

## A preliminary study on factors affecting software testing team performance

Tanjila Kanij  
Swinburne University of Technology  
Hawthorn, Victoria, Australia  
Email: [tkanij@swin.edu.au](mailto:tkanij@swin.edu.au)

Robert Merkel  
Monash University  
Clayton, Victoria, Australia  
Email: [robert.merkel@benambra.org](mailto:robert.merkel@benambra.org)

John Grundy  
Swinburne University of Technology  
Hawthorn, Victoria, Australia  
Email: [jgrundy@swin.edu.au](mailto:jgrundy@swin.edu.au)

**Abstract**—With the growth of the software testing industry, many in-house testing groups and outsourcing testing companies have been established. Underlying the success of these testing groups and companies are team(s) of testers. This research investigates the importance of different factors, diversity and experience on building a successful testing team. We collected the opinions of testing practitioners on these factors via a survey. The outcome strongly indicates the relative importance of different factors and that diversity is helpful for a testing team. The results also support the importance of suitable team experience.

**Keywords**-testing team; rank; diversity; experience

### I. INTRODUCTION

Testing services are increasingly outsourced to specialised testing companies. Most large software development projects also involve the use of specialist testers. Testers often work as teams in these companies and testing groups. The performance, and to some extent the revenue, of these testing groups and companies depends on the performance of their testing teams. However, the question of how best to build an effective testing team - which may or may not be the same as finding the best individual testers - has received surprisingly little research attention to date.

In this paper, we present a preliminary survey that examines what factors might be important for building testing teams. These factors include the influence of team diversity and experience. We collected the opinions of a range of testing practitioners via a survey instrument. Our research outcomes are helpful to the testing research community by providing some initial direction for further research in this area. It is also helpful for professional testers by providing some guidelines for building successful testing teams.

### II. METHODOLOGY

We used a personal opinion survey [1] for our research. This survey was administered as one part of a two-part survey. The other part of the survey investigated the factors that influence the performance of individual testers that has been reported separately [2].

Our survey was designed according to the six steps suggested by Kitchenham and Pfleeger [1]. The step by step design of the survey is described in the following sub sections.

#### A. Setting the Objectives

The main objective of this research is to identify the most important factors that determine the performance of software testing teams. We specifically examined whether diversity is important for a testing team. If so, what diversity do participants consider the most important. We also investigated whether experience working as a team is considered important to performance.

Existing research of the influence of different factors on general team building (not specific to software testers) [3] led us to hypothesize that there are multiple factors that can influence testing team performance. The relative importance of these factors, moreover, will vary. From our previous experience, a review of the team-building literature, and the feedback of a pilot survey, we listed seven factors - "Testing performance (performance of individual members)", "Interpersonal skill", "Team playing capability", "Experience in testing", "Training/certification in testing", "Knowledge of specific problem domain" and "Compatibility with other proposed team members (if known)" - that might be important for building a successful testing team.

Similar to the influence of these different factors, different types of team diversity has also been studied in general team building research [4]. We listed four types of diversity - "Diversity of personality", "Diversity of professional background/experience", "Diversity of age" and "Diversity of communication skill" - based on our experience and review of the team-building diversity literature.

#### B. Survey Design

We used an online survey containing a self-administered questionnaire.

#### C. Development of Survey Instrument

Kitchenham and Pfleeger [1] suggest searching for relevant literature before developing a survey instrument. Unfortunately, we did not find any research on software testing team building in the literature. However, relevant research investigating the influence of different factors on the performance of IS teams was helpful. Factors like diversity of personality [5], [6], experience [7], team playing capability and communication skill were included in the survey. These factors are considered to be influential on the performance of an IS team. We speculated they may apply to testing teams as well.

1) *Questionnaire Format*: The survey questionnaire included both closed and open questions. Most closed questions used a Likert scale with five possible responses (“Completely disagree”, “Somewhat Disagree”, “Neither disagree nor agree”, “Somewhat agree” and “Completely agree”).

2) *Questionnaire Design*: The survey instrument contained six closed and two open questions split in to three main sections.

