Under the Bridge: Trolling and the Challenges of Recruiting Software Developers for Empirical Research Studies

1st Ella Kokinda
School of Computing
Clemson University
Clemson, United States
ekokind@clemson.edu

2nd Makayla Moster
School of Computing
Clemson University
Clemson, United States
mmoster@clemson.edu

3rd Paige Rodeghero
School of Computing
Clemson University
Clemson, United States
prodegh@clemson.edu

Abstract—Much of software engineering research focuses on tools, algorithms, and optimization of software. Recently, we, as a community, have come to acknowledge that there is a gap in meta-research and addressing the human-factors in software engineering research. Through meta research, we aim to deepen our understanding of online participant recruitment and human-subjects software engineering research. In this paper we motivate the need to consider the unique challenges that human studies pose in software engineering research. We present several challenges faced by our research team in several distinct research studies, how they affected research, and motivate how, as researchers, we can address these challenges. We present results from a pilot study and categorize issues faced into three broad categories including participant recruitment, community engagement, and data poisoning. We further discuss how we can address these challenges and outline the benefits a full-study could provide to the software engineering research community.

Index Terms—software development, meta-research, software engineering research, research methods

I. INTRODUCTION

As part of the software engineering research community, we need to research software developers and their behavior within the profession. For early career researchers and graduate students, recruiting software developers is an arduous task, with researchers turning to online participant recruitment. Online recruitment methods vary from cold-emailing or messaging, posting on social media like Facebook, Twitter, or LinkedIn, as well as using community social platforms like Slack, Discord, Gitter, and Reddit. However, new and veteran researchers alike may unexpectedly encounter toxic responses from social media users during recruitment, seemingly out of nowhere and without an avenue to rectify the situation they are now in. Toxicity is well studied in social media use and within leadership contexts [1]–[6]. However, it is rarely shown or spoken of in a research context. How do researchers handle their work being undermined before it is published? How do we rectify push-back from communities where target participants reside? How do we approach developers and communities without alienating or overstepping boundaries? Finally, how should the research community report these challenges?

Recent meta-studies make strides toward understanding and defining improved strategies for participant recruitment, however, many do not address encountering difficult participants and unexpected data [7]. Our preliminary and proposed research aims to work hand-in-hand with recent studies regarding participant recruitment, help address the challenges that participant responses and behaviors pose, and provide transparency in research processes. Our goal is to invite the research community to provide feedback in the initial stages of this work. We seek to expand this work, find successful solutions to these challenges, and encourage those facing similar struggles to reach out. By understanding the challenges researchers face in online recruitment and data collection for human studies, we can inform and influence methods and approaches for participant recruitment and better understand empirical software engineering research approaches.

II. BACKGROUND

In this section, we discuss the importance of meta research and its impacts on improving research methodology. Additionally, we discuss participant recruitment in social sciences.

A. Importance of Meta Research

Meta research "aims to evaluate and improve research practices," and we believe that investigating our own research practices ensures effective and credible results. [8]. Understanding appropriate research methods is foundational to study success, however, it has been often over-looked when applying across research disciplines [8]. Through the analysis and adapting the methods of other research disciplines, such as social sciences and psychology, software engineering researchers can adapt research methodologies to their research studies.

B. Participant Recruitment in Software Engineering Research

Historically, research shows convenience sampling to be the most predominant sampling strategy within software engineering research and that many participants come from a few companies in industry [9]. These historical sampling strategies raise concerns for generalizablity and and overall

appropriateness of sample size. Recently, many researchers use paid platforms or other hosted sampling services and found that many of the same challenges sampling and recruiting appropriate and relevant participants and screen question challenges [10], [11]. Additional recent work aims to shed light on addressing recruitment challenges surrounding sampling, participant screening, and using student software engineers as proxies for full-time software developers [12]–[15]. Finally, it is important to note prominent one example of recruitment tension, and ongoing discussion on recruitment practices between researchers and software developers, where researchers were using a third-party site GHTorrent to solicit names and emails of developers [16], [17].

