

Lessons Learned from Conducting Industry Surveys in Software Testing

Tanjila Kanij
Swinburne University of Technology
Hawthorn, Victoria, Australia
tkanij@swin.edu.au

Robert Merkel
Monash University
Clayton, Victoria, Australia
robert.merkel@monash.edu

John Grundy
Swinburne University of Technology
Hawthorn, Victoria, Australia
jgrundy@swin.edu.au

Abstract—We have conducted a number of industry studies looking into aspects of software testing. These include a survey of practitioner view on software tester performance indicators and team formation issues; detailed worklog analysis of representative tasks conducted by software testers; feedback on a proposed approach to software tester performance evaluation; and collecting indicative personality factors of software testers vs other software developers. These have proved to be challenging exercises. Key issues to overcome include reaching appropriate testers in industry, gaining management and company consent, structuring surveys and questionnaires to balance time vs detail, and gaining sufficient response rate and response quality to be useful.

Index Terms—Survey, Questionnaire, Sampling, Response rate, Software testing

I. INTRODUCTION

Surveying is an empirical research method applied to collecting self-reported data from participants. A successful survey enables researchers to collect data from a large group of participants in a short time. However, with the technological focus of most software engineering research, surveys are under-represented in the software engineering literature [4]. This represents both an opportunity and a barrier to survey research in the field. While there are opportunities to provide new perspectives on many aspects of software engineering, the lack of published studies means that models for successfully conducting survey research in software engineering are limited. Researchers attempting to conduct studies may struggle to find out what methodologies lead to publishable data.

In this article, we discuss our experiences conducting four quite different Internet-based surveys on topics relating a particular area of software engineering: software testing. We focused on software testing due to the authors' interests and the lack of empirical evidence around software tester personality impact, team formation, tester performance assessment and tester practices. We discuss the design of these empirical data collection exercises; their relative success and limitations; and describe some key lessons we have learned from our experiences. We hope that this will guide future researchers in designing and running their own industry-based surveys.

II. SURVEYS

We describe the four industry surveys we have conducted. All were Internet-based anonymous web surveys, and hence

no face to face interaction between the researchers and participants occurred. A custom web site was developed for all the surveys and for each conducted a small pilot survey first.

A. Survey 1: Factors Influencing Software Tester Performance

This preliminary survey was designed to collect a range of opinions of software testers and related IT professionals on various factors influencing the performance of software testers [6], [5]. The approximate time to complete the survey for English speakers was 25-30 minutes.

The consent information statement was signed electronically, such that the homepage of the survey website contained the consent information statements with an "I agree" check box. Participants were required to check the check box in order to participate. No reward was offered for participation.

We used cluster sampling [2] and invited clusters of participants who were associated with software testing-related mailing lists and groups. To select the sample for the survey we have searched for software testing related LinkedIn and Yahoo! groups and listed 21 Yahoo! and 29 LinkedIn groups using purposive sampling [2]. Among those, moderators of 12 LinkedIn and 12 Yahoo! groups approved us to send invitation email to the group making the group response rate 41.4% and 57.1% for LinkedIn and Yahoo! groups, respectively. An invitation email containing the purpose of the survey, the expected outcome, benefits of the survey, the ethics approval details and the link to the online survey form was sent to those groups.

The survey was conducted during January-March 2010. In total, 104 participants completed the survey. The individual response rate is unknown and can not be measured since a participant can be member of more than one group. Nor is it known how many group members actually read the group emails.

B. Survey 2: Worklog Collection of Software Testers

To prepare a refined set of unit job responsibilities of software testers, we conducted a detailed survey collecting worklogs from a small number of testers. In broad terms, a worklog is a collection of units of work performed, tracking time spent in performing the different parts of the work. This was a time-consuming and long-range survey - participants were requested to log their work for at least two weeks.

Participation in the survey was voluntary. However, a risk in this survey was that the participants' employers may react negatively to their participation e.g. revealing company specific sensitive information. To address this issue, we asked for a signed consent information statement from the manager or supervisor or employer of the participant to authorize the participation.

