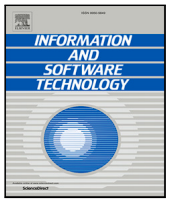
Information and Software Technology 133 (2021) 106514

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof



Case Study Research in Software Engineering—It is a Case, and it is a Study, but is it a Case Study?

Claes Wohlin

Blekinge Institute of Technology, Karlskrona, Sweden

ARTICLE INFO

Keywords:
Case study
Empirical
Misuse
Software engineering

ABSTRACT

Background: Case studies are regularly published in the software engineering literature, and guidelines for conducting case studies are available. Based on a perception that the label "case study" is assigned to studies that are not case studies, an investigation has been conducted.
Objective: The aim was to investigate whether or not the label "case study" is correctly used in software engineering research.
Method: To address the objective, 100 recent articles found through Scopus when searching for case studies in software engineering have been investigated and classified.
Results: Unfortunately, the perception of misuse of the label "case study" is correct. Close to 50% of the articles investigated were judged as not being case studies according to the definition of a case study.
Conclusions: We either need to ensure correct use of the label "case study", or we need another label for its definition. Given that "case study" is a well-established label, it is probably impossible to change the label. Thus, we introduce an alternative definition of case study emphasising its real-life context, and urge researchers to carefully follow the definition of different research methods when presenting their research.

1. Introduction

Case studies are common in software engineering, and guidelines have been provided, for example, by Runeson et al. [1]. They based their definition of case study on definitions from other areas including the definitions by Yin [2], Benbasat et al. [3] and Robson [4]. Runeson et al. [1] define a case study as follows within software engineering – "Case study in software engineering is an empirical enquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary software engineering phenomenon within its real-life context, especially when the boundary between phenomenon and context cannot be clearly specified".

The case study definition provided by Runeson et al. includes five essential aspects, which are common to several other definitions:

- *Empirical enquiry or investigation* is also covered by Robson [4] and Yin [2]. The empirical aspect is not covered explicitly in the definition by Benbasat et al. [3].
- *Contemporary phenomenon* is covered by both Robson [4] and Yin [2]. Benbasat et al. [3] do not include contemporary in the definition, although using the term phenomenon.
- Within its *real-life context* is included by Yin [2]. Robson [4] does not have real-life in his definition, while Benbasat et al. [3] define it as within its natural setting.
- Using *multiple sources* is covered by both Robson [4] and Benbasat et al. [3]. Yin [2] does not refer to multiple sources.
- *Boundary between phenomenon and context is unclear*, which is included by both Yin [2] and Benbasat et al. [3]. Robson [4] does not include the boundary in his definition.

The two aspects most clearly distinguishing a case study from other types of research studies are "contemporary phenomenon" and "real-life context". Many other methods include collecting empirical data, and use multiple sources of information, while it is more challenging to determine what it implies when the boundary between phenomenon and context is unclear. An aspect not covered by the definitions is the role of the researcher in relation to the phenomenon studied. In case study research, researchers work with participants in the study, while in action research, researchers are involved in the study.

The focus here is on evaluating the use of the label "case study" in relation to primarily contemporary phenomena and real-life context. A contemporary phenomenon is here defined as occurring at present, i.e. not primarily using purely historical information, for example, from data or open source repositories. Studies of only historical information, for example data or software, in any form does not meet the definitions of a case study. Moreover, real-life context means that the studies

E-mail address: claes.wohlin@bth.se.

https://doi.org/10.1016/j.infsof.2021.106514
Received 2 November 2020; Received in revised form 5 January 2021; Accepted 11 January 2021
Available online 18 January 2021

Table 1
Classification of 100 articles claimed to be case studies.

Case study	No real-life context	No real-life context and no contemporary phenomenon	Action research
53	33	13	1

focus on events existing or occurring in reality, for example, in an industrial context. The data is drawn from or drawing on actual events or situations, and it, in most cases, include people.

Recently, we experienced that a reviewer of an article asked us to rework our analysis as a case study in the Yin sense. Unfortunately, the study was not a case study. We presented a case, and it was a study, but it was not a case study. Our study presented an analysis of a research article, and hence it does not meet the aspect of a contemporary phenomenon, since it is an analysis of a published article. Furthermore, the study is not conducted in a real-life context. Thus, it is not a case study. The comment was aligned with our perception that the research method "case study" is all too often misused for studies that do not conform to the definitions of a case study. The comment of the reviewer gave rise to the following research question: How misused is the use of the label "case study" in software engineering research?