III. RESEARCH DESIGN

In this section, we outline our research goals for both the pilot study and subsequent research, as well as, methods approach for a full study.

A. Research Goals

Motivated by our experiences recruiting software developers for research studies, we believe the following research questions (RQs) can establish a baseline approach for understanding the recruitment challenges and their effects on research. The goal of our meta-research within software engineering is to deepen our understanding of approaches to online developer-participant recruitment and bring awareness to challenges faced by software engineering researchers which may not be readily apparent.

RQ_1: What challenges do software engineering researchers encounter when recruiting human participants for research studies?

RQ_2: How have these challenges impacted research?

RQ_3: How have these challenges impacted researchers' personal well-being?

RQ_4: What lessons have software engineering researchers learned from recruitment challenges encountered?

The rationale behind RQ_1 is to understand the extent and magnitude of challenges present in human-subjects software engineering research. From there, RQ_2 and RQ_3 aim to understand the impact of challenges on researchers. Finally, RQ_4 addresses how researchers can lessen the impact of challenges and understand what has helped when a challenge is posed during research.

B. Qualitative Research Approach

1) Pilot Study: For our pilot study, authors and previous members of our research group came together to share experiences and challenges encountered while recruiting for software development research. In total, we analyzed 7 different research studies and found shared and unique challenges from these studies. From there, the team classified challenges into 3 broad categories: participant recruitment, community engagement, and data poisoning. Table I provides a breakdown of the research studies and the challenges they pose.

2) Proposed Study: Using the RQs from our pilot study, we propose a survey and interview-based approach to meta-research within software engineering researchers. Our proposed survey will elicit both quantitative and qualitative data points regarding challenges and quantify their impacts on research from software engineering researchers who recruit human participants. We propose using standard statistical analysis for quantitative questions using methods outlined in [20]. For qualitative data from the survey and interviews we propose using open-coding techniques [21].

IV. PRELIMINARY RESULTS

In this section we outline experiences and challenges faced by the authors while recruiting human subjects in software engineering studies. Additionally we highlight the impacts of the challenges on research and well-being.

A. Challenges

Several types of challenges emerged from authors' experiences with many challenges surrounding approaching communities to solicit research, response from communities, concerns over oversampling from niche participant pools, and participants purposefully subverting their data.

Participant Recruitment – Necessary to human-subjects research are the participants, and as researchers we run into several issues with recruitment, especially online recruitment. One of the largest problems faced by our researchers, across all studies for in-person events or interviews were no-shows or ghosting[1]. Studies S3, S4, S5.1, and S7 faced the challenges of participants not showing up or ghosting after reaching out to reschedule. Participants who do not show up to an agreed time can hinder research progress such as study cancellations, demotivate future research projects, and can put researchers in negative or awkward situations with participants who arrived on time. In Study 5.1, in-person research needed to be cancelled due to lack of participants, when emailing with a weeks notice to those who signed up, researchers faced backlash from a participant that emailed over the course of several hours:

💬 *"You are doing research for A university that regularly researches stuff. You're telling me you have no money. Riiiighhhtttttt. Go f*** yourself."*

💬 *"I will personally make sure that everyone knows not to waste their time on [University]'s little programming thing that you have. My Venmo is @[VenmoHandle] if you want to compensate me."*

💬 *"I actually am a [Rival University] fan and this is the most Trump thing I could of possibly expected for from a good Ole Boys."*

💬 *"Just don't try to bring your bulls*** surveys to [city]. We have a beach and don't need to be dealing with snowflakes."*
Additionally, within S5.1 and S5.2, recruiting paid participants with appropriate experience for a software development study proved challenging. Participants expressed interest in the study

---

[1] Ghosting is ending contact with someone after prior communication or agreement without justification and any subsequent followup or acknowledgement

