The Key Considerations In Building A Crowd-testing Platform For Software Developers

Qamar Naith
The University of Sheffield, and University of Jeddah
Sheffield, United Kingdom
qhnaith1@sheffield.ac.uk,qnaith@uj.edu.sa

Fabio Ciravegna
The University of Sheffield
Sheffield, United Kingdom
f.ciravegna@sheffield.ac.uk

ABSTRACT

External testing of mobile software on a larger number of mobile devices by several users is often needed to ensure quality. Currently, the evidence as to what extent developers accept large-scale crowd-testing is limited. This paper aims to (1) gauge developers' perspectives with respect to the participation of the public and anonymous crowd testers, with varied experiences; (2) gather the developers' needs that could reduce their concerns of dealing with the public crowd testers and increase the opportunity of using the crowd-testing platforms. An online exploratory survey, conducted to included 50 Android and iOS developers from different countries with diverse experiences. This paper revealed several findings including the information that must be provided by developers and crowd testers for achieving effective crowd-testing process; the factors that can ensure the reliability and accuracy of the results provided by the public crowd testers. The findings conclude that (90%) of developers are potentially willing to perform testing via the public crowd testers worldwide. This on condition that several fundamental features were available which enable them to perform more realistic tests without artificial environments on large numbers of devices. The results also demonstrated that a group of developers does not consider testing as a serious job that they have to pay for, which can affect the gig-economy and global market. We aim at helping the individual or small development teams who have limited resources to perform large-scale testing of their products.

CCS CONCEPTS

• **Information systems → Crowdsourcing**; • **Human-centered computing → Human computer interaction (HCI)**; *Usability testing*; *Ubiquitous and mobile devices*; • **Software and its engineering →** *Software testing and debugging*; Software usability;

KEYWORDS

Mobile App testing, Public and Anonymous Crowd Testers, Large-scale crowd-testing, Gig-economy

ICCSE'19, October 18–21, 2019, Jinan, China

ACM ISBN 978-1-4503-7640-2/19/10…$15.00
https://doi.org/10.1145/3371238.3371247

1 INTRODUCTION

Testing mobile apps to ensure their quality remains difficult because of the diversity of mobile devices and operating system versions [7]. The research in this area is still in its early stages. This is due to a major barrier to large-scale deployment of the app by small and medium-sized enterprises (SMEs) as they do not have the ability to test a large number of devices. In the literature, several solutions have been proposed to address the issues of testing apps on a variety of mobile device models and OS versions [2, 7–9, 12, 14]. These solutions have been classified into two categories: (1) Automated testing frameworks or tools over the cloud environment, MAT [12], Testdroid [9], MobiTest [2]. It allows developers remotely accessing a pool of real mobile devices that are available and connected to the one main online server over a cloud (it covers only one geographical location). (2) Cloud-based testing service tools, RTMS [8], AppACTS [7], CTOMS [14], by comparison, allows developers to access a larger number of devices through different pools of real mobile devices connected to several online servers over many clouds (cover limited geographical locations).

Most of these solutions have tried to address this testing issue. Still, several challenges hamper its full success, such as lack of realistic tests, lack of providing a complete set of devices and OS versions to be tested, a limited number of users, a limited geographical location, limited users' behaviors. This made developers move to cooperate with traditional testing organizations instead, which also maybe still unable to cover the full breadth of global mobile devices the startup's target markets would use. Current crowd-testing organizations [e.g., uTest[1], 99tests[2], Mob4Hire[3], and BugFinders[4] etc] have served in minimizing this issue by leveraging the power of a specific group of testers belonging to their particular communities. However, developers still need to test their apps on a larger number of devices by several external users with different behaviors to ensure the quality of their apps.

To improve the solution of crowd-testing organizations and address this issue, we proposed in [11] a new solution based on open crowd-testing which participation of the public and anonymous testers over the world. It is likely that this solution will allow a larger number of users (crowd testers) with different behaviors and testing experiences, to participate and perform the test on their own devices. Thus, developers could cover more devices. In order to explore how developers look at this solution and to understand their perception on this matter, two research questions have been asked:

[1] https://www.utest.com/
[2] https://99tests.com/
[3] http://www.mob4hire.com
[4] https://www.bugfinders.com/

RQ1: *To what extent mobile app developers are keen to work with the public and anonymous crowd testers with various levels of experience to perform their testing tasks?*
RQ2: *What are the key factors required to build a large scale crowd-testing platform?*

To investigate these two research question, we carried out an online exploratory survey (questionnaire) shared on Twitter, LinkedIn, and Facebook. The questionnaire involved the participation of 50 Android and iOS mobile app developers. Mixed qualitative and quantitative analysis was conducted to analyze the collected data.

