# Recruiting participants and sampling items of interest in field studies of software engineering

Austen Rainer
*Electronics, Electrical Engineering and Computer Science*
*Queen's University Belfast*
Belfast, Northern Ireland
a.rainer@qub.ac.uk

Claes Wohlin
*Department of Software Engineering*
*Blekinge Institute of Technology*
SE-371 79 Karlskrona, Sweden
claes.wohlin@bth.se

***Abstract*—Software engineering research often samples participants for empirical studies. In this short paper, we briefly consider three concerns with sampling: (1) sampling participants may be misleading because the participant is not necessarily the item of interest; (2) existing guidelines do not seem to recognise the distinction between the recruitment of participants and the sampling of items of interest; and (3) existing guidelines provide limited concrete advice on recruiting "the right" participants.**

***Index Terms*—participant, sampling, recruitment, population**

## I. Introduction

In a recent article, Baltes and Ralph [1] provide a primer on sampling in software engineering (SE) research. They define sampling as the process of selecting a smaller group of items to study, the *sample*, from a larger group of items of interest, the *population*. A *sampling frame* is the available population list from which a sample can be actually drawn.

Many studies of SE use participants to study aspects of SE, e.g., sampling participants, in some way, for use in field studies such as case studies, interviews, or surveys. It appears that these studies tend to focus on sampling *participants*, and also focus on whether the *sample* of participants is, in some way, *representative* of practitioners. A representative sample is considered necessary for the findings from these participants to be generalised to the population. One ideal is that the sample is randomly drawn from the population.

Unfortunately, and taking Bouraffa and Maalej's [2] study as an example, it seems the most common, actual way of recruiting participants is through convenience sampling. Furthermore, there is a difference between sampling participants and sampling the actual item of interest. For example, if we are interested in critical success factors (CSFs) for software projects, the item of interest is the software project (or maybe the CSFs), and we sample participants for their experiences of particular software projects and for the beliefs they infer from those experiences. It is not clear how a convenience sample of participants maps to a representative sample of software projects.

As well as sampling participants, we therefore suggest the research community directs attention at sampling the items of empirical interest, e.g., requirements specifications, test cases, or development projects. We explore this concern in Section II, with an illustrative example: a paper by Curtis et al. [3]. The paper was published thirty years ago but the investigation by Bouraffa and Maalej [2] shows that these issues remain a challenge for SE research.

Moreover, it appears that current methodological guidelines on field studies in SE do not provide sufficient support for these concerns: guidelines do not seem to clearly distinguish between recruiting participants and sampling items of interest; the two issues seem to be conflated. We explore this concern, in Section III, with a brief review of published guidelines. In Section IV, we briefly describe future work and conclude.

## II. Illustrative example

Curtis et al. [3] report a field study of the software design process for large systems. The field study was part of a Software Technology Program established by MCC, the Microelectronics and Computer Technology Corporation. MCC was a research consortium with member companies.

For the field study, Curtis et al. [3] interviewed 97 participants from 17 projects in 9 of the member companies. Candidate projects were first identified by each member-company's liaison. Criteria were then used to select projects. Interviewed participants were the only source of information on the software design process.

Curtis et al. [3] structured their analysis using a Layered Behavioural Model of five layers: individual, team, project, company and business mileu. They clustered recurring problems identified from interviews into several themes, and then identified the three most salient themes for their paper: (1) the thin spread of application domain knowledge, (2) fluctuating and conflicting requirements, and (3) communication and coordination breakdowns. They include 43 quotations from participants, defined in terms of their functional role, to illustrate the themes. Thus, Curtis et al. [3] report their analysis along a two-dimensional matrix of three salient themes and five behavioural layers, supporting their analysis with participant quotes.

The illustrative example raises the following questions:

1) What is the item of interest? Is it the participant, the project, or the company, or the layers of the behavioural model, or the three salient themes, or some combination of these?

2) How does one trace observations from individual participants through to the final results about items of interest?

