EasyChair Preprint

№ 5997

Replications on the use of Graphical Abstracts to Support Study Selection in Systematic Mapping Studies

Alex Fernando Bonora, Katia Romero Felizardo,
Érica Ferreira de Souza and Nandamudi Lankalapalli Vijaykumar

July 3, 2021

Replications on the use of Graphical Abstracts to Support Study Selection in Systematic Mapping Studies

Alex Fernando Bonora[1], Katia Romero Felizardo[1], Érica Ferreira de Souza[1], and Nandamudi Lankalapalli Vijaykumar[2]

[1] Department of Computer, Federal University of Technology – Paraná (UTFPR), Cornólio Procópio/ Paraná, Brazil
[2] National Institute of Space Research – (INPE), São José dos Campos/ São Paulo, Brazil
Federal University of São Paulo, São José dos Campos/ São Paulo, Brazil
alexbonora@alunos.utfpr.edu.br; katiascannavino@utfpr.edu.br; ericasouza@utfpr.edu.br; vijay.nl@inpe.br

Abstract. Context: Little is known about the relevance of using graphical abstracts to select studies within Systematic Mapping Studies (SMS). Only one study (renamed as "original study – OS") was conducted to evaluate the conduction of selection review activity manually and using visual graphical abstracts. **Aim:** This research aims to build a body of empirical evidence through replications on the use of graphical abstracts to support the selection of primary studies in SMS. **Method:** A total of five replication studies were performed. Each replication was organized in two sessions (training – 9 studies to be classified as included or excluded – and execution – other 20 studies) and two groups (G1 – reading abstracts of the studies; G2 – analyzing graphical abstracts versions of the same studies used by G1). Performance (time) and effectiveness (number of studies correctly/incorrectly selected) were assessed. Qualitative analysis from the opinion of participants was also carried out. **Results:** Confirming the results of the OS, our results suggest that graphical abstracts speed up the selection activity and increase the performance of reviewers. Moreover, graphical abstracts are useful, and reviewers are completely satisfied by using them. **Conclusions:** Graphical abstract is an innovative and positive alternative for supporting study selection in SMSs.

Keywords: Graphical Abstract and Systematic Mapping Studies and Replication.

1 Introduction

The term replication has come into use to refer to a systematic repetition of an original study to double-check its results [2]. This definition implies that replication must be explicitly related to a previous study. Replications increase

the validity and reliability of the results yielded in an initial study [2]. Moreover, threats to the validity of a study can be addressed by the replication of this study.

Felizardo et al. [1] run one pilot study (renamed as "original study – OS") to evaluate the effects of using graphical abstracts to select studies during the conduction of Systematic Mapping Studies (SMSs) in Software Engineering (SE). The goals and the main contributions of the research presented in [1] were: (1) to compare performance (in terms of time taken); (2) effectiveness (in terms of correctness of the study inclusion/exclusion); and (3) to collect the opinion of participants on using conventional and graphical abstracts to select studies. This paper is an extension of [1].

For the purposes of the OS, Felizardo et al. [1] created a set of graphical abstracts. To assist the creation of graphical abstracts they proposed an approach using Concept Maps (CMs). In summary, Felizardo and the other collaborators [1] established a template based on CM and a set of guidelines to use this template to create graphical abstracts. The focus of Felizardo et al. [1] was not to carry out a study considering a large-sample. The study validated the template and the use of graphical abstracts based on CMs to support the selection activity of candidate studies. Therefore, one of the potential threats to the validity of the OS was related to the sample used (eight participants). It is often difficult to draw general conclusions from small-sample data. Therefore, the goal of this paper is to replicate the OS [1], involving a larger sample size of participants.

A summary of the OS is also presented herein, together with a comparison of its results and the results of the replications. The initial results of Felizardo et al. [1] have pointed out the value of graphical abstracts for selecting primary studies. Data from replications confirm that there are differences in performance and effectiveness for selecting studies manually or adopting graphical abstracts. An interesting finding is that participants widely pointed out that graphical abstracts are useful to support the selection activity, which is quite relevant information for researchers that intend to conduct SMSs.

The remainder of this paper is organized as follows: Section 2 describes related works on graphical abstracts. Section 3 reviews the study design, participants' task and metrics. Subsection 3.1 summarizes the results of replications. Conclusions are discussed in Section 4.