Our findings show that more than half of the developers never used crowd-testing before while a minority of them have little experience with. In related to their willingness to use the open crowd-testing method, (90%) of developers have agreed to perform their test by the public crowd testers worldwide. While (57%) of them thought that a critical mass of testing could be achieved if the public crowd-testing was used under several conditions. (68%) confirmed that they trust crowd testers from any part of the world if some serious cases are taken to ensure the reliability and accuracy of results. As (83%) of them stressed the importance of providing textual explanations about the entire testing process as evidence of the accuracy of results. The findings also clarify the developers' requirements to use the public crowd-testing including (1) the most important searching elements that developers focus on when they seek a solution to programming problems or specific testing issues; (2) the important pieces of information that must be exchanged between both developers crowd testers for achieving a correct crowd-testing process; (3) the list of potential rewarding that developers are willing to offer to crowd testers.

To summarize, this paper makes the following contributions: (1) It provides an in-depth investigation and an initial view about the developers' opinions and measures their willingness to work with a member of the public crowd testers for compatibility testing of their apps with mobile devices. (2) It shows the developers' needs to use the public crowd-testing and assesses the criteria that must be included in a the public and large-scale crowd-testing platform including the factors that will reduce developers' concern when dealing with the public and anonymous crowd testers and increase their confidence in them. (3) It helps developers to understand how to deal with the large-scale of crowd testers in terms of motivation and trust methods, the way of defining and submission tasks as well as required information during the testing process.

The rest of this paper is organized as follows. The method of designing the survey study is presented in Section 2.1. In Section 3 we present the results of analyzed data. Section 4 provides a discussion of the main findings. A list of recommendations is presented in Section 5. Finally, Section 6 concludes the paper.

2 METHODS

2.1 The Study Design

This exploratory study employed a mixed qualitative and quantitative descriptive design method[3] to examine collected data from the online questionnaire (using Google Forms) to provide a clear answer to the two research questions mentioned in this paper. This questionnaire was released online to the public for over two months. A Cross-sectional design method was used to collect the data from participants at a single point in time, rather than separate time periods [3]. The responses of the participant were considered as consent to their participation in this research study. The following sections briefly describe the protocol have been used to achieve this exploratory study.

2.2 Population and Sampling Strategy

Since the target population for this study is mobile app developers, we sampled the population by apply random sampling [4]. The sampling has been carried out through distributing randomly the questionnaire to a large variety of mobile apps development groups on Twitter, LinkedIn, Facebook. This online survey was fully completed by 50 random participant developers from different countries around the world (such as Saudi Arabia, Egypt, United Kingdom, Canada, Germany, Singapore, Sweden, Romania United Arab Emirates United States) who have various experiences in Android and iOS. This random probability sampling of participants assists in generalizing the gathered information from participant developers to the entire population without significant discrepancies [4].

2.3 Questionnaire Design

This questionnaire was designed to be strictly anonymous where no demographic or educational information for any participants was collected. Therefore, it was impossible to identify the details of participants or link any answers to a specific person. From the literature, gaps discovered in previous studies, we designed our questionnaire questions, with the two research questions discussed above in mind. This questionnaire was designed to include a mixture of *close-ended* (multiple choice, ranked, and rating questions) and *open-ended* questions [5]. The questionnaire involved 15 questions, split into six sections: (1) Developers' experiences with crowdsourcing; (2) Developers' expectations of the public crowd-testing; (3) Key requirements for the public crowdsourced testing platform; (4) Level of confidence in the public crowd testers; (5) Desired features in/from crowdsourced testing system; (6) Required information for the public crowd-testing process. For more information about questions included in each individual section in the questionnaire, the interested reader will find the online questionnaire available at Survey[5].

3 THE STUDY RESULTS

This section summarizes the main findings that emerged from the qualitative and quantitative analysis conducted on gathered information from the online survey. The responses of participants were highly insightful and illustrated how much mobile apps developers are agreeable to work with the public and anonymous crowd testers directly without the need for a manager or leader as in most crowd-testing methods used by testing organizations. In addition, it has highlighted a number of actions that must exist in proposed the public crowd-testing solution to be used which can enhance trusting of crowd testers. The discussion of the main survey findings: measuring the extent of using crowdsourcing platforms by developers; the developers' expectations from the public crowd testing method;

[5] https://goo.gl/forms/QEYjXibLySvo0GrT2

developers' requirements for using the crowd-testing method; measuring the level of integrity in the public crowd testers; desired features and outcomes from the public crowd-testing method; required information for achieving accurate and effective testing process. The details for all these findings are presented in the next subsections.

3.1 Developers' experiences with crowdsourcing

(Q.1.) How do you usually do your mobile app testing?

