the grammar and structure of requirements. Moreover, when students tallied their findings of good and bad requirements, there was little if any consensus on the quality of the requirements.

Using software to detect poor requirements

Realizing that the parsing process which was taking up much of the student's time in the workshop could be automated, a prototype software tool was developed to perform the function (Kasser, 2002). The tool described by (Kasser, 2002) parsed the text of the requirement looking for the presence of "poor words" that indicated that a defect or a potential for a defect existed. The list of poor words was gathered from experience and the literature (Hooks, 1993;Kar and Bailey, 1996;Kasser and Schermerhorn, 1994). The tool evolved into FRED (Kasser, 2004) , then Tiger and finally into Tiger Pro (Kasser, Tran, et al., 2003).

FRED and its successors tested the wording of requirements based on a set of requirements for writing requirements (Kasser and Schermerhorn, 1994). One of the categories of defects in those requirements is "a requirement shall not be vague". Antonyms of *vague* are "clear" and "understandable". Hence there should be no surprise to learn that in a later listing of the criteria of a good requirement (Schneider and Buede, 2000), one criterion was understandability. Now, there does not seem to be any way of directly measuring "vague" or "understandability". Understandability of a statement can be tested in the following two ways:

- Confirming that the statement conveys the same message to all readers.
- Does it contain more than the minimum amount of text needed to convey the desired information.

Measuring the readability of requirements

Readability seems to have been used in the field of education as a measure of the understandability of a written text. However, as part of upgrading ongoing courses in mid May 2005, Tiger Pro was used in a Requirements Workshop in which there were 85 participants who were professionals in the Defence Industry involved in writing or using requirements. In this workshop, pre- and post- workshop questionnaires were handed out to participants. The questionnaires comprised a set of 37 good and defective requirements. For the pre-workshop assessment, participants were required to read each requirement and decide if it was good or bad. In the post-workshop assessment, the same set of requirements was provided but the order was randomly mixed. In addition, the participants were required to read each requirement and decide if it was good or bad. They were also required to state the type of defect they found in the requirements.

Comparing the number of correct decisions from the post-workshop questionnaire with the pre-workshop questionnaire showed an improvement of 71%. It was also noticed that some requirements proved to be difficult for most participants and some participants were confused when identifying different types of defect. To assess whether the level of reading difficulty in the requirements given in the workshop had any affect on the participants' responses, both the pre- and post- workshop correct answers for each requirement were plotted against the Flesch-Kincaid Grade Level readability score provided in Microsoft Word for that requirement as shown in Figure 1 and Figure 2 (Tran and Kasser, 2005).