2 Related Works – Graphical Abstracts in Software Engineering

Graphical abstracts should concisely summarize the content of an article, through an image, to attract the reviewers' attention. Elsevier, which publishes the main Software Engineering (SE) journals, explains the purpose of a graphical abstract as "A graphical abstract should allow readers to quickly gain an understanding of the main take-home message of the paper and is intended to encourage

browsing, promote interdisciplinary scholarship, and help readers identify more quickly which papers are most relevant to their research interests."[3].

Currently, the existing Elsevier guidelines for the creation of graphical abstracts are limited to technical information, such as the size of the figure, the minimum number of pixels, appropriate screen resolution, file types (TIFF, EPS, PDF, or MS Office). In this context, Felizardo et al. [1] have proposed an approach to creating graphical abstracts using CMs in SE. They established a template based on CM and a set of guidelines to use this template to create graphical abstracts. CMs are diagrams containing meanings, i.e., diagrams that have meaningful relationships and conceptual hierarchy. Unlike the diagrams mentioned above, the purpose of CMs is not to classify concepts but relate and organize them hierarchically.

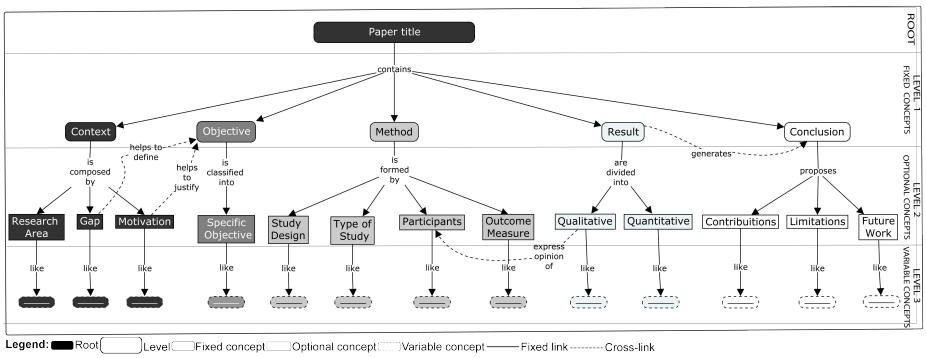

Fig. 1. CM template proposed by Felizardo et al. [1]

As shown in Figure 1, the CM template is a graphical and hierarchical view, containing parts of a scientific study (e.g., title, context, objective, method, result, and conclusion) as concepts and relationships between them. CM template contains the main concept (root) and other concepts related to this main concept are drawn. Concepts are also grouped into levels. The concept at the top is the most inclusive. More specific concepts are arranged hierarchically at the bottom. The lower the level is, the more details are described. The authors also proposed a set of guidelines to use their template to create graphical abstracts based on CMs. An exemplification of the use of the guidelines to create graphical abstracts and its evaluation are presented in Felizardo et al. [1].

Felizardo et al. [1] also evaluated the relevance of graphical abstracts based on CMs to support the selection of primary studies in SMSs. The current study also involves the use of graphical abstracts, however, its focus is to replicate

[3] https://www.elsevier.com/authors/journal-authors/graphical-abstract – Retrieved: 15/09/2020

Felizardo et al.'s study [1] to assess the relevance of using graphical abstracts for selecting primary studies in SE.

3 Replications: relevance of graphical abstracts for selecting primary studies

The Research Questions (RQ) of the original study [4](and that were also used in our replications) and their related hypotheses are:

- **RQ1:** *What type of abstract (textual or graphical) improve the performance of the study selection activity in SMS?*
  - ***H0:*** *Graphical abstracts has no effect on time taken to select studies.*
  - ***H1:*** *Graphical abstracts has a positive effect on time taken to select studies.*

- **RQ2:** *What type of abstract (textual or graphical) improve the effectiveness of the study selection activity in SMS?*
  - ***H0:*** *Graphical abstracts has no effect the number of studies correctly included/excluded.*
  - ***H1:*** *Graphical abstracts has a positive effect on number of studies correctly included/excluded.*

- **RQ3:** *What is the opinion of the participants in relation to the abstracts used?* – These questions will be carried out in a qualitative way.

The OS was organized in two sessions: training and execution. For training purposes, a small set of primary studies (Set 1 containing 9 primary studies, and their graphical abstracts) was used. To ensure that impressions and knowledge from the training would not interfere with the study, a different and larger set of primary studies (Set 2 with 20 studies and their graphical abstracts) was used for the execution stage. In summary, these sets of 29 graphical abstracts (Set1:9 + Set2: 20) were built for this study, based on SMS of Souza et al. [4] and randomly selected as Set 1 and Set 2.