Answer: Table 1 shows how frequently crowd-testing platforms and other testing methods such as private testing companies and automated/cloud testing tools are used by the mobile apps developers. As can be observed, over half of the developers (58%) never used crowd-testing platforms and (14%) they rarely used them. Meanwhile, (26%) of the participants said they sometimes used platforms while only (2%) said they often used them. The following are some of the most frequently used crowd-testing platforms as pointed out by participants who utilized crowd-testing: uTest[6], MyCrowd QA[7], 99tests[8], Mob4Hire[9], BugFinders[10], and TestIO.[11].

The data in Table 1 also showed how often developers used automated tools or cloud testing services. It is clear that a good proportion (34%) of developers sometimes used them while only (6%) they always used them. Meanwhile, (24%) replied that they never used while (22%) rarely used. A small minority (14%) of participants responded that they often used them. The data in the table also clarified that (34%) of developers always used testing companies. A similar proportion of developers answered that they used sometimes. Only (6%) of developers replied they rarely used them while (4%) they never used.

(Q.2.) Do you use any of the following crowdsourced programming websites (Stack Overflow, GitHub, and Stack Exchange) to look for solutions to programming problems that you face?

Answer: As we can see from Table 2, more than half of the participants indicated that they always use Stack Overflow for searching for programming issues and their solutions. It is interesting that none of them said they do not. While (26%) indicated that they very much use it and (12%) moderately. Only (8%) of the participants said they use it somewhat. GitHub is another crowdsourced programming platform that used slightly less than Stack Overflow as indicated by participants. (34%) of participants mentioned that they use GitHub all the time while around (24%) use it very much. It is surprising that none of the participants said they do not use GitHub while (26%) moderately used. (16%) of participants said they use it somewhat. The least popular website among the developers was Stack Exchange as (18%) of them said they use it all the time and a small minority (8%) indicated that they use it very much. (32%) of participants mentioned that they never use it.

[6]https://www.utest.com/
[7]https://https://mycrowd.com/
[8]https://99tests.com/
[9]http://www.mob4hire.com
[10]https://www.bugfinders.com/
[11]https://test.io/

3.2 Developers' expectations of the public crowd-testing

(Q.3.) Do you think you will be able to find enough a critical mass of testers if you used the public crowdtesting?

Answer: Interestingly, none of the participants answered a definite 'No' to this question. While (57%) of them believed that this was possible, (43%) were not quite sure and answered 'Maybe'. Upon being asked to the participants who answered with yes or maybe, what are the expected benefits that they could obtain "when dealing with the public crowd testers?" Most of the participants have shared a positive outlook on the use of this method.

From the outlooks, we observed that most of the participants agreed that the distribution of tests to the public and involving the public testers with different levels of experience and from different backgrounds and environments, would help to cover more mobile devices as well as discover more issues faster than traditional crowd-testing methods. The majority of participants also highlighted that this method will provide more results and give better feedback more quickly, which may reduce the time needed to finish the testing process. A few numbers of participants indicated that this way of testing will also enable the study of more human behaviors due to each crowd tester behaving in a particular way. Very few participants indicated that testing by the public crowd testers can give more useful results than individual testers and it would lead to improving the developers' skills based on different collected feedback. Two participants believed that the public crowd-testing will help in conducting testing many times in the early stages of apps development life-cycle. One respondent stated that "Because the test would be opened to whole testers in the world, this gives more variety in testing scenarios, techniques, and different tools in testing, which reduces the need for testing apps by companies".

3.3 The requirements of the public crowd-testing

(Q.4.) What are the first typical starting elements to search for a solution to any issue you may face during mobile app development?

Answer: The responses demonstrated that Mobile device model represents the highest percentage (45%), followed by OS Version (29%), mobile platforms (e.g., iOS or Android) (16%), and brand/manufacturers (10%). It is clear that the model of devices is the first and most important element that the developers look for during development and testing processes of the mobile app. Whereas, the brand is the least important element for them to search for.

(Q.5.) How would you prefer to post or define your problem?

Answer: (74%) of participants prefer to use title and general description similar to Stack Overflow for defining their tasks and problems. While (26%) of the participants found the structure form (divided into sections e.g., payment method) more suitable for them.

(Q.6.) The direct interaction between developers and crowd testers is important to perform an effective crowdtesting process on a large-scale:

Answer: None of the participant developers indicated that this is "Not important or Slightly important", while (70%) agreed that this was very important. (20%) of them answered that this is important,

Table 1: The proportional use of mobile apps testing methods from the participants' perspective

	Never	Rarely	Sometimes	Often	Always
Crowd-testing Platforms	58%	14%	26%	2%	0%
Automated/Cloud Testing Tools	24%	22%	34%	14%	6%
Testing Company	4%	6%	34%	22%	34%

Table 2: The proportional use of the three crowdsourced programming websites Stack Overflow, GitHub and Stack Exchange from the participants' perspective.

