

Software Engineering Researchers' Attitudes on Case Studies and Experiments: an Exploratory Survey

Dan Tofan, Matthias Galster, Paris Avgeriou
Department of Mathematics and Computing Science
University of Groningen
Groningen, The Netherlands
d.c.tofan@rug.nl, m.r.galster@rug.nl, paris@cs.rug.nl

Danny Weyns
Department of Computer Science
Katholieke Universiteit
Leuven, Belgium
danny.weyns@cs.kuleuven.be

Abstract— Background: Case studies and experiments are research methods frequently applied in empirical software engineering. Experiments are well-understood and their value as an empirical method is recognized. On the other hand, there seem to be different opinions on what constitutes a case study, and about the value of case studies as a thorough research method.

Aim: We aim at exploring the attitudes of software engineering researchers on case studies and experiments. Furthermore, we investigate how the perceptions of researchers vary along their views on what constitutes a case study.

Method: We performed an exploratory survey involving 26 software engineering researchers. We collected data using a paper-based questionnaire.

Results: We found that participants slightly prefer experiments over case studies. Moreover, participants believe there is more useful literature on experiments, than on case studies. By analyzing two different views on the nature of case studies, we found differences in the perceived validity of case studies.

Conclusions: The survey provided insights into the perceptions of researchers on case studies and experiments. Moreover, the results help reconcile different views on case studies.

Keywords - empirical software engineering; case studies; experiments; survey

I. INTRODUCTION

Researchers who conduct empirical software engineering (SE) studies can choose from a broad range of potential research methods. Easterbrook et al. (2008) consider five main classes of research methods: experiments, case studies, surveys, ethnographies and action research. According to Runeson et al. (2008), existing work on empirical SE research methodologies focuses on experimental research. This means, experiments seem to be an accepted and well-defined research method. In contrast, case studies seem to range from well organized and thoroughly performed studies to small toy examples (Runeson et al., 2008). Moreover, Easterbrook et al. (2008) identified a lack of understanding in the SE literature about what constitutes a case study. It appears that case studies are often interpreted as a plain working example of a newly proposed method.

This interpretation contrasts with Yin's (2003) definition of the case study as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident". Additionally, Zannier et al. (2006) reviewed 29 years of ICSE proceedings and noticed that, despite the high popularity of case studies, the term "case study" had frequently been used improperly, e.g., in studies that do not include a real world case.

We believe that the difference in opinion on what constitutes a case study has negative consequences on the quality of reported case studies. Furthermore, researchers may doubt the merits and the credibility of the case study as a rigorous research method. Consequently, they may use other research methods, even in situations when case studies would be more appropriate, e.g., when investigating a phenomenon in a software company, with many unknown factors and variables, and the context as a significant influence factor on the study. In such a case, experiments would not be applicable. Therefore, we propose two research questions:

RQ1. How do empirical software engineering researchers perceive the differences between case studies and experiments? This allows us to contrast the perceptions of researchers about the misunderstood notion of case studies with the more accepted notion of experiments.

RQ2. How do perceptions of researchers vary along their views on the nature of case study? This allows us to evaluate the difference between researchers that consider case studies as rigorous methods and those that perceive case studies as working examples.

This paper contributes to a better understanding of the perceptions of empirical SE researchers on case studies and experiments. Such perceptions play an important role when planning or reviewing empirical studies. In Section II of this paper, we present our research method used to investigate RQ1 and RQ2. Section III presents the findings for answering RQ1 and Section IV for RQ2. Section V discusses validity threats. Section VI presents future work and conclusions.

II. RESEARCH METHOD

We used the survey research method. According to Wohlin et al. (2003), a survey helps understand the population from which the sample of respondents was drawn. There are three

types of surveys: descriptive, explanatory, and explorative. Descriptive surveys enable the formulation of assertions about a population. Explanatory surveys aim at explaining assertions. Explorative (or exploratory) surveys act as an initial study to a more in-depth investigation. As our investigation is at an early stage, we use an exploratory survey.