Therefore, during the training session, the participants were given set 1 of studies to be analyzed (G1 – a list of abstracts; G2 – graphical abstracts), the inclusion and exclusion criteria, and a table to summarize the decision on whether to include or exclude a study. The design of the OS, previously described, was duplicated for replications without changes (2 groups, 2 sessions).

Qualitative analysis from the opinion of participants was carried out. The participants[5] in the five replications (85 in total) were not significantly differ-

[4] Laboratory package is available in:
https://www.dropbox.com/sh/qxea4866k48mt3l/AAAAKCvlt8gn-6hSRTzGnY0Ja?dl=0
https://www.dropbox.com/sh/lspjq3gtj3qc53p/AADLjS9Aq6Uz3QLleJfK_EfXa?dl=0

[5] The Informed Consent Form (ICF) used by us is available in
https://www.dropbox.com/preview/Public/Termo%20Consentimento.pdf?role=personal

ent from each other in terms of experience in conducting SMS (they had all conducted between 1–2 SMSs).

3.1 Results

This section reports the results of replications addressing our RQs. A summary of the results is shown in Table 1.

• ***RQ1: What type of abstract (textual or graphical) improve the performance of the study selection activity in SMS?***

Participants' performances were measured (see the second column of Table 1) to answer the first research question (RQ1). The time savings using graphical abstracts in replications was 7.40 minutes for each set of 20 studies. The answer to our RQ1 is that the use of graphical abstracts could speed up the selection activity.

Table 1. Summary of Results: Replications

Group	Time (min)	Cor. I	Cor. E	Cor. I and E	Inc. I	Inc. E	Inc. I and E
G1–ALL = 85	24.50	7.37	4.90	12.27	2.63	5.10	7.73
G2–ALL = 85	17.10	7.33	5.98	13.31	2.67	4.02	6.69
Legend: Cor. = Correct; Inc. = Incorrect; I = Inclusion; E = Exclusion							

We use the Wilcoxon-Mann-Whitney test to evaluate the group's values [3]. Wilcoxon-Mann-Whitney test[6] was used because results were not normally distributed. Moreover, we want to test whether one of the groups tends to have higher values (e.g. number of studies correctly included or excluded, etc) than the other group, or if they have the same median [7].

Regarding performance (RQ1), our results have shown that (see Table 2 – Performance) the use of graphical abstracts can improve the performance of the primary studies review activity ($p\text{-}value = 0.0001 < 0.05$). This means that the selection activity is faster using graphical abstracts.

• ***RQ2: What type of abstract (textual or graphical) improve the effectiveness of the study selection activity in SMS?***

Table 1 (see fifth column) shows averages of studies correctly included/excluded during the five replications. The answer to our RQ2 is that graphical abstracts could improve the effectiveness of selection study activity.

[6] For reproducibility purpose we adopted Action Stat, Version: 3.7, R Version: 3.3.2 as a statistical package.

[7] No statistical significance tests were used in OS due to the small sample employed (i.e., eight participants).

Table 2. Results for Man-Whitney test – Replications

Variable	*p-value*	Statistically Significant?
Performance	0.0001	Yes (p-$value < 0.05$)
Effectiveness – Correct Selection	0.0214	Yes (p-$value < 0.05$)
Effectiveness – Incorrect Selection	0	Yes (p-$value < 0.05$)

Regarding effectiveness (RQ2) (see Table 1), our results have shown that the use of graphical abstracts can improve the effectiveness of the primary studies review activity, i.e., increase the correct classifications and decrease the erroneous ones (p-$value$ – Correct Selection $= 0.0214 < 0.05$; p-$value$ – Incorrect Selection $= 0 < 0.05$).

• ***RQ3: What is the opinion of the participants in relation to the abstracts used?***

The replications also evaluated issues related to participants' opinions. A questionnaire with seven affirmative sentences was used. The sentences are detailed as follows:

- A1 – The abstracts are useful to select primary studies.
- A2 – The abstracts help me to select the primary studies.
- A3 – The abstracts make it easy for me to select the primary studies.
- A4 – The abstracts are easy to be read.
- A5 – I learned how to read the abstracts to select primary studies quickly.
- A6 – I like to read abstracts to select primary studies.
- A7 – I would recommend the reading of abstracts to another researcher conducting a systematic review.