	Never	Somewhat	Moderately	Very much	Always
Stack Overflow	0%	8%	12%	26%	56%
GitHub	0%	16%	26%	24%	34%
Stack Exchange	32%	22%	20%	8%	18%

and (10%) fairly important. Overall, it can be said that all of the surveyed developers agreed on the importance of the direct interaction between testers and developers rather than the need for middleman crowd manager or leader during the testing process.

(Q.7.) How much you think the easy/difficulty design of the results reporting template will affect the encouragement and participation of the crowd testers?

Answer: The responses demonstrated that (55%) of participants have strongly agreed and (31%) agreed that this will significantly affect while (4%) disagreed that this would have any effect. The remaining (10%) responded neutrally that this may have negative impacts on the participation of the public crowd testers.

3.4 Measuring Levels of Trust

(Q.8.) Would you trust the public and anonymous crowd testers from any country in the world to perform your testing tasks?

Answer: (68%) of participant developers trust to perform their tests by crowd testers from any part of the world while (32%) did not. About (69%) of the developers who replied negatively to this question justified " why?" Their responses covered a range of reasons. Most participants mentioned that their reason for the lack of trust is linked to the security of data (lack of identity and guarantees), as one of the developers said: "an idea could easily be stolen and published before finishing the app development process". Other participants' reasoning is linked to the level of education and technological development of some countries over the world. While a few numbers of participants indicated that the main reason for not trusting is that the participation of the public crowd testers can only be to get money. Only one developer had a somewhat neutral response, which mentioned that trusting the public crowd testers from different countries depends fundamentally on the specific region of the world that the mobile app targeted; in that case, the developer does not trust other testers from the particular region.

For the (68%) of the participants who gave positive responses, a further question "how much they trust the information provided by the public crowd testers?" It was surprising that the participants' responses were (50%) moderate and (50%) very much, while none of them answered with a little.

(Q.9.) Consider you are working with the public and anonymous crowd testers, how do you will know that they have really performed the test correctly?

Answer: Most of the participants mentioned that they could know that through the detailed description of the testing plans and test cases or testing scenarios that are reported by crowd testers. Almost a half of participants mentioned that repeating the testing steps implemented by the public crowd testers to reproduce the same issues can also be a possible solution. A minority of the participants indicated that they have to integrate a tracking tool to capture and record the testing results, process, and activities carried out by crowd testers. Three participants believed that the backgrounds or practical experiences of the developers may help in that situation. Two participants considered that making an issue or more intentionally in the app can be one of the best solutions for measuring if the crowd testers actually executed the test. Interestingly, one participant pointed out that asking one or two precise questions at the last stage of the testing process is an accurate method of measuring if the crowd testers actually conducted the testing process.

(Q.10.) How do you want them to prove that the results they provided are correct?

Answer: The participants' responses gave a set of different possible solutions. The provided solutions included the provision of images, video recordings and textual reporting and automatic reports (e.g., log file). Most of the participants given the solution required a textual explanation as evidence of the accuracy of results. While a good proportion of participants mentioned the screenshots of issues as evidence. Another group of participants required video recording as a solution. The remaining possible solution as indicated by a minority of participants was to the importance of automatic reports to prove that the results are correct.

3.5 Desired Features and Outcomes

(Q.11.) What are the elements that will attract you to work with the public crowd testers to execute your tests and would make you leave working with testing companies?

Answer: Only (90%) of the participants responded to this question. The responses covered a broad range of views that are organized into three six categories as follow:

1) Better quality of results: Most of the participant agreed that

obtaining fast and accurate results could be the main reason to deal with the public crowd testers.

2) Lower cost: Another group of the participants mentioned that the lower payment cost will be another reason for that.

3) Flexibility: Two participants referred to the flexibility for repeating the test more than once and any time during the development process as another reason that would motivate them to work with the public crowd testers;

4) Diversity: This is related to the need to cover a wide variety of environments, cultures, processes, and steps for testing mobile apps. From participants' responses Four sub-categories were identified: a *Test diversity:* The majority of participants mentioned the need to use a diverse set of real-world testing scenarios, test cases, techniques and steps for testing mobile apps; (1) *Hardware resources diversity:* Other participants pointed out that the ability to cover a large variety of mobile devices models and OS versions is another reason to use with the public crowd-testing method; (2) *Human knowledge diversity:* Two participants considered the accessibility of various levels of crowd testers' experiences is also an important factor. As one participant said that " the ability to find crowd testers adapted to many different functions or activities is really important"; (3) *Human behavioral diversity:* Only one participant had considered the possibility of covering a large diversity of end-users behaviors as an important feature;

5) Organization and User-friendliness: A small number of participants expressed that good organization of the testing processes, issues reporting mechanisms, and supporting free automated testing tools were also considered important features that motivate them to leave working with testing companies and start working with the public crowd testers. As one participant said," using TFS tools to list the bugs to make the developer aware of the complete problem is important". Three participants mentioned the importance of the ease of use of the crowd-testing platform. Two participants indicated the need for a good communication method between the public crowd testers and developers;

6) Other responses: There were two interesting responses; the first response was: "The patience in repeating questions and frequent communication without increasing service charges or feeling bored is considered one of the significant reasons to work with crowd testers". The second response was a neutral "choosing the testing method between either companies or crowdsourcing depends on the type of the app itself whether it's allowed to be tested by the crowd"

(Q.12.) What are the incentives that you would be willing to offer to the public crowd testers to motivate them to participate more?