Study #	Ref.	Recruitment Details	Challenges
1	[18]	Reddit	Participant Recruitment, community engagement
2	Under Review	Reddit, Discord, Other Social media	Participant Recruitment, Community Engagement, Data Poisoning
3	Under Revisions	Email, Twitter Direct Messages	Community Engagement
4	Under Review	Personal Contacts	Participant Recruitment
5.1*	[19]	Flyers, Social Media	Participant Recruitment
5.2**	[19]	Social Media	Participant Recruitment
6	Confidential Research	Social Media, Flyers	Data Poisoning
7	Accepted, to be published 2023	Reddit, Linked-In, Twitter	Participant Recruitment, Community Engagement, Data Poisoning

but lacked the appropriate programming experience to participate, with 9 out of 89 initial participants (10%) ultimately being rejected from the study. With niche participant requirements or unique research interests, each researcher encountered concerns surrounding over-fishing problems (reaching out to the same communities for participants) and ultimately reaching enough participants to make research generalizable.

Community Engagement – In the case of S1, S2, and S6 recruitment using online communities posed several issues regarding community rules related to research solicitations, community moderator responses, and online platform automated spam filters. When reaching out to online communities it is best to understand the community norms and rules, which some define clearly and others do not. We found that many communities who would be appropriate to invite to participate in research do not allow for these types or solicitations or to contact the moderation team for permission. When contacting the moderator teams for permission when communities ask or when rules are unclear, we have faced several challenges - no responses and hostile reactions. In S7, the research team found that 202 out of 1000 (20.2%) subreddit moderator teams reached out to were willing to allow advertisements or post. When presenting to moderators our team would give an overview of the research and its affiliated university, why we are reaching out, outline benefits and protections for participants, state its IRB approval, and often a link to the survey for it to be previewed. While most communities would not respond or politely decline, others took a hostile approach with verbal abuse and name calling:

- 💬 *"You are dogsh**."*
- 💬 *"You are thieves stealing from this website, thats a fact"*
- 💬 *"Come back with cooler ideas and maybe I will reconsider."*
- 💬 *"No, and [if] you post I will remove it."*

Additionally, we found we would be banned from communities for reaching out to moderation inboxes with no further explanation or communication from the moderation team.

Data Poisoning – While not a problem for everyone, several authors run into "troll" or disruptive behaviors from research participants through their answers on survey questions and in responses to interview questions. In S7, researchers encountered those using gender self-description boxes to voice their opinions like *"there are only two genders"* or non-sensible answers like *"dinosaur."* Another researcher, in S2, encoun-

tered hostile survey responses from consenting participants and aimed at subverting the research, with participant responses to questions such as *"f*** you"* and *"My answers have derailed this survey to the point my answers are probably meaningless to your research."*

B. Impact of Challenges

Impacts from these challenges varied from researcher to researcher, but all expressed frustration and demotivating about the research. When data subversion occurred, in the cases of S7 and S2, researchers indicated uncertainty if subsequent data in those responses were genuine and should be included. Particularly in the case of S2, researchers reported that data from these participants could be used in some capacity but questioned overall inclusion of these responses, as they did not have a large impact on overall generalizability of results, and were often outliers.

For those who faced participant recruitment challenges, over-fishing the same communities, finding appropriate participants, and concerns over small sample size led researchers to question if online recruitment methods are worth the hassle and appropriate channels to pursue for research. Additionally, the frustrations of challenges led to overall demotivation to continue research, hesitation toward using certain social media platforms, and takes a toll on personal well-being.

V. DISCUSSION

Research emphasizes the need to reach broader and diverse participants and achieve larger representative sample sizes, researchers must understand how to filter, address, and/or use data from participants who may provide difficult data [7]. From our example studies, participant response data that proved challenging, we did not want to completely exclude, but often did not make any point other than being outlier data.

It is imperative to acknowledge that researchers may be the root cause of their own problems, and that hostile behavior from participants or communities may reside in how we, as researchers, interact, interface, and subsequently treat participants and communities. When soliciting developer communities to participate in research, we recommend to provide concise and digestible information about the study as well as clearly define why developers should participate when reaching out to communities. We found it helpful and respectful to reach out to moderation teams when questioning if research

solicitations were appropriate for the community or when reaching out to individuals. Below we provide exemplar texts from reaching out to community moderators and individuals:

Community Moderators Boilerplate

"Hey [Community] Mods!