The possible choices to these sentences vary from "strongly disagree" to "strongly agree", in a scale based on the Likert Scale method.

Table 3 summarizes participants' of replications 1–5 who agree or strongly agree with the seven affirmative sentences of our qualitative analysis.

3.2 Threats to validity

In conducting our replications, we identified some threats to validity, they are described as follows.

One of the potential threats to the internal validity of our replications is related to our assumption that the researchers who originally conducted the SMS made 100% correct decisions on the inclusion and exclusion of studies. However, SMSs conducted by different researchers on the same question sometimes led to different conclusions. Most of the classifications performed by the participants of replications match the classification conducted by the experts. For example, two studies were classified as included by the expert who conducted the SMS and all participants who participated in the replication. In the same way, the other three studies were classified as excluded by the expert and all participants.

Table 3. Summary of Results of Qualitative Questions

Sentence	*Manual Reading*	*Graphical Abstracts*
The abstracts are useful to select primary studies.	94%	92%
The abstracts help me to select the primary studies.	95%	96%
The abstracts make it easy for me to select the primary studies.	70%	91%
The abstracts are easy to be read.	48%	86%
I learned how to read the abstracts to select primary studies quickly.	48%	82%
I like to read abstracts to select primary studies.	52%	94%
I would recommend the reading of abstracts to another researcher conducting a systematic review.	78%	88%

One other study was originally classified as included by the expert and only one participant did not classify it the same way. Additionally, more than 90% of the participants agreed on the classification of the other five studies. Even so, we cannot affirm that all expert's opinions are correct.

Another threat to internal validity is related to the number of studies or graphical abstracts to be analyzed. A large number of studies could affect the motivation of the participants. We do not consider this threat serious since we limited the number of studies used as set 2 (20 studies).

One threat to construct validity was the participants' level of experience in KM/Software Testing. It is possible that the results could have changed if the participants of the replications had more research experience in the topic of the analyzed primary studies. Another threat was that the native language of the participants was different than the language of the study materials. The participants spoke Portuguese as their native language and the abstracts were written in English. The level of experience in English could affect the capability for selecting studies, especially for reading abstracts using the traditional approach. In graphical abstracts, there is little text to be read, only the concepts and links. Therefore, a new replication, with English-native speakers could be conducted to explore it.

It is also important to mention the replication of this study in a Systematic Literature Review (SLR). One of the main differences between SLR and SMS is related to the selection activity of the studies. The scope of an SMS is broader, and the selection criteria of studies are more general than in an SLR, where studies are selected/analyzed in a greater depth. It is difficult to draw general conclusions from our results, hence we cannot affirm that graphical abstracts also speed up the selection activity and increase the effectiveness of reviewers in SLRs. A new study is required to validate this assumption.

4 Conclusions

The main contribution of this research is five replications of a study to compare Ph.D. and Master students' performance, effectiveness, and participants' opinions in reviewing primary studies manually and using graphical abstracts. The results show that the answer to RQ1 is "The performance of the participants that used graphical abstracts is higher than that of the participants that used the manual method. The answer to RQ2 is "The effectiveness of the participants that used graphical abstracts is higher than that of the participants that used the manual method." Finally, the answer from RQ3 is that "The participants who use the graphical abstracts found the abstracts useful and they are completely satisfied by using them."

We concluded that graphical abstract is an innovative and positive alternative for supporting study selection in SMSs. As future work we intend to: (1) conduct a new replication with English-native speakers; (2) assess the use of graphical abstracts in SLR context.

References

1. Felizardo, K.R., Souza, E.F., Hesae, S., Vijaykumar, N.L., Nakagawa, E.Y.: Analysing the use of graphical abstracts to support study selection in secondary studies. In: Experimental Software Engineering (ESELAW Track). pp. 1–10 (2017)
2. Juristo, N., Gomez, O.: Replication of software engineering experiments. In: LASER Summer School. Lecture Notes in Computer Science, vol. 7007, pp. 60–88. Springer (2010)
3. Neuhäuser, M.: Wilcoxon–Mann–Whitney Test, pp. 1656–1658. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
4. Souza, E.F., Falbo, R.A., L., V.N.: Knowledge management initiatives in software testing: A mapping study. Information and Software Technology **57**(1), 378–391 (2014)