*Important Factors for Building Testing Team*: This section had three closed questions. The first question asked the participants to rank seven factors described in Section II-A according to their importance for recruiting team members. The participants could assign a rank from a range of 1 to 7 to each of the factors in ascending order (lower number implies higher rank). They could also assign the same rank to more than one factor, if they considered the factors equally important. An accompanying open question asked participants to list more factors that they think are also important. The second and third questions asked the participants whether they think all members of a testing team should be good testers and good team players. Due to limited space, responses to the second and third questions are not discussed here in detail. In short, responses to the second question were inconclusive. Respondents thought that testers should be good team players.

*Diversity of the Testing Team*: This section had two closed questions. The first question asked participants whether they think diversity in a testing team helps to improve performance with two possible responses- “Yes” and “No”. The second question asked participants what diversities they look for building a testing team. The question had four types of diversity listed as described in Section II-A and each of the diversities had 5-point Likert scale choices attached to them. Participants could also report other types of diversity (not present in our list) that they considered important, via an open question.

*Experience of the Testing Team*: This section contained one question that asked participants whether they think that a testing team performs better when they have experience working as a team.

#### D. Evaluation of Survey Instrument

We administered a pilot survey on a selected sample software engineers. In the pilot survey the questionnaire was sent to seven software engineers requesting them to fill out the survey and to comment on the questionnaire. Five of them gave us their opinion about the questionnaire. Depending on their feedback changes such as adding some additional factors and questions were carried out.

#### E. Obtaining Valid data

To select the sample for the survey we searched LinkedIn and Yahoo! groups with the keyword “software testing” and listed the first 1000 groups. We read the description of

the listed groups and selected 29 LinkedIn and 21 Yahoo! groups using purposive sampling. We sent a request to the moderators of these selected groups to allow us to invite group members to take the survey. Moderators of 12 LinkedIn and 12 Yahoo! groups approved our request. The group response rate was 41.4% for the selected LinkedIn and 57.1% for the selected Yahoo! groups. 104 participants completed the survey; however 4 of them did not give any response for testing team building questions. 100 responses are therefore reported here. 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.

#### F. Data Analysis

The Likert scale responses usually represent ordinal data [1]. The responses were converted to numerical values (5 implies “strongly agree” and 1 implies “strongly disagree”) for analysis. Kruskal-Wallis and Tukey’s Honestly Significant Difference (Tukey’s HSD) tests were used for statistical analysis.

### III. RESULTS

#### A. Demographics

The majority of the respondents (73%) were male. 76% of the respondents were between 18-40 years of age.

60% of respondents were employed by large<sup>1</sup> IT companies; with 18% of the sample employed by smaller IT companies, and a similar proportion (17%) employed by larger non-IT companies. Among the participants 77% reported their job responsibility includes testing software/programs developed by others. 45% reported that they manage testers<sup>2</sup>. Half of the respondents (50%) had more than five years of job experience.

#### B. Important Factors for Building testing team

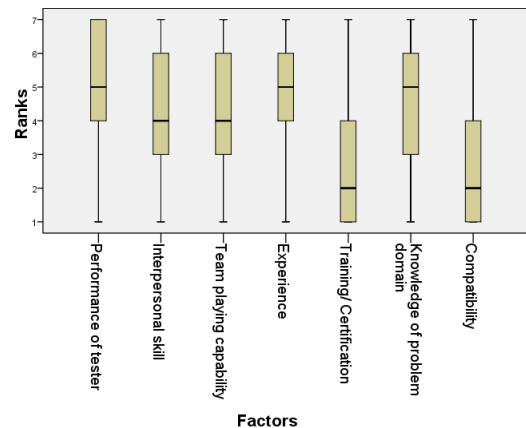


Figure 1. Rank of important factors (Question 8.1)

<sup>1</sup>Large defined as having more than 50 employees

<sup>2</sup>Respondents were able to select more than one job responsibility

Figure 1 depicts the distribution of ranking of the factors to be considered for building testing team according to their importance. To improve the clarity of the graph we have reversed the ranks so that a larger number indicates higher rank.

Some participants (25%) listed other factors that they consider while building a testing team via the accompanying open question. Among them, 3% of participants suggested that, they look for “Learning ability”, 2% suggested that they look for “Programming skill”, “Analytical skill” and “Communication skill”.

A Kruskal-Wallis test ( $p < 0.01$ ) showed there is a significant difference between the mean rank of the factors.

We have used Tukey’s HSD as Post-hoc test to find which factors’ mean ranks differ significantly. The test results indicate that “Training/certification in testing” and “Compatibility with other team members” are ranked significantly ( $p < 0.05$ ) lower than all other factors. “Experience in testing” is significantly ( $p < 0.05$ ) ranked higher than “Interpersonal skill”.