Given that the Curtis et al. [3] paper was published many years ago, we would hope that the two questions we enumerate above would have been answered since then, at least to some degree. But as shown by Bouraffa and Maalej [2], this is not the case. One hypothesis for the lack of answer is that methodological guidelines provide limited support for recruiting participants. To investigate this hypothesis, we briefly reviewed a set of guidelines and other advisory sources. Our review is presented in the next section.

## III. Guidelines on participant selection

To briefly investigate existing advice on recruiting participants separate from sampling items of interest, the two authors independently searched for guidelines and methodological advice on participant recruitment. We prioritised guidelines published *after* 2009, to be consistent with when the case study guidelines were published [4]. We had difficulties finding guidelines on interviews, so identified two "proxy" articles. We found ten sources.

We downloaded PDF copies of each article and the second author manually searched each PDF for explicit guidance on participant selection. We used nine keywords for that search. The objective of the searches was to identify formulations in the guidelines etc. in relation to advice concerning recruitment of participants for field studies. We were particularly looking for concrete advice, e.g., beyond general statements concerning the importance of recruiting representative participants. We searched for the following stemmed words:

Person: `subject*`, `partici*`, `respond*` and `contribu*`

Activity: `select*`, `identif*`, `sampl*`, `find*` and `recruit*`

A summary of our searches of the ten articles is presented in Table I. The counts shown in the table refer to the raw frequency ($f$) of occurrence of the nine keywords and to the relevant (e.g., in-context) frequencies and percentages of occurrence.

Overall, we find high level recommendations concerning, for example, representativeness and sampling strategy, but little concrete guidance on how researchers should recruit "the right" participants, and very little guidance recognising the distinction between participants and their recruitment, and items of interest and their sampling. As one example, Salleh et al. [5] present recommendations for recruiting practitioners. These recommendations concern, for example, how to interest practitioners in the research, how to find practitioners, and the importance of ensuring confidentiality, etc. Their recommendations do not consider the relationship of practitioners to items of interest. As a second example, Molléri et al. [6] frame their discussion of sampling in terms of selecting *respondents*. They also recognize the importance of stating research questions. It is the research questions that "determine" the *items of interest*, which are not necessarily the respondents. Again, there is no clear discussion of the relationship between the two.

TABLE I
Summary counts of searches of stemmed words.

| Article | Research method | Raw $f$ | Relevant $f$ | Relevant % |
|---|---|---|---|---|
| Runeson [4] | Case study | 125 | 8 | 7 |
| Verner [8] | Case study | 87 | 8 | 9 |
| Strandberg [9] | Interviews | 101 | 3 | 3 |
| Hove [10] | Interviews | 78 | 4 | 5 |
| Molléri [6] | Survey | 312 | 6 | 2 |
| Salleh [5] | General | 293 | 8 | 3 |
| Kontio [11] | Focus group | 138 | 8 | 6 |
| Kitchenham [12] | General | 114 | 4 | 4 |
| Zhang [13] | Ethnography | 116 | 1 | 1 |
| Ralph [14] | Standards | 45 | 1 | 2 |

Notably, the guidance provided in the recent SIGSOFT Empirical Standards does not explicitly recognise these issues. For example, for the *Sampling* Supplement online, the guidance states: "This standard applies to empirical research in which the researcher selects smaller groups of items to study (a *sample*) from a larger group of items of interest (the *population*) using a usually imperfect population list (the *sampling frame*)" [7]. There is no explicit recognition of the issues we raise here, i.e., the distinction between participants and items of interest.

## IV. Future Work and Conclusion

In ongoing work [15], we are formulating a framework for thinking about credible participants. The framework distinguishes participants from the items of interest, recognises that participants can have different roles in research, and recognises that participants may provide facts or beliefs *about* the items of interest.

Future work might conduct a critical review, in which the focus is not on the retrospective analysis of frequencies or prevalence of prior research, but rather on the identification of exemplar studies, on how the respective researchers distinguished participants from items of interest, and on how they explicitly traced observations from individual participants through to final results about items of interest. As two examples: Karlström and Runeson [16] present a model of the flow of information, from participant observations to final results, in the research process for their study; and Bjarnason et al. [17] distinguish the selection of companies (cases) from the selection of participants for interview, reporting interviewee demographics (including experience), and discussing threats to construct validity, arising from the selection of interviewees, and threats to external validity, arising from the selection of companies.