I am a human factors and software development researcher from [University] and am looking to recruit participants for an IRB approved research study on [details here].

This study focuses on [Additional detail of your study and why you are reaching out to this community]. From this research I am looking to understand [explain the benefits of your research and why participants from here should participate]

I am more than happy to answer any questions y'all may have, and I appreciate your time!

All data is anonymous. You can find the survey here: [link to survey]."

Individual Boilerplate

"Hi [Individual's Name]!

My name is [Name Here], and I am a [Current Position] at [Affiliated University/Company] under advisement of [Advisor/Research Group] ([Links to personnel or research group homepages]). I am looking to recruit participants for an [Interview/Survey] that will take [Time Commitment] for research related to [topic]. I am really interested in [explain why you are reaching out to this person in particular]! We are investigating [things you think will help you move forward with your research with this individuals help].

What this interview would entail would be a short Q&A session regarding [overview of what you plan to ask about].

I appreciate your time in reading this, and am more than happy to answer any questions you may have!"

Recommendations based upon observations for Software Developer Research Recruitment
- Introduce the work concisely and in plain, understandable language
- Identify researchers, affiliated institutions
- Explain why you are approaching this person or community
- Motivate benefits of participation
- Clear and concise impacts to participants, if any
- Identify how you handle or anonymize data
- Provide contact information for questions and concerns
- Provide a link to survey or interview questions for moderator preview (if applicable)

These outlined recommendations for participant recruitment were used in part for studies 2, 3, and 7, and fully resulted from feedback we received from participants or community moderators while soliciting participants for these same studies.

Additionally, we believe that efforts to increase sample sizes and pull from diverse participant pools may help alleviate challenges from reaching out to communities, find broader participation and support from online software development communities, and lead to more generalizable research results.

Brown's research outlines a standalone website where researchers may post research opportunities along with unique incentives for participation [22]. Additionally, we propose as part of our future work to extend presence on current social media outlets and create a community for software developers to engage and interact with researchers on proposed and upcoming work. Through this we hope to foster meaningful interactions between developers and researchers and address identified challenges in current software engineering research recruitment methods.

Though everyone faces challenges differently, we believe that providing transparency in research challenges will encourage new researchers, re-motivate those who feel discouraged, and compliment recent work with respect to participant sampling approaches. By exploring alternative methods to participant recruitment, we may also alleviate and lessen the impacts from these observed challenges.

VI. FUTURE PLANS

For future work, we propose a two-part study using interviews and surveys continuing meta-analysis within the software engineering research community investigating research methods and researcher approaches to communities, as well as working with software developer community moderators and leaders to understand how researchers approach and interface with their communities. Through this we hope to understand how and why hostile responses from participants and communities may arise, and what, if anything, researchers can do to mend relationships with participants and communities if given the opportunity.

Continuing meta-analysis should investigate participation fatigue - how often are researchers reaching out to communities, do communities feel spammed or exhausted in being continually asked for their time? It may be beneficial to work with these communities to understand the types of requests they are receiving from researchers and non-researchers alike who are turning to these communities for input. Understanding what is asked of these communities with respect to their participation in research can help researchers and these communities form a trustworthy, and mutually beneficial relationship. However, as researchers, we need to understand that some communities may not work to work with us, they do not owe us their participation or time, and we must respect that.

Future work will self-reflect and compare to other research disciplines on our approaches to participant recruitment and relationships with developer communities. Are participant recruitment methods clear, concise, and relevant to the developers we want to reach? Are recruitment solicitations upfront about their impact on participant developers and justify time spent participating?

VII. Conclusion

By sharing our research teams' experiences, we highlight the challenges that software engineering researchers face and show how these challenges impact research. We believe that by sharing these experiences, the software development research community will come together and share their experiences with online recruitment and human subjects research, and contribute to progressing research methodologies and strategies for future research. Through subsequent research, our work will contribute to re-framing recruitment methods and engage in research practices that are more generalizable and reflective of the populations we study. Finally, we invite feedback and collaboration from researchers in the software engineering community who may have experienced similar recruitment difficulties and want to collaborate towards meaningful contributions and solutions.