### C. Diversity of the testing team

89% of the participants agreed that diversity in a testing team helps to improve performance. 9% of participants suggested the opposite. 2% of participants did not respond to this. Figure 2 shows the distribution of responses for the four types of diversity.

In the accompanying open question, 19% of participants listed other types of diversity that can be helpful for a testing team. The responses to the open question included (in order of frequency of occurrence): “Cultural diversity” (4%), “Knowledge of diverse domain” (4%), “Gender diversity”(3%) and “Diversity of academic discipline” (3%).

A Kruskal-Wallis test indicated that there is a significant ( $p < 0.01$ ) difference in the influence of different type of diversities.

A Tukey’s HSD post-hoc test revealed that “Diversity of professional background/experience” was considered significantly ( $p < 0.05$ ) more influential than the other three type of diversities. “Diversity of personality” is significantly ( $p < 0.05$ ) more influential than “Diversity of age” and “Diversity of communication skill”. There was no significant difference between “Diversity of age” and “Diversity of communication skill”.

### D. Experience of the testing team

Figure 3 shows the level of agreement of the participants to the statement “A team performs better when they have experience working as a TEAM”. The responses indicate that 74% of participants at least somewhat agree that a team performs better when they have experience working together.

## IV. THREATS TO VALIDITY

Misinterpretation of the survey questions by respondents threatens our study’s internal validity. However, the results

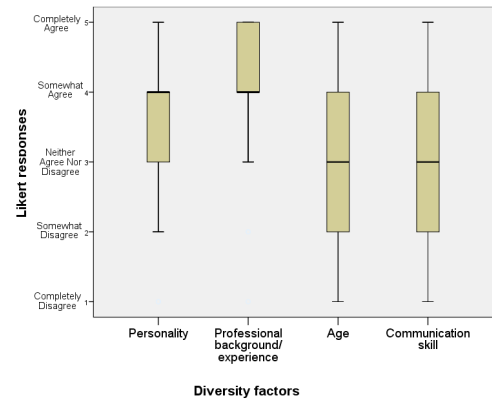


Figure 2. Level of agreement on different type of diversity (Question 8.5)

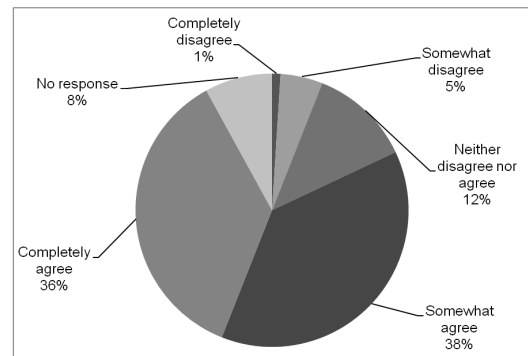


Figure 3. Level of agreement on “A test team performs better when they have experience working as a TEAM, rather than gathering experience as individuals” (Question 8.6)

show no evidence for this. Another potential threat is random or less than candid survey responses, which is a common issue in this kind of study. We see no evidence that it occurred in our data, and no particular motivation for participants to do so.

One possible threat to the external validity of the survey outcome is the representativeness of the respondents. As a voluntary survey with an unknown individual participation response rate, the survey does not represent any kind of random sample. Another threat to the external validity of our research is that it seeks only the thoughts and views of expert testers. They may reflect the “common wisdom” of the profession, however that common wisdom may well be wrong. The possibility of a response bias, based on the construction of the questionnaire, is also a plausible threat to the external validity of our research.

## V. DISCUSSION

Statistical test results reported in Section III-B indicate that there is difference in importance among the seven factors for building a testing team. However, we cannot prepare a list ranking the seven factors from the responses since the differences are not significant for all pairs of factors.

However, we can conclude that “Training/certification in testing” and “Compatibility with other team members” are considered significantly less important than all other factors. The responses to “Training/certification in testing” are not surprising, as the benefits of certification in software engineering are not supported by all [8]. However, responses to “Compatibility with other team members” are not consistent with Schutz’s hypothesis that states compatibility has a positive influence on team productivity [9]. This raises the question of why compatibility is not given much importance in a testing team. Further research is required to substantiate this result and shed light on the reasons behind it.