We followed the survey process proposed by Ciolkowski et al. (2003): “(1) Study definition – determining the goal of the survey; (2) Design – operationalizing the survey goals into a set of questions; (3) Implementation – operationalizing the design to make the survey executable; (4) Execution – the actual data collection and data processing; (5) Analysis – interpretation of the data; and (6) Packaging – reporting the survey results”. We detail the first five steps in the following subsections (step 6 is done in this paper). Please note that overall these steps were performed consecutively. However, we performed survey definition and design iteratively, i.e., the design shaped the definition and vice versa.

A. Survey Definition and Design

As discussed in the introduction, we noticed disagreements in the literature about how case studies are understood in SE research. However, we did not find studies that investigate the perceptions of SE researchers. As experiments seem to be well understood, we decided to conduct an exploratory study to understand the perceptions of empirical SE researchers on case studies and experiments.

Our target population consists of empirical SE researchers. Therefore, we applied purposive sampling for selecting survey participants that are both knowledgeable and interested in empirical SE research, in particular in case studies and experiments. Moreover, because two of the authors attended the LASER (2010) summer school on empirical SE, we used convenience sampling to recruit participants. Ideally, we should have compiled an international list of such persons, and randomly select subjects from it. We asked the LASER participants to act as our sample. From 40 distributed paper questionnaires, 26 were returned, resulting in a response rate of 65%. We believe the high response rate is an indicator for the participants’ strong interest in empirical SE. The favorable timing also played an important role, as the survey took place during a summer school, therefore not interfering with the participants’ busy daily schedules. We discuss the validity threats introduced by our sampling approach in Section V.

To operationalize the attitudes of the researchers, we defined eight dimensions. Two of the authors identified the initial dimensions and their questions through brainstorming. Then we had further discussions with other researchers to refine the survey. We used existing literature (Easterbrook et al. 2008, Runeson et al. 2008, Yin 2003) for additional improvements, like the scenarios in Q8 and Q9. Based on the dimensions, the subjects could express their attitudes. For simplicity, we do not differentiate between “attitudes” and “perceptions”.

TABLE I. DIMENSIONS, DESCRIPTIONS, AND DERIVED QUESTIONS

Dimension	Description	Questions
Nature of case study	Nature of case studies, as a research method	Q1. In academic papers, working examples like the use of a tool to prove some point of the paper is a valid example of a case study
Preference	Preferred method of the participant	Q2. Whenever possible, I prefer to use experiments instead of case studies in my research
Peer acceptance	Perceived acceptance of a research method by other researchers	Q3. If I use experiments to study some phenomenon then I have better chances of getting my next paper accepted, compared to using case studies Q4. In general, my peers recommend choosing experiments over case studies for investigating various complex phenomena in software engineering
Validity	Perceived validity of a research method	Q5. For me, case studies are a valid research method Q6. For me, experiments are a valid research method Q7. In my opinion, when investigating some phenomenon, the results of experiments are less biased than the results of case studies
Suitability	Situations when to decide on what method to use	Q8. A beginner researcher is interested in understanding the effectiveness of a new type of a file navigator, which might improve the productivity of software developers. I recommend him/her to prepare some experiments to investigate the problem, rather than case studies Q9. A beginner researcher is interested in understanding how developers in the industry use UML diagrams during software design. I would recommend him/her to prepare some case studies to investigate the problem, rather than experiments Q10. I believe that case studies are only appropriate for the exploratory part of an investigation
Related work	Availability of examples and references on the research method	Q11. I believe there are too few examples of good case studies, which I could use as example for my own research Q12. I believe there are too few examples of good experiments, which I could use as example for my own research Q13. There are good references in the literature that I could use for preparing an experiment. Q14. There are good references in the literature that I could use for preparing a case study
Future plans	Plan for using a research method in the future	Q15. I plan to perform an experiment in the next 12 months. Q16. I plan to perform a case study in the next 12 months.
Personal network	Methods used by peers in the participant’s personal network	Q17. In your opinion, in your personal network of researchers, what is the roughly estimated percentage of researchers who used case studies as a research method at least once? Q18. In your opinion, in your personal network of researchers, what is the roughly estimated percentage of researchers who used experiments as a research method at least once? Q19. In your opinion, in your personal network of researchers, what is the roughly estimated percentage of researchers who would favor experiments over case studies, as research methods?