As we needed detailed, high quality data and small datasets were useful, the participants for this survey were recruited from the personal contacts of the researchers using snowball sampling [2]. Our initial invitation email, which included the purpose, aim and other details of the survey, was sent to 16 contacts. Most of them who were managers or academics were not requested to directly participate but were instead requested to nominate participants. They nominated 17 more participants 6 of them finally participated in the survey. The individual response rate was 41.2%. Invitations were sent from November 2011 to July 2012 and participation took place during January 2012 to May 2012.

C. Survey 3: Testing Manager Opinion on a New Performance Appraisal Form for Software Testers

This survey was designed to validate a performance appraisal form for software testers proposed by the researchers and to collect opinions of participants on the practice of performance appraisal of software testers. The overall survey was estimated to take around 45-60 minutes to complete.

This survey also requested that supervisor or employer consent be obtained, along with participant consent. In this survey the consents were collected by email and the participants were given a unique code to log in to the site to participate.

We used both cluster and snowball sampling in this survey. Permission to send invitation email was requested from the groups only that had approved us in the first survey, described in Section II-A. 3 LinkedIn and 4 Yahoo! groups permitted, making the group response rate 25% and 33.3%, respectively. An invitation email including the purpose of the survey, the expected outcome, benefits of the survey and the ethics approval details was sent to the groups. Of 10 group members that showed interest to participate, 3 finally completed the survey. In the snowball sampling process initial invitation was sent to 5 contacts. Unfortunately, we found no participants from this process.

Due to the poor response rate, we modified the survey design to make it more lightweight. The new survey took around 15-20 minutes to complete. We used snowball sampling for the new light weight survey and obtained 3 participants so far. Data collection is still in progress. The initial survey began in August 2012.

D. Survey 4: Survey of Personality of Software Testers and Related IT Professionals

This survey collected the personality profiles, as measured by the 50-item IPIP NEO personality assessment test [3], of software testers and other people involved in the software development process to find significant differences in personality traits, if there was any. It took around 15-20 minutes for

participants to complete the survey. The consent information statement was signed electronically.

We used cluster and purposive sampling. In cluster sampling we requested permission from the 12 LinkedIn and 12 Yahoo! groups, that gave us permission in first survey. 5 Yahoo! and 3 LinkedIn groups approved us. As part of the process we attended and invited the participants of a software testing related industry conference in Australia and posted to an industry conference email list related to software development. We also tweeted on the Twitter feed of the developer conference with the help of the organizers.

We received 180 responses from November 2012 to January 2013. The data collection for this survey is still open. The major differences to previous surveys were the use of professional software tester conference organizers to help disseminate the call for participation and the use of a prize as an inducement.

The details of all four surveys discussed above are summarized in Table I.

III. LESSONS LEARNED FROM OUR EXPERIENCES

In conducting the four web-based surveys described in Section II, we came across several factors that we believe had an effect on our response rates and hence quantity and quality of data, which we discuss below.

A. Participant Recruitment

All the surveys were conducted as part of a PhD thesis. As such, the time to design, to get approval from the University human research ethics committee, to collect data, and to analyze the results were limited. Hence we searched for a centralized body of software testing professionals where we could send invitations to participate in our survey in order to be able to recruit large numbers of participants in a relatively short time. Unfortunately, there is no such centralized board of software testing professionals.

The relevant email lists provided by LinkedIn and Yahoo! were helpful in this regard. We believe this process of recruitment helped us to get a good number of participants in the short available time, for two of our surveys.