Our results also suggest that “Experience in testing” is significantly more important than “Interpersonal skill”. The relatively higher importance of experience is consistent with the observations of Beer and Ramler [10], reporting experience resulted in higher domain knowledge that helped testing in case of insufficient or inaccurate specifications. However, the perceived benefit of experience was observed and reported for individual testers and not for a team. The relative importance of these two factors according to our survey suggests research is needed to investigate the relation in more detail.

According to the responses to the closed question asking whether diversity of a testing team is helpful or not, we can conclude that diversity is considered helpful for a testing team to perform better since almost 90% participants selected the option “Yes”. Our next goal was to find what type of diversity participants think is helpful. The distribution of the responses to different type of diversity and our statistical tests of the responses suggest that not all diversities are equally important. Specifically, “Diversity of professional background/experience” and “Diversity of personality” seem to be more important than the others. The influence of “Diversity of personality” on IS team building has been supported by [6] and [11]. Here the authors of the studies suggest different IS roles for people of different personality types. However, our finding that “Diversity of professional background/experience” is considered influential is, to our knowledge, new and needs to be investigated further.

The majority of the participants tend to agree with the statement that “A testing team performs better when they have experience working as a TEAM, rather than gathering experience as individuals”. However, the confidence of the agreement seems to be low, since a greater number of participants “somewhat agree” with the statement than the number of participants who “completely agree”. One possible interpretation is that practitioners believe that team experience makes some contribution to performance, however other factors are of greater importance.

## VI. CONCLUSION

Our study indicates that respondents consider multiple factors important for building testing team. It indicates that

some factors are perceived as more important than others, though the differences did not often reach statistical significance. We also conclude that diversity is desired for a testing team. “Diversity of professional background/experience” and “Diversity of personality” are two important type of diversity to be considered for building testing team. However, the specific influence of these on testing team building and performance needs further empirical study. We also found that the practitioners consider experience as a testing team is helpful for better performance, although not as important as other factors.

Not all of our results were conclusive. Nevertheless, we believe that the commercial importance and the lack of research to date in this area justifies further, more detailed work.

## REFERENCES

- [1] F. Shull, J. Singer, and D. I. Sjøberg, *Guide to Advanced Empirical Software Engineering*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.
- [2] R. Merkel and T. Kanij, “Does the individual matter in software testing?” <http://www.swinburne.edu.au/ict/research/sat-technicalReports/TC2010-001.pdf>.
- [3] S. G. Cohen and D. E. Bailey, “What makes teams work: Group effectiveness research from the shop floor to the executive suite,” *Journal of Management*, vol. 23, no. 3, pp. 239–290, 1997. [Online]. Available: <http://jom.sagepub.com/content/23/3/239.abstract>
- [4] S. K. Horwitz and I. B. Horwitz, “The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography,” *Journal of Management*, vol. 33, no. 6, pp. 987–1015, Dec. 2007. [Online]. Available: <http://dx.doi.org/10.1177/0149206307308587>
- [5] J. H. Bradley and F. J. Hebert, “The effect of personality type on team performance,” *The Journal of Management Development*, vol. 16, no. 5, pp. 337–353, 1997. [Online]. Available: <http://dx.doi.org/10.1108/02621719710174525>
- [6] N. Gorla and Y. W. Lam, “Who should work with whom?: building effective software project teams,” *Commun. ACM*, vol. 47, pp. 79–82, June 2004. [Online]. Available: <http://doi.acm.org/10.1145/990680.990684>
- [7] S. Faraj and L. Sproull, “Coordinating expertise in software development teams,” *Manage. Sci.*, vol. 46, pp. 1554–1568, December 2000. [Online]. Available: <http://portal.acm.org/citation.cfm?id=970301.970322>
- [8] A. Kolawa, “Software certification debate: Certification will do more harm than good,” *Computer*, vol. 35, pp. 34–35, 2002.
- [9] W. C. Schutz, *The Interpersonal Underworld*. Palo Alto, CA, USA: Science and Behavior Books, 1966.
- [10] A. Beer and R. Ramler, “The role of experience in software testing practice,” in *SEAA '08: Proceedings of the 2008 34th Euromicro Conference Software Engineering and Advanced Applications*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 258–265.
- [11] L. F. Capretz and F. Ahmed, “Making sense of software development and personality types,” *IT Professional*, vol. 12, pp. 6–13, 2010.