Table I contains the dimensions, their descriptions and the concrete questions derived for them (as listed on the questionnaire). For each dimension, we derived between one and four questions. Moreover, we added questions on the profile and background of participants. Later, we discarded a few questions from the original questionnaire, due to their lack of relevance and unclear wording (e.g. “As research methods, comparing case studies to experiments is like comparing apples to oranges.”). In total, we used 28 questions, out of which 9 related to the profile of the subject, and 19 to their attitudes. The questions on the *personal network* dimension (Q17-19) required percentage values. Q1-16 used a Likert scale with the following values: 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree) and 5 (strongly agree). Additionally, we offered a value of 0 (not applicable/I don’t know).

B. Survey Implementation

As we aimed at obtaining data from a larger sample, conducting interviews was not feasible. Therefore, we decided to use a questionnaire for data collection (see above). Implementing a survey is rarely a sequential process. Thus, we needed three iterations until deciding on the final questionnaire. We piloted the intermediate versions with a few researchers, who gave us important feedback, i.e., suggestions on adding some questions.

C. Survey Execution

During one of the last sessions of the summer school, we asked the audience to participate in the survey. We distributed the printed questionnaires. Some persons filled it in and returned it the same day, while others returned it the next day. Two of the authors were available to answer questions from participants about the survey.

D. Survey Analysis

We use descriptive statistics to analyze our data. As statistical tools, we use Excel and SPSS. For each dimension, we use frequency analysis to study the distribution of the answers. Details on the data analysis are provided in Sections III and IV, where we address RQ1 and RQ2, respectively. The raw data is available online (Tofan, 2011).

III. EXPERIMENTS AND CASE STUDIES

A. Profile

Twenty-one (80%) of our subjects conducted their research in Europe, two (8%) in North America, and one (4%) in Asia. Two subjects did not provide any information about their profile. We conclude that the geographical diversity of the participants is low, making it a more representative sample for Europe. To understand the research experience of our sample, we asked subjects about their academic background. Two persons did not pursue PhD studies. 16 (61%) subjects were enrolled in a PhD program, and 7 (27%) hold a PhD degree.

Twenty-two (85%) of the participants conducted SE research in the previous year, and 23 (88%) were involved in SE research, when answering the questionnaire. The high numbers indicate that our sample consists of active researchers.

Moreover, we are interested in the publishing venues, targeted by the participants. We asked them to indicate the

three major venues they usually aim for. The most popular one is the International Conference on Software Engineering, with 15 researchers aiming for it. Four persons indicated the International Symposium on Software Testing and Analysis. Each of the other venues was indicated by three or less researchers. Additionally, we asked each person to list the most important keywords describing his or her research interest. The wide range of publication venues and research interests suggest that the sample of researchers is representative for the SE community.

Around half of the sample (14 persons) had published at least one article that used an empirical research method. Also, 12 researchers conducted at least one experiment, while 20 performed at least one case study. In each of these two groups, there were 7 (for experiments), respectively 15 (for case studies) junior researchers. These numbers suggest that junior researchers (no PhD degree or PhD in progress) are more likely to have used case studies, rather than experiments.

B. Attitudes

In the following, we discuss the results obtained for each dimension to answer RQ1. Figure 1 summarizes the answers of questions Q1-16, along with frequency distributions and basic descriptive statistics (mean and standard deviation).

1) *Nature of case study*: This dimension measures if the subject perceives the case study as a systematic research method, as described by Easterbrook et al. (2008), Runeson et al. (2008) and Yin (2003), or as something different. We obtained 8 respondents for the first category (let us call them the Yin group), and 18 for the rest (the non-Yin group). These numbers suggest the Yin group forms an important minority. As the majority holds a different perspective, the Yin view is not broadly accepted among study participants.