Answer: The participants' responses covered vast ideas of the possible incentives, the most significant proportions among these incentives was gift cards or vouchers (e.g., restaurants, shopping, buying electronic devices, traveling, Amazon, etc.) (56%), providing money (41%), provision of free apps or allowing the use of a paid version of the app (38%), and points and reputation (33%). While invitations to attend events, or training courses (8%), providing certifications (4%) and providing more knowledge related to testing scenarios and activities (2%) are also possible incentives, but with a little lower proportion from the first group of incentives.

(Q.13.) What are your desires from the public crowd-testing?

Answer: Only 96% of participants responded to this question. Their responses showed a broad set of desires. The most common desire was the ability to perform more realistic tests quickly without artificial environments and to find more issues in a short time. This demonstrates that the time factor is critical for most of the participants. Some participants mentioned that they hope the use of the public crowd-testing can enhance testers' and developers' skills as well as mobile app development domain by providing more testing information and knowledge, storing the performed testing scenarios and cases by the public crowd testers for later use. A few participants have indicated that the ability to distribute a test on a large-scale to cover a variety of devices and operating system (OS) versions is one of the critical hopeful outcomes from the public crowdsourced testing. Another group of participants showed their strong desire to obtain good testing reports including all possible issues, exceptions or non-logic operations; One participant said that "provide a secure way to test apps with protection for identity especially for app ideas" is a significant factor in using crowdsourced testing. Another participant said: "I hope that testers have a good technical /programming background as this may lead us to perform Gray-box Testing which is better than Black-box testing". The last desire extracted from the responses was the possibility of providing a dashboard to display a good sampling of data and metrics.

3.6 Required Testing Information

(Q.14.) What are the four important parts of the information that developers must provide it to crowd testers as part of task defining stage?

Answer: Only (88%) of the participants responded to this question, (5%) they just indicated that the testing requirements are important without any explanation. While the remaining (83%) of them provided interesting responses identified as follow:

1) Functional Behaviors: Most of the participants mentioned that the functional requirements of apps or its component and expected behaviors (output) is significant information and must be defined when announcing any testing task.

2) Mobile device information: They are related to the need to provide the mobile platform, model, and OS versions details that need to be tested against the app as mentioned by most of the participants;

3) Timing information: A small number of participants referred to the importance of providing an estimation of time needed for each test cycle and the deadline for completely submitting the test reports and obtaining fast results;

4) App information: Other participants indicated that type of app and URL is necessary to provide, as there are many apps recently launched and all may have the same name. Only one participant said: "logo or image of the app is significantly important for crowd testers to know which app they need to test";

5) Test information: A good proportion of participants indicated the need for a full description of the whole app's and test's steps that crowd testers need to follow. In contrast, very few participants indicated the importance of providing testing scenarios or test cases by developers sometimes rather than always created by crowd testers which can help beginner crowd testers to perform an

accurate test;
6) Issues solved: The minority of participants highlighted the importance of clarifying the errors discovered and solved previously, and the parts of the app that may be influenced by the performed amendments as important information that needs to be presented at defining testing tasks;
7) App development information: Only two participants mentioned the importance of providing the source code of the app when defining a specific type of testing tasks (if needed) to understand if the issues belong to the mobile device characteristics or in the code itself;
8) Users characteristics: Very few participants indicated that target users and their characteristics (e.g. location, language, age, working domain, etc.) is also important information for defining the task.

(Q.15.) What are the four important parts of the information that testers should provide in their testing reports?
Answer: (88%) of participants answered this question. (7%) of participants did not provide clear information while (81%) of them provided enlightening responses identified as follow:
1) Tester information: A small minority of participants mentioned that due to dealing with the public and anonymous crowd testers, personal information (including name and contact information) and geographical information would be very helpful if included in submitted reports;
2) Testing environment: The details of mobile devices used in testing (platform, model, OS version) and its characteristics is also important and must be included in the submitted reports as indicated by the majority of participants;
3) Execution information: This relates to the need to submit information about the testing process that was performed: test cases or scenarios used (84%), a clear description of the steps that crowd testers followed (69%), error messages (37%), which enhance the quality of the report. Only two participants mentioned the importance of providing the number of test repetitions and time taken for each test cycle;
4) Issues information: Most of the participants highlighted the importance of receiving a clear description of issues within submitted reports, including issue id, issue name, category or type of issue, a priority of issues, severity, and actual results. A minority of them (22%) mentioned that videos or screenshots of issues are also important and must be included in the submitted testing reports;
5) Supplementary information: (25%) of participants stated that receiving additional information such as solutions or suggestions for solving issues, or expected causes of issues within submitted reports would be important.

