Protocol for an Empirical Study of Structured Abstracts

Barbara Kitchenham, Pearl Brereton, David Budgen, Mark Turner, Stuart Charters and Stephen Linkman

Abstract

Background. We have been undertaking a series of systematic literature reviews as part of a research program aimed at assessing evidence-based software engineering. Medical standards for systematic reviews suggest that initial searches of candidate primary studies can rely on the title and the abstract to determine the eligibility of primary studies. Our experience indicates that software engineering abstracts are of such poor quality that it is often impossible to assess the eligibility of a primary study without reading some parts of the study itself. Medicine and psychology recommend the use of structure abstracts to improve the quality of abstracts. There have been several experimental studies of structured abstracts that have confirmed their value in psychology.

Objective. This protocol defines a research plan aimed at assessing whether structured abstracts exhibit better structural properties than conventional abstracts for software engineering articles.

Method. Conventional abstracts from two published proceedings of the EASE conference will be rewritten as structured abstracts. Metrics such as length, average sentence length, the Flesch readability index and amount and nature of added information will be obtained from different versions of the abstracts. The metrics for conventional and structured abstracts will be compared.

Resource requirements. The experiment will require a research supervisor and two student researchers for a period of 8 weeks, and, in addition, support from other experienced research staff from the EBSE project.

Background

A requirement for Evidence-based Software Engineering is the ability to find, evaluate and aggregate sources of evidence. The Evidence-based paradigm relies heavily on Systematic Literature Reviews to assemble evidence required to address a research question. Systematic Literature reviews require exhaustive searches of the literature to identify potentially relevant primary studies. Searches involve two stages. Firstly researchers need to perform a wide search to identify as many primary studies as possible. Secondly research must undertake a more detailed review of the candidate primary studies against specific inclusion and exclusion criteria. The first step of the search process is expected to identify a large number of studies of which many are actually irrelevant. Current standards suggest a review of the title and abstract should be sufficient to decide whether or not a paper is relevant to the question of interest [2]. However, recent attempts to use systematic literature reviews to address specific software engineering questions have reported difficulties identifying whether or not primary studies are relevant because information contained in abstracts is incomplete [1].
One approach to improving the standard of abstracts is to adopt the use of structured abstracts. This approach is strongly advocated in medical research and other domains such as psychology that are adopting the evidence-based approach to abstracts pioneered in medicine [3]. Experimental evidence suggests that structured abstracts are a potentially valuable approach to improving the readability and value of traditional abstracts [4]. We believe that it would be useful to obtain some empirical evidence to evaluate the potential impact of structured abstracts in software engineering. This protocol describes an empirical study aimed at assessing the quality of abstracts in software engineering based on submissions to the EASE Conference.

Empirical study context
The Evaluation and Assessment in Software Engineering (EASE) conference is a forum for empirical software engineering research. EASE06 actively encouraged authors to use structured abstracts in its call for papers. We received 29 submissions to the EASE06 Conference of which 9 included a structured abstract and 20 provided a conventional abstract. Of those papers, 11 were accepted unconditionally of which 4 had structured abstracts and 2 were conditionally accepted of which 1 had a structured abstract (see Table 1).

Table 1 Outcome of EASE submission review process

<table>
<thead>
<tr>
<th>Structured Abstract</th>
<th>Accept</th>
<th>Conditional accept</th>
<th>Poster</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>No</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>11</td>
<td>29</td>
</tr>
</tbody>
</table>

Fishers exact test does not indicate any significant difference in the acceptance rate for papers with or without structured abstracts but there is insufficient data to be confident of that result. However, it does seem clear that using a structured abstract did not reduce the probability of acceptance for the EASE Conference.

Although we can (and will) compare the readability, length and completeness of the structured and unstructured abstracts of the EASE06 accepted papers, there are too few data points to provide any very convincing results. We therefore propose a study based on the method used by Hartley to assess clarity of psychology abstracts [3]. The aim of the study will be to determine whether or not structured software engineering abstracts are more readable than unstructured abstracts with respect to selected readability metrics.

Proposed Study
The study will be organised as an 8 week summer project with a supervisor and two second year students working as research assistants. Members of the Evidence-Based Software Engineering (EBSE) project will be available to support the project (i.e. Professor Barbara Kitchenham, Professor Pearl Brereton and Mr Steve Linkman).