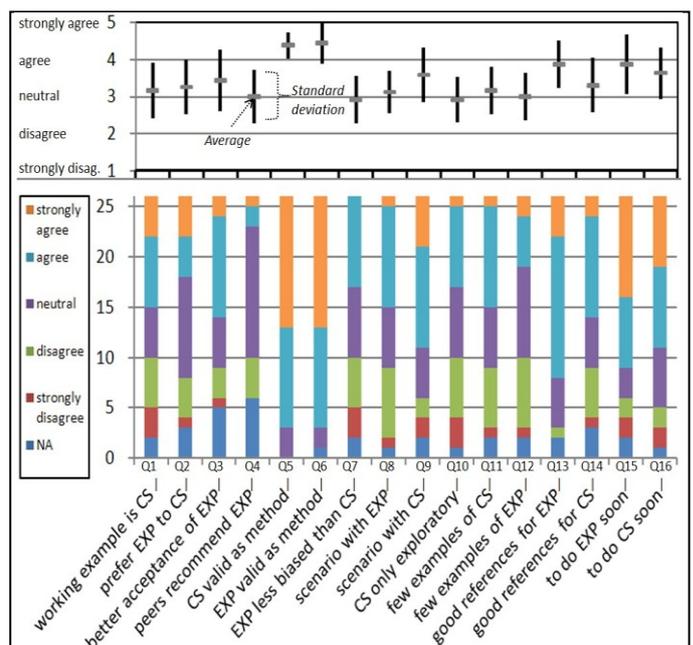


Figure 1. Upper chart: averages and standard deviations for questions 1 to 16. Lower chart: frequency distributions for each question.

IV. PERCEPTIONS ON CASE STUDIES

2) *Preference*: In Figure 1, for Q2 we see that most respondents were neutral. Still, the average has a small tendency to agreement. This suggests that researchers do not have a strong personal preference towards either case studies or experiments.

3) *Peer acceptance*: For Q3, the results suggest researchers believe that using experiments in papers slightly increases acceptance chances, compared to using case studies. For Q4, we have a strong neutral stance (13 subjects), suggesting that researchers do not perceive that their peers recommend one research method over the other.

4) *Validity*: As visible in Figure 1, at the answers for Q5 and Q6, the respondents highly agree on the validity of both case studies and experiments. This result may indicate that even if researchers have different understandings of a case study, they agree on their high validity. For Q7, the respondents are neutral, indicating their perception that other factors than the research method influence the bias of a study.

5) *Suitability*: The scenarios in Q8 and Q9 were discussed in detail by Easterbrook et al. (2008). Both case studies and experiments are suitable for the first scenario, while case studies particularly for the second one. The neutral result for Q8 confirms that both research methods are applicable, as described in Easterbrook et al. (2008). For Q9, we have a confirmation as well.

For Q10, the results are neutral. This suggests that some researchers believe case studies may be appropriate for more than the initial phase of an investigation. This would confirm Runeson et al. (2008) who mention the applicability of case studies not only for exploratory purposes, but also for descriptive or explanatory studies.

6) *Related work*: We notice that for Q11 and Q12, more researchers agree and strongly agree that there are too few examples of reported software engineering case studies, compared to experiments. The results suggest that researchers need more quality reports of case studies. The answers to Q13 and Q14 suggest that respondents believe that there is better literature for preparing experiments, than case studies.

7) *Future plans*: For Q15 and Q16, the results indicate that researchers plan to use both research methods in the future. As shown in Figure 1, slightly more persons indicate plans for experiments compared to case studies. Overall, the results suggest similar levels of popularity among the researchers.

8) *Personal network*: The questions for this dimension concern the opinion of the respondent on his or her personal network of researchers. For Q17 and Q18, the results are that an average of 64% of the participants used case studies at least once. On the other hand, an average of 43% used experiments at least once. These results suggest that case studies are more popular than experiments. Question Q19 resulted in an average of 32% of peer researchers who would favor experiments over case studies.

In the previous section, we discussed the differences in perception between case studies and experiments (RQ1). Next, we describe how such perceptions vary between the Yin and non-Yin groups, to answer RQ2. First, we discuss the interval data (Q17-19), then the ordinal data (Q1-16). Additionally, we explore the perceptions on case studies along the experience of the subjects. Finally, we offer a potential solution to unify the perspectives.