B. Low Response Rate

A crucial factor in survey research is the response rate, calculated as the ratio of the number of participants who completed the survey to the number of participants who we sent invitations to. A good survey aims to keep the non responses to a minimum. In the surveys (except the second survey) we conducted, it was impossible to calculate an accurate response rate, since most of the invitations were sent to the groups and the number of group members who actively read emails cannot be obtained. However, in comparison to the number of people we sent invitations to, the number of responses we received for the these surveys is very poor. A common reason behind this low response can be the fact that IT professionals may be reluctant to participate in research studies like a survey! However, this assumption has not been supported by any specific experimental evidence. Additionally, we found that academic paper referees are generally very keen to see specific response rates calculated and presented

TABLE I
SURVEYS

Survey	Intended participants	Questionnaire length	Sampling	Invitation	Consent collection	Number of responses	Period of data collection
1	Software testers	25-30 minutes	Cluster	Large email sent to groups	Participant consent signed electronically	104	January-March 2010
2	Software testers	15 days	Snowball	a. Large email sent to groups; b. Large email sent to personal contacts	a. Participant consent signed electronically; b. Manager consent collected via email	6	January 2012 to May 2012
3	Software development project managers	45-60 minutes	Cluster and snowball	a. Large email sent to groups; b. Large email sent to personal contacts	Participant and manager consent collected via email	3	August 2012
4	Software development related stack holders	15-20 minutes	Cluster and purposive	Small text sent to groups, mailing lists, twitter, conference attendees	Participant consent signed electronically	180	November 2012 to January 2013

in papers using survey techniques. Being unable to calculate these due to recruitment techniques via boards and lists is not always well-received.

In conducting surveys, along with the survey response rate, the *representativeness* of the sample is also very important. According to Cook et al. [1], the assurance that the chosen sample represents the population is particularly important when a sample of convenience is chosen. In the surveys we conducted, the captured demographic information of the participants increased our confidence on the representativeness of the sample.

C. Invitation Email

We believe the nature of the invitation played an important role in recruiting participants for our surveys. We sent large invitation emails to the participants in the first three surveys. However, in the fourth survey, we sent a much shorter, catchy slogan and obtained far more participants compared to our other surveys.

D. Consent Collection

According to standard human research ethics protocols, participants are generally required to indicate their informed consent by signing Consent Information Statement (CIS) before participating. In our first, second and fourth surveys, participants could sign CIS electronically. However, in the third survey we requested participants to send a signed CIS to the researchers via email. We think the necessity of making two separate efforts (sending email and participating online) for participating might have negatively influenced participation.

Getting consent from the supervisor or manager or employer added an extra work burden on the participants. The reaction on getting manager consent was illustrated with the comment of one interested participant- “...getting the consent from employer or manager is hard job and would prefer to avoid it. But if you need my opinion do send me the link, I can fill it up and can happily provide the information.”

E. Questionnaire Length

In the series of surveys we conducted, we found more respondents when the stated length of the survey was around

20-30 minutes. For the longer duration surveys, such as the worklog (at least 15 days) collection, we got very few participants.

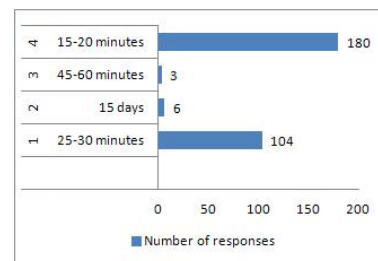


Fig. 1. Number of responses compared to questionnaire length

Figure 1 shows the number of responses compared to the length of the survey. The length of the second survey cannot be compared directly to the others since the time needed to complete a worklog might be only a few minutes (depending on the work load of the participants). However, participants were required to spend that few minutes each working day for at least 15 days. As such the exact time needed to complete the survey is unknown. Also the responses to the third survey are not conclusive, as due to the low number of responses, we stopped collecting data for the survey and modified the survey to make it less time consuming. Data collection for the new light-weight survey is still in progress.

F. Nature of Participation

We obtained a much higher number of participants in our first and fourth surveys. Compared to these two surveys the responses we obtained for our second and third surveys were very low (and very disappointing). One factor behind this might be the nature of participation requested in the surveys. The first and fourth survey was prepared with straight forward questionnaire. On the other hand, the second and third survey requested participation such that participants had to recall information from memory which took more time to form

responses to questions. This may be a significant causative factor in the low response in those two surveys.