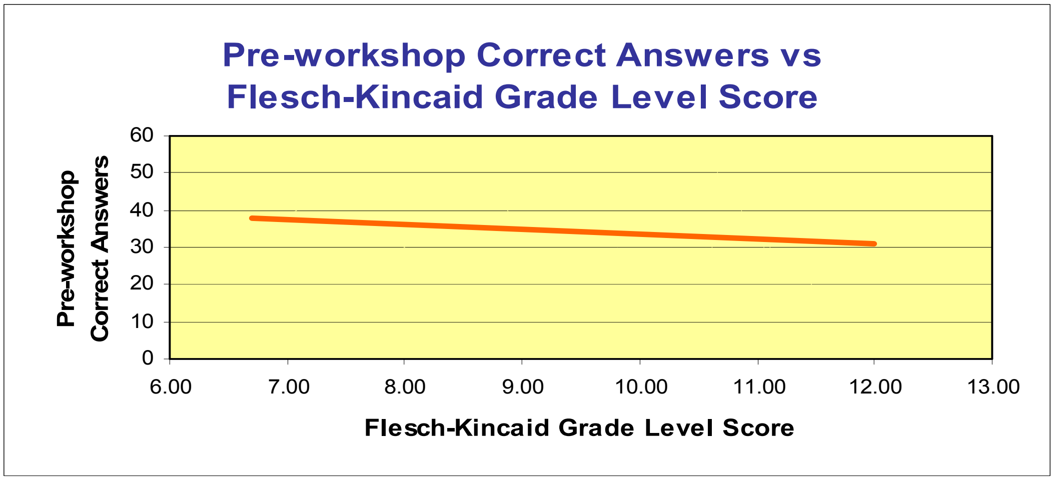

Figure 1 Pre-workshop correct answers vs Flesch-Kincaid Grade Level Score

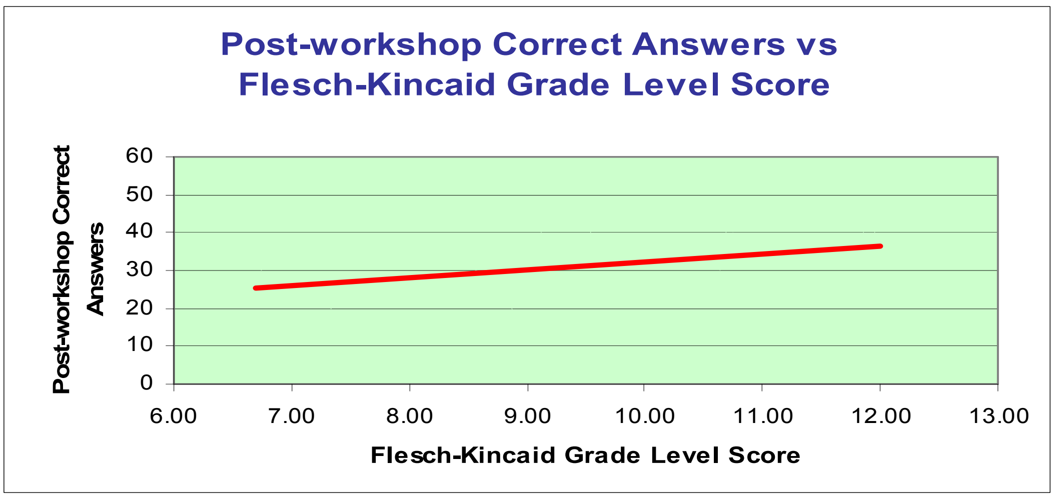

Figure 2 Post-workshop correct answers vs Flesch-Kincaid Grade Level Score

The following deductions can be made from Figure 1 and Figure 2.

- In the pre-workshop survey (Figure 1), as the reading grade level of the requirements increases, the number of correct answers decreases.
- In the post-workshop survey (Figure 2), as the reading grade level of the requirements increase, the number of correct answers also increases.

The first finding seems to be intuitive once articulated. However, the only reference to the reading level of a requirement that has been found in the systems engineering literature is (Wilson, Rosenberg, et al., 1997) which discusses the development of an Automated Requirements Measurement (ARM) tool by the Software Assurance Technology Center at the Goddard Space Flight Center. The ARM tool focused on the grammar of the sentence and the use of “weak phrases” or “poor words” such as “large”, “rapid” and “many”. The ARM tool did not attempt to assess the correctness of the document, rather it assessed the structure of the requirements document and the vocabulary used to state the requirements based on the desirable characteristics for requirements specifications (IEEE 830-1993). While the ARM tool does provide the four readability statistics provided by Microsoft Word for the requirements specification document, as of January 29, 2003 it did not seem to make use of the statistics.

The second finding needs to be researched. The obvious conclusion is that the workshop focussed on the more complicated statements and improved their comprehension. However, this is probably not true and further research is needed.

Determining a metric for understandability

One avenue may be in the difference between readability and understandability. The formula for the Flesch-Kincaid Grade Level score is (.39 x ASL) + (11.8 x ASW) – 15.59 where:

- ASL = average sentence length (the number of words divided by the number of sentences)
- ASW = average number of syllables per word (the number of syllables divided by the number of words)

However, in the wording of requirements, short words such as “it” and acronyms, while improving the Flesch-Kincaid Grade Level score tend to reduce the understandability of the text. The pronouns have the potential to introduce ambiguity and unfamiliar acronyms increases the difficulty of understanding the text hence lowering the understandability of the requirement text. Thus well-written requirements may not have good Flesch-Kincaid Grade Level scores. This also applies to any readability metric that incorporates a measure of the number of syllables in the sentence such as the Flesch-Kincaid Reading Ease score, the Gunning-Fog index and the SMOG index. This should also apply to the Coleman-Liau Index and the Automated Reliability Index both of which rely on factors based on the number of characters per word.