The remainder of the article is structured as follows. Section 2 presents related work concerning critique towards case study research. In Section 3, the research approach for studying the use of the label "case study" in research articles is detailed. The results from the analysis of the articles are presented in Section 4. Validity threats to the study are presented in Section 5. Finally, the conclusions are provided in Section 6.

2. Related work

Guidelines for case study research have been published in a software engineering context, for example by Verner et al. [5] and Runeson et al. [1], as well as other more general sources as above-mentioned with their definitions of what constitutes a case study.

A search was conducted in Google Scholar and the reference lists of Verner et al. [5] and Runeson et al. [1] were examined to identify articles being critical towards case studies. Case study research has been criticised, although primarily from the point of view of the conduct of case studies. The criticism concerns selection and interpretation bias, and that they are hard to generalise as pointed out by, for example, Easterbrook et al. [6] and Runeson et al. [1]. Flyvbjerg [7] identifies five misunderstandings concerning case studies, although none of them relates to the problem of how the term "case study" is used when describing research studies.

No criticism concerning the use of the label "case study" was identified. However, it may be mentioned in passing in an article, but to the best of our knowledge, the research question has not been the focus of a research article. The objective here is to fill this research gap.

3. Method

The objective is to investigate a sample of 100 recently published articles concerning the use of the label "case study" in software engineering. The following procedure was applied.

A search was conducted in Scopus using the following search string: "case study" AND ("software engineering" OR "software development"). Furthermore, the search was limited to the publication type "article", and to seven software engineering journals. The selection of journals is based on the journals used historically in the ranking of institution and scholars, as described by Wong et al. [8]. We chose to limit the time interval from 2011 to present to assess the current status, assuming that it would result in at least 100 articles. Moreover, we excluded (1) methodology articles and (2) articles referring to a case study presented in another article. Within the articles found in the search in Scopus, we investigated the use of the label "case study", and decided whether or not the article should be included. All included articles are classified concerning how the label "case study" was used. The following classification was used:

- Case study that exhibits the five essential aspects listed in Section 1.
- A case, but not in a real-world context.
- A case, but not a contemporary phenomenon.
- A case, but it ought to be classified as action research, due to the role of the researcher.

4. Results

The search in Scopus was conducted on October 25, 2020, and it resulted in identifying 368 articles using the method described in Section 3. The number of articles identified per year in the seven journals does not vary much over the years. The lowest number of articles is 19 in 2011, and the highest is 46 in 2013 and 2020. Five articles are listed as published in 2021. The articles are sorted from newest to oldest since we want to investigate the current status.

Based on the sampling method described in Section 3, the analysis involved having to assess 106 articles since six articles were excluded. The six articles were excluded since they were either methodology articles or literature reviews, although referring to "case study" in the article. The 100 articles included were published from May 2018 to the five articles listed as published in 2021, i.e. we assessed the 100 most recent articles published as case studies and available in Scopus.

The classification showed that if a study was done in a real-life context, then it also studied a contemporary phenomenon. Real-life context and contemporary phenomenon most often go hand in hand; for example, an evaluation of a proposal (e.g. a new process, approach, method, technique or tool) is performed on data or software in repositories. The context is not real-life, and it is not a contemporary phenomenon, since the evaluation is solely based on information in repositories, i.e. historical information. In some cases, the researchers developed or constructed an artefact to be able to evaluate their research, which may be interpreted as a contemporary phenomenon, although not studied in a real-life context.

The outcome of the classification is shown in Table 1. It turns out that only 53 studies were classified as case studies while 47 studies did not conform to the essential aspects in the definitions for being case studies. Thus, the investigation of 100 recent articles illustrates that the label "case study" is often used without meeting the definitions.

Typical examples claiming to be case studies, but not meeting the definitions are: (1) Evaluations or illustrations on a limited scale, for example, in a laboratory. These could be labelled evaluations or illustrations given their primary objective, and (2) Studies of existing information in open source repositories, defect databases of other database sources. These ought to be labelled as archival analysis.

