

On Recruiting Experienced GitHub Contributors for Interviews and Surveys on Prolific

Felipe Ebert*
Jheronimus Academy of Data
Science, Tilburg University
The Netherlands
f.ebert@jads.nl

Alexander Serebrenik
Eindhoven University of
Technology
The Netherlands
a.serebrenik@tue.nl

Christoph Treude
The University of Melbourne
Australia
christoph.treude@unimelb.edu.au

Nicole Novielli
University of Bari
Italy
nicole.novielli@uniba.it

Fernando Castor
Utrecht University
The Netherlands
f.j.castordelimafileho@uu.nl

ABSTRACT

Software engineering researchers have been using general purpose online tools for crowd-sourcing for quite some time. Those tools can be useful to recruit participants for research studies as they are paid for their time. However, those tools should be used carefully. In this paper, we have described the issues we faced when recruiting participants on Prolific. We used Prolific to recruit open-source developers with experience in submitting and reviewing pull requests (PRs).

However, we did not succeed in obtaining valid participants for either the interview or the survey, which led us to change the approach of our study and not use Prolific anymore. For example, a major issue we faced on Prolific was that even using the specific filter related to participants with software development knowledge, the screening study revealed the majority of them did not have any development knowledge. Based on this experience, we formulate recommendations for future research when using such crowd-sourcing tools.

ACM Reference Format:

Felipe Ebert, Alexander Serebrenik, Christoph Treude, Nicole Novielli, and Fernando Castor. 2022. On Recruiting Experienced GitHub Contributors for Interviews and Surveys on Prolific. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Software engineering researchers have been recruiting participants for studies for a long time. The input provided by

*This work has been conducted while the first author was affiliated with Eindhoven University of Technology.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . \$15.00
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

those participants is essential for research. However, recruiting participants in the area of Software Engineering can be challenging, hard, and exhausting. In the last few years, online tools for crowd-sourcing such as Prolific¹ and Mechanical Turk² have been aiding in the recruitment of participants for scientific studies [3, 4]. These platforms enable researchers to recruit participants matching the desired profile, in line with the research goals. Furthermore, they enable researchers to pay participants for performing specific tasks, similar to recruitment in scientific studies in other areas [2]. For example, Prolific, which is described to “connect those doing research (academics, startups, data scientists, organizations) with their target participants. Participants earn rewards for participating in studies on Prolific.”³, allows researchers to choose the amount to be paid to participants within a certain range, but the system helps and indicates a fair value based on how much time the task will require. While crowd-sourcing platforms are *generic* as the participants available vary in their demographics (*e.g.*, nationalities, age, education), they also provide filters so that researchers can target a specific sub-population. For example, Prolific has a specific filter option to identify participants with “*knowledge of software development techniques*”.

In this *position* paper, we describe the issues we faced while using Prolific to recruit participants for an interview and a survey for an on-going study (*i.e.*, not published yet), which from now on we will refer to as the *original study*. While those tools provide an *easy and quick* way to recruit participants, we have observed that the responses can be quite noisy depending on the study goal. For example, one issue we faced was establishing how experienced potential participants actually were in contributing to open-source projects and performing code reviews. Finally, we present suggestions for future studies based on the lessons we learned.

2 RECRUITING ON PROLIFIC

Our original study aimed at interviewing developers who had contributed to open-source projects by both submitting

¹<https://prolific.co>

²<https://www.mturk.com>

³<https://researcher-help.prolific.co/hc/en-gb>

Table 1: Results of the interview invites.

Invites sent	54
Accepted invite	15
Declined invite	19
No answer	20
No-shows	6
Opt-out	1
Interviews conducted	8
Interviews disregarded	6
Interviews valid for the study	2

and reviewing PRs. As we were able to recruit only three participants from the open-source community, we decided to try Prolific to recruit additional open-source developers.

Following the recommendations⁴ from Prolific, we decided to pose a screening survey before inviting participants for the interview in order to increase the chances of inviting only participants of interest for our study. We used the filter on Prolific regarding software development in all our studies. A screening survey with 300 participants was conducted with two questions: *Have you contributed to any open-source project?* and *Have you reviewed code changes in any of those open-source projects?* The number of participants who answered *yes* to both questions was 74. From those, we invited a total of 54 random participants for the interview due to our budget restrictions. Table 1 depicts this process.