For the *personal network* dimension, we explain the higher percentage for case studies by the different perspectives on case studies. The Yin group gave an average of 49%, and the other group 73%. As the non-Yin group uses a more relaxed view of case studies, they report more peer researchers using them (i.e., they categorize studies as case studies that would not be classified as case studies by the Yin group). The Yin group obtained an average of 20%, and the non-Yin group 35%. We interpret this as follows. The non-Yin group regards case studies as a less rigorous research method, compared to the Yin group. As researchers prefer rigorous methods, they favor experiments over case studies.

In the previous section, we use the average of Likert scales for comparing the various perceptions. While useful in showing trends in the data, such approach has limitations [10], as the intervals between values cannot be presumed equal (e.g. the distance between neutral and agree is not necessarily the same as the one between agree and strongly agree). Therefore, we cannot use averages to compare the Yin and non-Yin groups, due to the ordinal nature of the Likert scales. A better approach is to use non-parametric tests, as we cannot assume that the distribution of our data is normal.

To compare the Yin and the non-Yin group, we use the Mann-Whitney U-test, available in the SPSS statistical tool. We compare the answers for Q1-16, to see if the two groups gave different answers. A low p -value, i.e., $p < 0.05$, indicates a statistically significant difference between the two compared groups. As expected, for Q1 the p -value is very low (0.001) as this question is used to separate the sample in the two groups. We obtained an important difference for Q5 ($p=0.02$), suggesting that the Yin group perceives case studies as more valid, compared to the non-Yin group. Some interesting but weaker differences exist for Q6 ($p=0.11$), i.e., on the perceived validity of experiments, and Q12 ($p=0.09$), i.e., the perceived availability of examples of experiments in the literature. We conclude that perceived validity is an important dimension for differentiating the two groups, while the other dimensions are less relevant.

Next, we explored how the perceptions of the respondents vary along their experience as researchers, as that may influence the perception on case studies. We divided the respondents in two groups: experienced (already having a PhD degree) and junior (the others). Using the Mann-Whitney U-test, we obtained differences for Q5 ($p=0.06$), and Q10 ($p=0.009$). We conclude that experienced researchers are more likely to perceive case studies as a valid research method (Q5), than junior researchers. Additionally, experienced researchers are more likely to regard case studies as valid for more than the

exploratory part of an investigation (Q10), compared to junior researchers.

We notice that the proposed dimensions provided some differentiation between the attitudes of researchers on case studies and experiments. The disagreement or confusion around the case study as a research method is mostly visible in its perceived validity. Furthermore, experienced researchers seem to agree with the high validity of case studies. To reduce the disagreements around case studies, we propose the idea of a *case study maturity model* with two levels. On the one hand, we can have low maturity case studies, as less systematic, but requiring low preparation efforts. Such case studies comprise working examples, or toy applications, useful mostly in the early, exploratory phases of an investigation. On the other hand, we can have high maturity case studies, as more systematic approaches, as described in the literature (Easterbrook et al. 2008, Runeson et al. 2008, Yin 2003). In our study, the Yin group uses the high maturity perspective on case studies, while the non-Yin group follows the perspective of low maturity. The main benefit of the proposed maturity model is clarifying the terminology around case studies, thus reducing disagreements.

V. VALIDITY THREATS

Our study is subject to validity threats (Wohlin et al., 2003). To reduce threats to its internal validity, we piloted the survey, and we were available for questions from the respondents. As we did not investigate causal relationships in our exploratory survey, internal validity concerns are low.

The survey's external validity is affected by the small size of our sample that included only 26 subjects. Moreover, all participants attended the summer school on empirical research methods (LASER, 2010), thus potentially influencing their perceptions. The sessions at the summer school covered experiments extensively, while case studies received little attention. We organized a 'Birds of a Feather' session to discuss informally about case studies, attended by five survey participants. Further studies, using a larger sample with a better geographical distribution, are needed to increase the external validity of our survey.