4 MAIN FINDING AND DISCUSSION

We draw some initial conclusions to address the two selected research questions discussed in Section 1.

RQ1: *To what extent mobile app developers are keen to work with the public and anonymous crowd testers with various levels of experience to perform their testing tasks?*

The responses to (Q.1.) which are presented in Table 1 were not expected. More than half of the participants said they have never used crowd-testing platforms while (90%) had used testing companies before. Indeed, majority of the large and common testing companies such as Clarity, uTest, Telcom Italia, Pass Brains, and Bug Finders prefer to use crowd-testing in addition to their own testing teams. This provides them which a much more in-depth testing process as mentioned in [6]. This high percentage (58%) of developers who have never used the crowd-testing suggests a lack of their knowledge about the method used by the companies and their indirect use of a crowd-testing platform. This means that most likely there are some developers do not know that the companies and their websites they have used are really crowd-testing based companies.

The responses to (Q.2.) proved the participants' lack of knowledge relating to crowdsourcing or indirect use of the crowdsourcing method in general. It can be inferred from Table 2 that most likely all the participants use programming websites such as Stack Overflow, GitHub, and Stack Exchange. The data also shows that Stack Overflow is the most commonly-used platform as clarified by (56%) of participants, followed by GitHub with (34%). It is surprising to see that (34%) of developers use always GitHub to address coding issues. Since these websites are the public crowdsourcing platforms or repository for programming/coding, which deal with the public and anonymous crowd programmers [15]. Furthermore, none of the participants indicated that they never use them, this demonstrates some indication of how much mobile apps developers accept the idea of working with anonymous crowd workers, not only for programming but also for testing.

Despite the lack of knowledge, we noted that all participants provided a positive outlook in relation to finding sufficient critical mass when dealing with the public crowd testers (answer to (Q.3.)). They provided their point of views about future expectations or the benefits that could be produced when dealing with the public crowd testers (answer to (Q.3.)) as well as their desires (answer to (Q.3.)), which were grouped in six main categories: (1) Wider distribution of the test; (2) Reduce Consumed Testing Time; (3) Wider understanding of Issues; (4) Increases in Knowledge and Experience; (5) Transfer knowledge from the industrial side to the academic side. This implies an increased interconnection and cooperation of developers and testers between industrial and academic. (6) Enhanced Social networking. This means that most of the developers and testers will have a chance to recognize each other and this can increase more social networking among developers and testers of mobile apps and helps in getting more insight and knowledge.

Although significant numbers (90%) of participants were willing to move and work with the public crowd testers and stressed the need to include the fundamental testing features on the public crowd-testing platform (answer to (Q.11.)). On the contrary, (32%) of participants still had concerns and hesitations about work with the public crowd testers. These concerns included a set of reasons (answer to (Q.8.)), while the highest percentage was linked to the security of data (lack of identity and guarantees) followed by the weak level of education and technological development in some countries over the world, as well as, random execution of the test by the public crowd testers for making money only. Whereas, (68%) of participants said that they trusted very much or extremely the crowd testers from any part of the world very much (answer to

(Q.8.)). Furthermore, these participants presented many possible solutions that enable them to trust the public and anonymous crowd testers (answer to (Q.9.)). The suggested solutions can use the data that crowd testers have submitted in the reports, or some other methods, such as a deliberate mistake or deep questions related to one of the test steps. In addition to the availability of the several other possible methods, which serve as a guide for the public crowd testers to prove the validity of their results (answer to (Q.10.)). The combination of a very high percentage (90%) of willing participants with moving and the possible solutions suggested and desires point out to the positive outlooks on the idea of testing mobile apps by the public and anonymous crowd testers around the world. In conclusion, there are good indications and clear evidence of how much mobile apps developers acceptance to the concept of the public crowd-testing and their willingness to engage with the public and anonymous crowd testers.

RQ2: *What are the key factors required to build a large scale crowd-testing platform?*

The most important topics discussed in this survey are the diversity in searching criteria, definition style of the task and issue, direct interaction, difficulty level of the system, the variety of incentives, and the necessary information to be provided by both testers and developers are the most important topics discussed in this survey. Based on answers to (Q.4.), it was found that different groups of participants seeking their testing issues via different searching elements of the mobile device. It is interesting that the higher percentage was for the phone model. It could be because of that the model can indicate to the brand and platform at the same time. Therefore, building a searching mechanism with diversity in searching criteria will be very important to developers that are looking for different solutions for a wide variety of phone models. This is due to the fact that the diversity criteria in the searching mechanism may reduce the time required for searching for the complex problems at the testing and programming stage. Additionally providing a large set of solutions that do not immediately appear. Furthermore, it provides the ability to easily reach target solutions for which the developers are looking.