From the 54 invitations sent, only 15 participants answered positively and scheduled the interview. Then six of the participants did not show up, one decided to opt-out before the interview, resulting in eight interviews being conducted and only two of these interviews were valid for our study, *i.e.*, only two participants actually had experience submitting and reviewing PRs in open-source projects. For the participants we excluded, they had a *divergent* definition of what constitutes an open-source project.⁵

Based on the low number of valid participants for the interview, we decided to modify the study approach and pose our questions in a survey aiming for higher interest and engagement of participants on Prolific. We used our remaining budget to run a screening survey, which was enough to recruit 22 participants. This time we decided to pose more specific questions about general programming and open-source development knowledge based on the recommendations of Danilova *et al.* [1].

For example, we asked questions such as:

- Q1) Which of these websites do you use most frequently as an aid when programming?
- Q2) Choose the answer that best fits the description of a compiler's function
- Q3) Which of these values would be the most fitting for a Boolean?
- Q4) Do you have a GitHub account?
- Q5) How many pull requests have you submitted on GitHub? — we asked participants to login on GitHub, access the URL⁶, and then type sum of the PRs.
- Q6) Have you ever submitted a pull request to any open-source project?
- Q7) Have you ever reviewed pull requests on any open-source project?
- Q8) How experienced would you consider yourself at programming?
- Q9) What is your profession?

⁴<https://researcher-help.prolific.co/hc/en-gb/articles/360010165173-Can-I-screen-participants-within-my-study->

⁵<https://opensource.org/osd>

⁶<https://github.com/pulls>

Surprisingly, six participants explicitly answered they do not program (Q1), nine wrongly answered question Q2, ten wrongly answered Q3, 16 participants do not have a GitHub account (Q4), only one participant informed us that they had submitted a PR to GitHub (Q5), but for Q6 and Q7 no one answered positively for submitting and reviewing PRs. The answers about the respondents' profession show only five participants working on software development. The remaining ones have different professions, such as artisan, store manager, and photographer.

3 ISSUES & RECOMMENDATIONS

Issue: on Prolific, when a person registers to participate in studies, they are not required to select any topics regarding their knowledge but they can select as many topics as they want (Prolific does not provide any means to verify this information). In addition, Prolific participants are using the platform to earn money. We believe those factors could motivate some Prolific participants to indicate knowledge on topics they actually do not have: we have seen that most of the participants presumably knowing software development techniques do not really know programming (questions Q2, Q3), do not program (Q1), and are not familiar with the dominant open-source forge GitHub (Q4, Q5, and Q6). **Recommendation:** researchers should always conduct screening studies to confirm participants' knowledge on the topic.

Issue: we noticed inconsistent answers when asking participants to type the number of PRs they have submitted (Q5), one of them answered "six" but the answers to the questions whether they submitted or reviewed PRs (Q6 and Q7) were negative. Participants might be tempted to lie in their answers. Alternatively, participants can simply perpetrate typographic mistakes due to retyping. **Recommendation:** researchers should validate participants' answers whenever possible. For example, we have asked participants to visit the GitHub URL which shows their PRs and type the number they see on it.

Issue: we needed to exclude several participants after conducting the interviews because they had a *divergent* understanding of the definition of open-source project than the one from the Open Source Initiative. Asking clear and concise screening questions while obtaining useful answers is hard. **Recommendation:** researchers should run pilots with the target population. However, a sufficient number of participants in the pilot is difficult to determine. In our case, we have run three pilots, and despite this there were still misconceptions regarding the definition of an open-source project.

Issue: on Prolific, when one starts a study, all participants eligible can register to it in the order of *first come first served* until all the spots are filled in. In our case, it resulted that in the first study (where we did not filter out students) the majority of the participants were students, maybe because they have more *free time* to be checking available studies on Prolific. **Recommendation:** researchers should involve multiple recruitment rounds at different times (weekend vs. working day, times of the day, etc.).

REFERENCES

- 233
234 [1] A. Danilova, A. Naiakshina, S. Horstmann, and M. Smith. 2021. Do you Really Code? Designing and Evaluating Screening Questions for Online Surveys with Programmers. In *ICSE*. 537–548. 291
235
236 [2] Y. J. Lee, J. A. Arida, and H. S. Donovan. 2017. The application of crowdsourcing approaches to cancer research: a systematic review. *Cancer medicine* 6, 11 (2017), 2595–2605. 292
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
- [3] A. M. Mellis and W. K. Bickel. 2020. Mechanical Turk data collection in addiction research: utility, concerns and best practices. *Addiction* 115, 10 (2020), 1960–1968. 293
[4] T. Wise, T. D. Zbozinek, G. Michelini, C. C. Hagan, and D. Mobbs. 2020. Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. *Royal Society Open Science* 7, 9 (2020), 200742. 294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348