In a few cases, we identified that the authors only used one data collection method, for example, only interviews, but we still accepted them as case studies. Finally, if the investigator participates in the case studied, then it ought to be labelled as action research (as identified in one case in Table 1).

5. Validity threats

The main validity threats to the investigation are selection and researcher bias. The selection threat was addressed by searching in seven journals. The quality of the articles published in the seven journals

ought to be better than a random publication in the field. Thus, the percentage of articles not being case studies ought to be lower than in many other venues. It implies that the result ought to be on the conservative side. The researcher bias was addressed by analysing the 100 articles in less than a week to ensure consistency.

All investigated articles are listed in Appendix A. The listing includes both a motivation for including or excluding them and their classification as stated above. Thus, the investigation may be replicated. Overall, the results are clear, and the threats are not likely to affect the conclusions, although some articles may be judged differently by someone else.

6. Conclusions

Given that 47% percent of the investigated articles do not qualify as case studies, we have a problem with the label and its definitions. Thus, we conclude that neither the label nor its definitions are suitable for their purpose, or the definitions are not used as they should. However, given that the label "case study" is very well-established, it is unlikely that it can be changed. Thus, we have two suggestions concerning the use of the term "case study" in software engineering.

First, we suggest that "case study" is better defined as follows:

A case study is an empirical investigation of a case, using multiple data collection methods, to study a contemporary phenomenon in its real-life context, and with the investigator(s) not taking an active role in the case investigated.

The objective with the definition is to make it clear that evaluations or illustrations, and analysis of information or artefacts in repositories are not case studies. On the one hand, more appropriate classifications exist for studies not following the formal definition of "case study". On the other hand, the definition of "case study" does not incorporate all different types of studies possible to conduct in a real-life context.

We have also emphasised multiple *data collection methods* to clarify that triangulation is essential. It is essential since it is close to impossible to control the independent variable in a real-life context. To interview more than one person does not make it a case study. In several existing definitions of case study, the concept of "multiple sources" is unclear, and may be interpreted differently by different researchers. Furthermore, by introducing the role of the investigator in the definition, we would like to make the boundary with action research clearer.

Second and finally, we want to urge researchers, reviewers and editors to carefully assess the use of different labels for research methods. In particular, researchers should provide a reference to the definition used. Hopefully, this will lead to more studies being more correctly classified.

CRediT authorship contribution statement

Claes Wohlin: Conceptualization, Methodology, Validation, Investigation, Writing - original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I would like to express my sincere thanks to Prof. Austen Rainer for his constructive feedback. Furthermore, I would like to express my gratitude to the reviewers and the editor for valuable comments that improved the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.infsof.2021.106514.

References

[1] P. Runeson, M. Host, A. Rainer, B. Regnell, Case Study Research in Software Engineering: Guidelines and Examples, Wiley Publishing, 2012.
[2] R.K. Yin, Case Study Research and Applications: Design and Methods, second ed., Sage Publications, 1994.
[3] I. Benbasat, D.K. Goldstein, M. Mead, The case research strategy in studies of information systems, MIS Q. 11 (3) (1987) 369–386.
[4] C. Robson, Real World Research : A Resource for Social Scientists and Practitioner-researchers, Wiley Publishing, 2002.
[5] J.M. Verner, J. Sampson, V. Tosic, N.A.A. Bakar, B.A. Kitchenham, Guidelines for industrially-based multiple case studies in software engineering, in: Proceedings Third International Conference on Research Challenges in Information Science, IEEE, 2009, pp. 313–324.
[6] S. Easterbrook, J. Singer, M.-A. Storey, D. Damian, Selecting empirical methods for software engineering research, in: F. Shull, J. Singer, D. Sjøberg (Eds.), Guide to Advanced Empirical Software Engineering, Springer, London, UK, 2008, pp. 285–311, (Chapter 11).
[7] B. Flyvbjerg, Five misunderstandings about case-study research, Qual. Inq. 12 (2) (2006) 219–245.
[8] W.E. Wong, T. Tse, R.L. Glass, V.R. Basili, T. Chen, An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008), J. Syst. Softw. 84 (1) (2011) 162–168.