References

[1] H. Almerekhi, S. b. B. J. Jansen, and c.-s. b. H. Kwak, "Investigating toxicity across multiple reddit communities, users, and moderators," in *Companion proceedings of the web conference 2020*, 2020, pp. 294–298.

[2] S. Mohan, A. Guha, M. Harris, F. Popowich, A. Schuster, and C. Priebe, "The impact of toxic language on the health of reddit communities," in *Canadian Conference on Artificial Intelligence*. Springer, 2017, pp. 51–56.

[3] J. Blackburn and H. Kwak, "Stfu noob! predicting crowdsourced decisions on toxic behavior in online games," in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 877–888.

[4] K. L. Pelletier, "Leader toxicity: An empirical investigation of toxic behavior and rhetoric," *Leadership*, vol. 6, no. 4, pp. 373–389, 2010.

[5] C. J. Carpenter, "Narcissism on facebook: Self-promotional and anti-social behavior," *Personality and individual differences*, vol. 52, no. 4, pp. 482–486, 2012.

[6] J. Golbeck, Z. Ashktorab, R. O. Banjo, A. Berlinger, S. Bhagwan, C. Buntain, P. Cheakalos, A. A. Geller, R. K. Gnanasekaran, R. R. Gunasekaran *et al.*, "A large labeled corpus for online harassment research," in *Proceedings of the 2017 ACM on web science conference*, 2017, pp. 229–233.

[7] A. Rainer and C. Wohlin, "Recruiting credible participants for field studies in software engineering research," *Information and Software Technology*, vol. 151, p. 107002, 2022.

[8] J. P. Ioannidis, D. Fanelli, D. D. Dunne, and S. N. Goodman, "Meta-research: evaluation and improvement of research methods and practices," *PLoS biology*, vol. 13, no. 10, p. e1002264, 2015.

[9] 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*, 2020, pp. 71–77.

[10] F. Ebert, A. Serebrenik, C. Treude, N. Novielli, and F. Castor, "On recruiting experienced github contributors for interviews and surveys on prolific," in *International Workshop on Recruiting Participants for Empirical Software Engineering*, 2022.

[11] D. Russo, "Recruiting software engineers on prolific," *arXiv preprint arXiv:2203.14695*, 2022.

[12] M. Endres, W. Weimer, and A. Kamil, "Making a gamble: Recruiting se participants on a budget," 2022.

[13] A. Danilova, A. Naiakshina, S. Horstmann, and M. Smith, "Do you really code? designing and evaluating screening questions for online surveys with programmers," in *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, May 2021, pp. 537–548, iSSN: 1558-1225.

[14] M. Smith, A. Danilova, and A. Naiakshima, "A meta-research agenda for recruitment and study design for developer studies," 2022.

[15] C. Brandt and A. Zaidman, "Strategies and challenges in recruiting interview participants for a qualitative evaluation," 2022.

[16] G. Gousios, "The issue32 incident," May 2016. [Online]. Available: https://speakerdeck.com/gousiosg/the-number-issue32-incident

[17] ——, "The issue 32 incident – an update," Mar 2016. [Online]. Available: https://gousios.org/blog/Issue-thirty-two.html

[18] M. Moster, A. Chandra, C. Chu, W. Liu, and P. Rodeghero, "In the zone: An analysis of the music practices of remote software developers," in *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2022, pp. 313–318.

[19] J. Dominic, B. Tubre, C. Ritter, J. Houser, C. Smith, and P. Rodeghero, "Remote pair programming in virtual reality," in *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2020, pp. 406–417.

[20] Z. Ali and S. B. Bhaskar, "Basic statistical tools in research and data analysis," *Indian journal of anaesthesia*, vol. 60, no. 9, p. 662, 2016.

[21] J. Corbin and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[22] C. Brown, "Nudging developers to participate in se research," 2022.