G. Motivation

We found that some IT professionals are highly motivated to participate in the surveys we conducted. The enthusiastic responses of the participants in the first survey encouraged us to design this series of surveys. In order to store the lengthy textual responses to open-ended questions in the first survey, we had to increase the size of our database field during data collection. Along with filling out the surveys, some members of Yahoo! groups, carried out discussion in the groups on the survey topic. We received the following response from one member to the group invitation indicating their interest “*Fascinating! It’s an interesting study. I’d love to hear any conclusions you draw from this*”. Members also informed their colleagues about the surveys without being explicitly asked to do so.

Some participants, on the other hand, participated for personal self-interest. For example, one participant was interested in postgraduate research and asked for help from the student researcher. The following comment is indicative: “*I am interested about SQA (Software Quality Assurance) and want to (do) research. I read your mail but i am not fully understand(ing) about the mail purpose. I want to participate in the research study about SQA (Software Quality Assurance). Can you help me?*”.

H. Analysis of Data

The surveys were all originally conceived and designed to allow detailed quantitative analysis of the collected data. However, due to the limited responses received to our second and third studies, we were limited to performing qualitative analysis of the data instead. The qualitative analysis was very time-consuming although we did obtain useful and interesting results. However, had we known that we would need to use qualitative analysis, we would have asked some questions differently and some different questions.

I. Process of Ethics Approval

In the process of obtaining human research ethics committee approval, we had to submit a filled out protocol form accompanied by templates of the informed consent letters, website design documents and so on. The committee typically took a few weeks to review the application. We found the opinions of the committee members helped to improve the quality of the surveys. Delays were, however, compounded when it was necessary to make changes, often requiring another round of ethical review. Our new light-weight survey on software tester performance appraisal was an example of this.

J. Data Security

In web based surveys, data usually is temporarily stored on a web server. The security of the data is a concern when those reside on an external server. While the data we collected were anonymous, even if the data would have been compromised,

those would not be of any personal interest to others. In case of highly sensitive data like email addresses, we encrypted the data before storing. The university ethics committee appraisal of our applications was highly concerned to ensure very secure data management.

IV. SUMMARY AND SUGGESTIONS FOR OTHERS

In this article, we reported our experiences of conducting a series of web surveys on software testing-related topics. We found the most difficult part of the process is to convince participants. Although the group invitations helped to reach a large number of potential participants, we did not achieve a high response rate for two surveys. In the case of our third survey, we had to significantly modify the survey itself - and thus the research we were attempting to undertake - in order to get more responses!

Based on our experiences we make the following recommendations to others who want or need to use surveying of IT professionals in their research programme -

- 1) Group invitations are helpful - but moderators need convincing of the worth.
- 2) Surveys designed with straightforward questionnaires that can be answered without too much thinking are more popular than surveys requiring more time to form responses - deeper research questions may massively reduce response and quality.
- 3) Marketing is important - a short, catchy slogan is more likely to draw the attention of participants.
- 4) Simple methods for participants to indicate informed consent are helpful in reducing drop-out rates.
- 5) Surveys should be designed with different data analysis methods in mind.
- 6) A significant amount of time should be allocated to the process of human research ethics approval.
- 7) In web surveys, the method of ensuring the security of data should be given a high priority.

REFERENCES

- [1] C. Cook, F. Heath, and R. L. Thompson. A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*.
- [2] M. Denscombe. *The Good Research Guide for Small-Scale Social Research Projects*. Open University Press, Milton Keynes, UK, 2003.
- [3] L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84 – 96, 2006. Proceedings of the 2005 Meeting of the Association of Research in Personality.
- [4] A. Hfer and W. Tichy. Status of empirical research in software engineering. In V. Basili, D. Rombach, K. Schneider, B. Kitchenham, D. Pfahl, and R. Selby, editors, *Empirical Software Engineering Issues. Critical Assessment and Future Directions*, volume 4336 of *Lecture Notes in Computer Science*, pages 10–19. Springer Berlin Heidelberg, 2007.
- [5] T. Kanij, R. Merkel, and J. Grundy. A preliminary study on factors affecting software testing team performance. In *ESEM*, pages 359–362, 2011.
- [6] T. Kanij, R. Merkel, and J. Grundy. Performance assessment metrics for software testers. In *CHASE*, pages 63–65, 2012.