While the general description was considered the most preferred method for the developers to define their testing issue due to the flexibility in describing the testing issues well enough to be understood by the crowd testers or other developers (answer to (Q.5.)). A small group of participants considered the reduction of time needed for typing is more important and so they preferred to use a structured form (divided into sections e.g., payment method). Whereas, the majority of the surveyed developers agreed on the importance of direct interaction between testers and developers (answer to (Q.6.)). This is probably due to the fact that they will deal with the public and anonymous testers from different societies and geographical locations over the world. Consequently, the developers may need to: (1) Recognize who performed the tests; (2) Understand more information about the results including bugs or defects [10]; (3) Understand their expressions and explanations for testing issues are sometimes due to the differences in cultures and spoken languages. From the responses (answer to (Q.6.)), we extracted that direct interaction between the public crowd testers and developers will lead to faster tests compared to the existing crowd-testing methods, as highlighted out by some crowd testers in [6] that the delays could happen because of test managers or leaders. In addition, building an online space for developers and the public testers to share and discuss the results of testing will lead to a seamless environment and a better understanding of the testing results (including causes of the problems occurred) [11]. Thus, this would enhance and facilitate the development process and deliver better-accepted apps of high quality. As the public testing method deals with crowd testers from the public who represent the end-users. This will be consistent with that stated by [1] that direct communication and collaboration between software developers and end-users during the development process is important in order to develop better-accepted software. Besides, the difficulty of the issues reporting system has been indicated as one of the major obstacles face the public crowd testers with various level of experience, which can negatively affect the enthusiasm of the public crowd testers to participate and use the crowd-testing platform as completely agreed by (86%) of the participants in (Q.7.). This is due to the different behaviours and interactions of crowd testers with the system. This is according to [13] the difficult interfaces of the system can reduce the use of the system by target users and at the same time shows their dissatisfaction to continue using the system.

The methods of motivating crowd testers are considered an important factor in crowd-testing systems. Although, the different types of incentives provided in (Q.12.), it is interestingly to notice that a group of participants think that the testing is not the main serious job, it is only simple cooperation. Therefore they want to offer not to pay to testers and preferred to provide goods instead. Unfortunately, this is not fair, do you think that tester who need to earn a living are not interesting in having free access to the app, gift card, ability to work in other apps, etc. Developers should be mindful of cost with regards to gig-economy and globalization and testers need to pay according to the standard living in their countries. Testing is not easy and required a lot of experience and many qualification exams to be professional similar to any other job. This raises a question for future researchers in the importance of studying the nature of the testing process depending on what testers do. All types of incentives listed provide (answer to (Q.12.)) could be an additional solutions to motivate crowd testers further when developing a new crowd-testing platform, but not as the main reward.

In the end, it can be said that these responses brought attention to the necessity of providing a way to enable developers to provide the incentives, which they see fit for encouraging more crowd testers to participate in the published test cases or scenarios. To provide suitable incentives for crowd testers and to facilitate the public crowd-testing activities, sufficient information needs to be introduced to crowd testers to execute the test correctly and so submit an accurate report. According to [6] a good number of the crowd testers indicated that the amount of information provided to them by the testing companies is not sufficient to carry out the test well. They only receive information about the app itself such as test scenarios, information about specific inputs they have to test, and occasionally information about the devices that need to be tested. This study addressed this issue by providing a broader view of the information that the developers need to provide to the public crowd testers as well as the information that must be received from them

within testing reports. (1) For the developers, to ensure that the testing task is defined in a clear and structured way to crowd testers, and to understand more information about the results including bugs or defects [11], all pieces of information that are reported in (Q.14.) must be included. (2) For the public crowd testers, to understand more about the testing tasks that they need to conduct, and to ensure that the testing report has completely included all important details of a task's results and that the report will not be rejected, all pieces of information that were reported in (Q.15.) must be included. Such information will help developers to clearly understand the issues. we see that all these pieces of information presented in (Q.15.) are the key factors for measuring the quality of testing reports collected by crowd testers and so provide suitable incentives. Ultimately, we see that the diversity in searching criteria, free space to define the issue, direct interaction between testers and developers, easy-use of the system interfaces, the variety of incentives, and the necessary information provided by both testers and developers, as recommendations for the participants to accept to work with the public and anonymous crowd testers.