To briefly conclude: in this paper, we distinguished two kinds of item of interest, i.e., practitioners vs. software engineering phenomena. Sometimes the practitioners *are* the software engineering phenomena, but often practitioners are asked to provide insights into the phenomena. We suggest that software engineering research needs to better distinguish these two kinds of items of interest, and better understand how they relate.

## REFERENCES


[1] S. Baltes and P. Ralph, "Sampling in software engineering research: A critical review and guidelines," https://arxiv.org/abs/2002.07764, 2020.

[2] A. Bouraffa and W. Maalej, "Two decades of empirical research on developers' information needs: A preliminary analysis," in Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, ser. ICSEW'20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 71–77.

[3] B. Curtis, H. Krasner, and N. Iscoe, "A field study of the software design process for large systems," Communications of the ACM, vol. 31, no. 11, pp. 1268–1287, 1988.

[4] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," Empirical Software Engineering, vol. 14, no. 2, p. 131, 2009.

[5] N. Salleh, R. Hoda, M. T. Su, T. Kanij, and J. Grundy, "Recruitment, engagement and feedback in empirical software engineering studies in industrial contexts," Information and Software Technology, vol. 98, pp. 161–172, 2018.

[6] J. S. Molléri, K. Petersen, and E. Mendes, "An empirically evaluated checklist for surveys in software engineering," Information and Software Technology, vol. 119, p. 106240, 2020.

[7] Empirical Standards, "Sampling," downloaded 1 March 2022. [Online]. Available: https://acmsigsoft.github.io/EmpiricalStandards/Supplements/?supplement=Sampling

[8] J. M. Verner, J. Sampson, V. Tosic, N. A. Bakar, and B. A. Kitchenham, "Guidelines for industrially-based multiple case studies in software engineering," in Proceedings of the 3rd International Conference on Research Challenges in Information Science, 2009, pp. 313–324.

[9] P. E. Strandberg, "Ethical interviews in software engineering," in Proceedings of the 13th International Symposium on Empirical Software Engineering and Measurement, 2019, pp. 1–11.

[10] S. E. Hove and B. Anda, "Experiences from conducting semi-structured interviews in empirical software engineering research," in Proceedings of the 11th International Software Metrics Symposium, 2005, pp. 10–23.

[11] J. Kontio, J. Bragge, and L. Lehtola, "The focus group method as an empirical tool in software engineering," in Guide to Advanced Empirical Software Engineering, F. Shull, J. Singer, and D. I. K. Sjøberg, Eds. Springer, London, 2008, pp. 93–116.

[12] B. Kitchenham, S. Pfleeger, L. Pickard, P. Jones, D. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering," IEEE Transactions on Software Engineering, vol. 28, no. 8, pp. 721–734, 2002.

[13] H. Zhang, X. Huang, X. Zhou, H. Huang, and M. A. Babar, "Ethnographic research in software engineering: a critical review and checklist," in Proceedings of the 27th Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 2019, pp. 659–670.

[14] P. Ralph, "ACM SIGSOFT Empirical Standards Released," SIGSOFT Softw. Eng. Notes, vol. 46, no. 1, p. 19, Oct. 2021.

[15] A. Rainer and C. Wohlin, "Recruiting credible participants for field studies in software engineering research," 2021. [Online]. Available: https://arxiv.org/abs/2112.14186

[16] D. Karlström and P. Runeson, "Integrating agile software development into stage-gate managed product development," Empirical Software Engineering, vol. 11, no. June, pp. 203–225, 2006.

[17] E. Bjarnason, P. Runeson, M. Borg, M. Unterkalmsteiner, E. Engström, B. Regnell, G. Sabaliauskaite, A. Loconsole, T. Gorschek, and R. Feldt, "Challenges and practices in aligning requirements with verification and validation: a case study of six companies," Empirical Software Engineering, vol. 19, no. 6, pp. 1809–1855, 2014.