There is thus the need to research alternative ways of measuring the understandability of the text of requirements. This research would be based on the assumption that readability and understandability are different metrics, they may be correlated, but we have to use different measurement algorithms for each. Avenues for further research include testing readability indices designed for languages other than English such as the Fernandez-Huerta index for Spanish, and the Douma index for Dutch. However, at this time, we feel that the most promising avenue to explore is based on modifying the Halstead Metrics.

Modifying the Halstead Metrics

From the perspective of this research, the text of requirements can be treated and hence measured in a similar manner to computer software source code which is written in a semi-formal “descriptive” language. Based on these assumptions, we should be able to modify and apply metrics for software source code developed (with some modifications) to estimate the understandability index of requirements. The research question is “can we apply a derivative of the Halstead Metrics to the measurement of the understandability of requirements?” A preliminary approach is outlined in Table 1

Table 1 Suggestions for modifications to the Halstead Metrics for the measurement of the understandability of the text of requirements

Halstead Notation	Programming language equivalent	Text of requirement
n_1 – unique operators	Unique variables	nouns or just keywords from the domain?
n_2 – unique operations	Unique operations (operators) , like +, - and keywords, like if, case …	all verbs or just specific ones, such as "shall"
$n = n_1 + n_2$ Vocabulary	Sum of unique operands and operations	To be determined.
N1 – total number of operands	All variables in the code	All keywords or all words?
N2 – total number of operations	All operations in the code	All verbs or specific ones?
N=N1+N2	Program length	Length of the requirement text
V= N*log(n)	Program volume (or information in the code)	V=N * log(n) – Does this correlate with understandability?

The research approach

The planned frame work for the research is as follows

1. Develop a series of standardized reference requirement documents of various qualities.
2. Incorporate measurements of the various readability indices into future versions of Tiger Pro.
3. Analyze the data from the workshop questionnaires and correlate the various readability scores for the requirements in the questionnaire with the student data.
4. Run tests on the series of standardized reference requirement documents to determine which of the metrics provide the best measurement of the understandability of the requirement text.
5. Modify the Halstead Metrics to determine if a good measurement can be made.
6. Incorporate the modified Halstead Metrics into future versions of Tiger Pro.

Summary

This paper summarized the concept behind a prototype educational and research software tool that has helped postgraduate students to understand the difficulty in writing good requirements, reported on some results in the classroom, discussed the difference between readability and understandability and proposed a research program to develop a metric for the understandability of requirements.

References

Carson, R. S., "*Keeping the Focus During Requirements Analysis",* the 11th International Symposium of the International Council on Systems Engineering, Melbourne, Australia, 2001.

Hooks, I., "*Writing Good Requirements",* Proceedings of the 3rd NCOSE International Symposium, 1993.

Jacobs, S., "*Introducing Measurable Quality Requirements: A Case Study",* the IEEE International Symposium on Requirements Engineering, Limerick, Ireland, 1999.

Kar, P. and Bailey, M., "*Characteristics of good requirements",* 6th International Symposium of the International Council on Systems Engineering (INCOSE), Boston, MA, 1996.

Kasser, J. E., "*A Prototype Tool for Improving the Wording of Requirements",* Proceedings of the 12th International Symposium of the INCOSE, Las Vegas, NV, 2002.

Kasser, J. E., "*The First Requirements Elucidator Demonstration (FRED) Tool*", Systems Engineering: The Journal of the International Council on Systems Engineering, Volume, **7** (2004), no. 3.

Kasser, J. E. and Schermerhorn, R., "*Determining Metrics for Systems Engineering",* The 4th Annual International Symposium of the NCOSE, San Jose, CA, 1994.

Kasser, J. E., Tran, X.-L. and Matisons, S., "*Prototype Educational Tools for Systems and Software (PETS) Engineering",* Proceedings of the AAEE Conference, 2003.

Schneider, R. E. and Buede, D. M., "*Properties of a High Quality Informal Requirements Document",* the 10th International Symposium of the INCOSE, Minneapoilis, MN, 2000.

Tran, X.-L. and Kasser, J. E., "*Improving the recognition and correction of poorly written requirements",* the Systems Engineering Test and Evaluation (SETE) Conference, Brisbane, Australia, 2005.

Wilson, W. M., Rosenberg, L. H. and Hyatt, L., "*Automated Analysis of Requirements Specifications",* the IEEE International Conference on Software Engineering, Boston, MA, 1997.