Review of: "Analyzing Students' Perceptions of Collaborative Tools for Automated Assessment of Programming Assignments in Distance Education"

Shee Mun Yong[1]

1 KDU College Sdn Bhd

Potential competing interests: No potential competing interests to declare.

This manuscript specifically references Mattermost as a collaborative tool supported by DSLab as an automated assessment tool. As much as it is is good to review collaborative tools used in education, I have several concerns with the approach and methods used in undertaking this study:

1. It is not clear as to the method and calculation used in determining the sample size of 271 from the total population. Furthermore, the sampling method seems like random cluster sampling and not just random. The criteria of individuals choosing their own pairs is not ideal as it might result in biased sampling. However, it is noted that the influence on the overall result would not be significant given the low number of groups.
2. From a research ethics standpoint, 168 students have signed the data protection consent form to participate in the experiment. Although the survey reflects the exact number of respondents, i.e., 168, log analysis was carried out on 271 participants. It is important to explain the reason for extracting logs for the participants that did not consent to having their data analyzed.
3. It is not clear how the improvement in the learning process using Mattermost is deduced when the results of indicators' mean values shown on Table 9 are not consistent. For instance, TU (Q11 and Q12) show significantly higher mean values for CG compared to EG, inferring a less effective tool utility for Mattermost.
4. There seems to be a contradiction in using Pearson Correlation for bivariate analysis on data that is considered non-parametric data, as evidenced in the use of the Mann-Whitney U test. Either way, the approach should be consistent. There is also no comparison made between the experiment and control groups that would determine the learning process of both groups.
5. It is not clear why multivariate, i.e., EFA, is carried out at the tail-end of the analysis, since understanding of the model underlying structure would not allow any remedial action to be taken. For instance, indicators of SL, CT, CI, and TU are loaded/cross-loaded into two factors, whereas this study covers four different variables.
6. In terms of the results, despite stating that deliveries in the experimental group are proportionally better than the control group, there was no evidence found as to the "significance" of these differences to support the outcome. Also, whether the percentage computation of the subsequent phases in Table 15 should be based on the passing numbers of the previous phase instead, e.g., 42 EG students participated in phase 3 and therefore the base used in the percentage computation should be 58 (phase 2), giving a result of 72.4% passes, as opposed to using the original

number of 78.