Construct validity issues focus on the correct interpretation and measurement of the perceptions. For our study, there might be other dimensions which we did not consider, and that might have proven relevant, e.g. necessary resources for performing either type of study. Also, some dimensions may have deserved more attention, i.e., 'suitability' addresses only partially the strong dependency between the research question and the appropriate method for answering it. Additionally, some questions might have been misunderstood by the participants, due to ambiguous phrasing. We mitigated these risks by not only piloting the questionnaire, but also removing unclear and irrelevant questions during the analysis phase. Moreover, measuring attitudes or perceptions suffers from bias risks from the researchers, who need to interpret the responses. We tried to mitigate this problem by involving more researchers in the analysis of the results. Although we assumed experimental research as accepted and well-defined, the literature survey of Sjøberg et al. (2005) found that the reporting of experiments is often vague, unsystematic, and lacking consistent terminology.

VI. CONCLUSIONS AND FUTURE WORK

The contribution of our exploratory survey is an insight into the perceptions of empirical SE researchers about case studies and experiments. For example, we obtained insights on the high popularity of the two research methods, and the need for more literature on case studies. Furthermore, we investigated two different views on case studies: the systematic understanding of the Yin group, and the relaxed view of the non-Yin group. We found that the two groups perceive differently the validity of case studies as a research method. We also noticed that experienced and junior researchers may also have different perceptions on case studies. To deal with the disagreements, we discuss a new approach to classify case studies, i.e., a case study maturity model.

As the proposed case study maturity model is at an early stage, future work is needed to refine it in more levels, including criteria for classifying case studies. These levels can then be calibrated using existing case studies and literature on the case study methodology (such as case study checklists). Also, more studies are needed to further understand the perception of researchers on case studies, with the ultimate goal of increasing the quality of reported case studies.

ACKNOWLEDGMENT

We thank the participants and organizers of the LASER 2010 summer school for their help in conducting this study.

REFERENCES

- Ciolkowski, M., Laitenberger, O., Vegas, S. & Biffi, S. (2003). Practical experiences in the design and conduct of surveys in empirical software engineering. *Empirical Methods and Studies in Software Engineering*. Springer, p. 104–128. Available from: <http://www.springerlink.com/index/2T8U4BUE0MXLMXV5.pdf>.
- Easterbrook, S., Singer, J., Storey, M.-A. & Damian, D. (2008). Selecting empirical methods for software engineering research. *Guide to advanced empirical software engineering*, p.285–311. Available from: <http://www.springerlink.com/index/n815725515063p2m.pdf>.
- LASER (2010). Summer School on Software Engineering. [Online]. Available from: <http://laser.inf.ethz.ch/2010/> [Accessed: January 2011].
- Runeson, P. & Höst, M. (2008). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*. 14(2) p.131-164. Available from: <http://www.springerlink.com/index/10.1007/s10664-008-9102-8>.
- Sjøberg, D.I.K., Hannay, J.E., Hansen, O., Kampenes, V.B., Karahasanovic, A., Liborg, N.-K. & Rekdal, A.C. (2005). A survey of controlled experiments in software engineering. *Software Engineering, IEEE Transactions*. 31(9) p.733–753. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1514443.
- Tofan, D. (2011). *Questionnaire data for EASE 2011 short paper*. [Online]. Available from: <http://www.cs.rug.nl/~dan/EASE2011/> [Accessed: January, 2011].
- Wohlin, C., Höst, M. and Henningson, K. (2003). Empirical research methods in software engineering. *Empirical Methods and Studies in Software Engineering*. 9(3) p.7–23. Available from: <http://www.springerlink.com/index/UFKGYWMMVBTPC4M.pdf>.
- Yin, R.K. (2003). *Case study research: Design and methods*. Third edition. Sage Publications.
- Zannier, C., Melnik, G. and Maurer, F. (2006). On the success of empirical studies in the international conference on software engineering. *Proceeding of the 28th international conference on Software engineering - ICSE '06*: p.341. Available from: <http://portal.acm.org/citation.cfm?doi=1134285.1134333>.