5 RECOMMENDATIONS

Based on the results discussed above, we provide the following recommendations:

(1) The need for a new crowd-testing platform to serve individual developers or small development teams, and provide opportunities to the public crowd testers with different levels of experience is evident from a number of participants' responses; the need for different levels of experience (answers to Q.3 and Q.11), searching mechanism with different criteria (Q.4), direct interaction between developers and crowd testers during testing processes (Q.6), information for the apps' expected behaviors on various mobile device models and different perceptions of human behaviors and their interactions with mobile apps (Q.11). Therefore, a new practical crowd-testing platform should be designed to address all the aforementioned needs. **(2)** The direct interaction between mobile apps developers and crowd testers without the need for an intermediary leader or manager is perceived positively such as in (Q.6.); in addition, the need for online space for both developers and the public testers to share more knowledge about apps' behavior and discuss the useful testing information and results is also evident in a number of participants' answers, such as in (Q.3, Q.5, Q.12, Q.13, Q.15). Therefore, to provide evidence of the benefits that individual developers or small development teams may gain from this direct interaction and online space available, a new platform needs to be built that covers these two needs. **(3)** An empirical validation needs to be performed to evaluate such this platform including recommendations mentioned in (1) and (2), (in the case of constructed).

6 CONCLUSIONS AND FUTURE WORK

Although crowdsourcing has gained much attention among mobile app developers, much still needs to be done to change the perspectives of participants who are still concerned about the use of the public crowd-testing for mobile apps testing. This paper has presented an exploratory study investigating developers' point of views in agreeing to work with the public and anonymous crowd testers in crowd-testing processes of mobile apps. In addition, it has identified the desirable features or properties required by developers in order to use the crowd-testing platform for testing their apps. In total, we analyzed responses from 50 mobile app developers with different experience in iOS and Android from different countries around the world.

The results reported that app developers are willing to use the public crowd-testing platforms if some challenges can be addressed. Further, the study concludes that the direct interaction and development of trust between the public crowd testers and the mobile app developers is key to performing an effective testing process and to establishing a long-term working relationship between these two groups. The external validity of this study is verified as it involves participating in mobile app developers from different countries around the world. In addition, this paper discussed the results in detail and have given several recommendations that software developers need to consider when developing a new crowd-testing platform. This study has helped us to better understand the key requirements and issues that the participants concerns about in the public crowd-testing method. However, there is still a need for further research as the links used in sharing this only survey may not have reached large-scale potential participants with significant experience in this field.

REFERENCES

[1] I. Alvertis, S. Koussouris, D. Papaspyros, E. Arvanitakis, S. Mouzakitis, S. Franken, S. Kolvenbach, and W. Prinz. User involvement in software development processes. *Procedia Computer Science*, 97:73–83, 2016.

[2] I. Bayley, D. Flood, R. Harrison, and C. Martin. Mobitest: a cross-platform tool for testing mobile applications. 2012.

[3] A. Bryman and E. Bell. *Business research methods*. Oxford University Press, USA, 2015.

[4] A. Bryman, B. Burgess, et al. *Analyzing qualitative data*. Routledge, 2002.

[5] M. Denscombe. *The good research guide: for small-scale social research projects*. McGraw-Hill Education (UK), 2014.

[6] F. Guaiani and H. Muccini. Crowd and laboratory testing can they co-exist?: an exploratory study. In *Proceedings of the second international workshop on CrowdSourcing in software engineering*, pages 32–37. IEEE Press, 2015.

[7] J.-f. Huang. Appacts: Mobile app automated compatibility testing service. In *Mobile Cloud Computing, Services, and Engineering (MobileCloud), 2014 2nd IEEE International Conference on*, pages 85–90. IEEE, 2014.

[8] J.-f. Huang and Y.-z. Gong. Remote mobile test system: a mobile phone cloud for application testing. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, pages 1–4. IEEE, 2012.

[9] J. Kaasila, D. Ferreira, V. Kostakos, and T. Ojala. Testdroid: automated remote ui testing on android. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, page 28. ACM, 2012.

[10] Q. Naith and F. Ciravegna. Hybrid crowd-powered approach for compatibility testing of mobile devices and applications. In *Proceedings of the 3rd International Conference on Crowd Science and Engineering*, page 1. ACM, 2018.

[11] Q. Naith and F. Ciravegna. Mobile devices compatibility testing strategy via crowdsourcing. *International Journal of Crowd Science*, 2(3):225–246, 2018.

[12] C. M. Prathibhan, A. Malini, N. Venkatesh, and K. Sundarakantham. An automated testing framework for testing android mobile applications in the cloud. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on*, pages 1216–1219. IEEE, 2014.

[13] M. B. Rosson and J. M. Carroll. *Usability engineering: scenario-based development of human-computer interaction*. Morgan Kaufmann, 2002.

[14] O. Starov. Cloud platform for research crowdsourcing in mobile testing. 2013.

[15] B. Vasilescu, V. Filkov, and A. Serebrenik. Stackoverflow and github: Associations between software development and crowdsourced knowledge. In *Social computing (SocialCom), 2013 international conference on*, pages 188–195. IEEE, 2013.