We propose the following study procedure:
- Select all unstructured abstracts from published papers in proceedings of EASE04 and EASE06 (both of which are formally published proceedings with
an ISBN). EASE04 accepted 17 papers of which one used a structured abstract. This means we would have a total of 23 papers with unstructured abstracts.

- Obtain (or develop, if required) a software tool to extract basic readability metrics (Length, average sentence length, use of passive, Flesch reading Ease score). This tool will be tested on the EASE06 structured abstracts.
- Employ two students (on 8 week summer bursaries) to re-write each of the abstracts in an agreed structured format. The students will require training for this task, based on [5]. The re-writing exercise is defined in the Appendix 1. With the exception of 3 papers chosen at random for an additional training exercise (see Appendix 2), the papers will be allocated at random to each of the students for the re-writing exercise.
- Send the structured abstracts to the authors of the papers and ask them to verify that they are an accurate summary of the paper. (In the event of an inability to contact or obtain response from any of the authors within the project timescales, Kitchenham, Brereton or Linkman will assess the accuracy of the abstract).
- Identify the nature of the information added to each structured abstract compared with the original unstructured abstract.
- Compare the readability metrics obtained for the structured abstracts with those obtained from the original abstracts.

Note. The results of the research will be converted into a research paper after the completion of the 8-week summer project.

The use of students to construct the abstracts addresses a limitation of Hartley’s original study due to the fact that he himself constructed the structured abstracts. It could be argued that Hartley’s expertise means that the structured abstracts were better than those another researcher might have produced. This would have biased the results in favour of detecting a difference between the structured and unstructured abstracts.

The construction of the pairs of structured and non-structured abstracts will provide the input to a follow-on experiment to assess reader-based measures including an information checklist score and clarity ratings.

**Data Collation and Analysis**

A preliminary exercise will collate information about the abstract provided by each of the 13 papers accepted for EASE06. Basic data will be collated as shown in Table 2. The only analysis performed on this data will be to present the average values for the 5 papers that provided structured abstracts and the 7 papers that provided conventional abstracts.

**Table 2 Data collation form for EASE06 papers**

<table>
<thead>
<tr>
<th>Study Id</th>
<th>Structured abstract</th>
<th>Length</th>
<th>Average sentence length</th>
<th>Flesch Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes/No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The main results of the study will be a comparison of a structured and unstructured version of papers accepted at EASE04 and EASE06 that provided conventional abstracts. For each abstract, the students will collect readability metrics from the unstructured version of the abstract and from a structured version of the abstract. The data will be collated as shown in Table 3. The data will be used to test the null hypothesis that there is no difference in the readability metrics between conventional and structured abstracts with the alternative hypotheses that structured abstracts are longer but more readable than conventional abstracts.

The data will be analysed using a t test if the data are approximately Normally distributed. The non-parametric Mann-Whiney test will be used if the data are non-Normal.

A regression analysis will be used to check whether factors such which student constructed the abstract or which EASE conference the abstract originated from influenced the result. I.e. we will apply the following regression model

\[ y = b_0 + b_1 \text{Student} + b_2 \text{Source} + b_3 \text{x}_1 \]

where \( y \) refers to a readability metric for the original abstract, \( \text{x}_1 \) refers to the corresponding readability metric for the structured abstract

Student and Source are dummy variables used to identify the student who revised the specific abstract and the EASE conference where the abstract was published

\( b_i \ (i=0,\ldots,3) \) are coefficients of the regression equation.

If \( b_1 \) and \( b_2 \) are not significantly different from zero, we can assume that student and source do not affect the outcome of the experiment.

**Table 3 Data collation form for abstract comparison experiment**

<table>
<thead>
<tr>
<th>Study Id</th>
<th>Source EASE04 or EASE06</th>
<th>Student id</th>
<th>Abstract version</th>
<th>Length</th>
<th>Average sentence length</th>
<th>Flesch Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Original</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Structured</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data will also be analysed to determine whether there are any trends in the type of information omitted from conventional abstracts. This will be done by assessing the information added to construct structured abstract in terms of additional words required in the following categories: background, objective, method, results, and conclusions. The data will be tabulated as shown in Table 4. The students must record the number of words added in each category. The data will be analysed by comparing the null hypothesis that the new information is distributed equally among all categories (i.e. the average percentage of new words in each category is approximately 20%) with an alternative hypothesis that the new information is not distributed equally amongst all categories. It will be important to determine whether conventional abstracts are likely to omit information that is particularly important when assessing eligibility for systematic reviews such as results and conclusions than other types of information.
Table 4 Data collation form for missing data trend analysis

<table>
<thead>
<tr>
<th>Study id</th>
<th>Total New words</th>
<th>New Words per section</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Background</td>
</tr>
</tbody>
</table>

**Limitations of the study**

Using EASE papers restricts our assessment to empirical software engineering papers. However, if structured abstracts are unable to demonstrate any benefit for empirical software engineering papers it is unlikely that they would be useful for other types of paper. If the results of this study are favourable, future studies should assess whether other forms of software engineering paper could also benefit from structured abstracts.

The use of students to construct the abstracts poses a risk that the structured abstracts would not be of high a quality as those produced by experienced researchers. This might bias the study against finding a difference between structured and unstructured results. It will therefore be important that the students are properly trained and receive feedback on their initial attempts at constructing abstracts.

**Project Timetable**

Project Start Date: 3 July 2006  
Project Duration: 8 weeks  
Project End Date: 25th August 2006.

<table>
<thead>
<tr>
<th>Project Task</th>
<th>Start Date</th>
<th>End Date</th>
<th>Task description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary Training</td>
<td>3 July</td>
<td>14 July</td>
<td>The student research associates will be introduced to the project and given training in writing structured abstracts.</td>
</tr>
<tr>
<td>Tool construction or familiarisation</td>
<td>3 July</td>
<td>14 July</td>
<td>The student research associates will either construct or learn how to use a tool for extracting readability metrics.</td>
</tr>
<tr>
<td>Experimental process training</td>
<td>17 July</td>
<td>19 July</td>
<td>The student research associates will undertake training in re-writing and the experimental procedures outlined in Appendix a1 and Appendix 2.</td>
</tr>
<tr>
<td>Re-writing abstracts &amp; Data extraction</td>
<td>20 July</td>
<td>11 August</td>
<td>The student research associates will perform the re-writing exercise, extract the data and contact the researchers who wrote the original abstracts</td>
</tr>
<tr>
<td>Final correction and</td>
<td>14 August</td>
<td>18 August</td>
<td>The student research associates will perform the correction and final extraction.</td>
</tr>
</tbody>
</table>

Version 3, 1 March 2006  5
| data extraction | associates will make any changes to the abstracts requested by the original authors and re-run the data extraction tools |
| Collation of experimental results | 21 August | 25 August | The student research associates will collate and present the numerical results of the experiment. Note. They will not be expected to perform any statistical analysis on the results. This will be undertaken by the supervisor subsequent to the project. |
| Post project activities | 29 August | 22 September | The supervisor and EBSE staff will produce a research paper based on the project results. |

**References**


**Appendix 1. Re-writing Procedure**

The re-writing process will involve five stages:

1. The information in the original abstract will be re-organised to fit the structured abstract format.
2. The basic readability metrics will be extracted from this version of the abstract (version 1).
3. Version 1 will be supplemented by any information required by the structured abstract format but missing from the original abstract to obtain Version 2 of the abstract.
4. Any information required by the structured abstract format missing from Version 1 will be recorded.
5. The basic readability metrics will be extracted from Version 2.
Appendix 2. Re-writing Training

The students will require training in the rewriting process. This will be organised as follows:

- Three papers will be selected at random.
- For each paper:
  - Both students will complete stage 1 of the re-writing process (see Appendix 1).
  - They will meet with the supervisor and construct a joint Version 1 and run the metrics extraction tool on the jointly produced abstract.
  - They will then individually construct Version 2 of the abstract.
  - They will meet with the supervisor and construct a joint Version 2 of the abstract and together identify the missing information and run the metrics extraction tool.

If the supervisor is confident that the students understand the process and are producing abstracts of similar quality the training exercise will finish otherwise the process will be repeated.