Author’s response to reviews

Title: Tutor Assessment of PBL Process: Does tutor variability affect objectivity and reliability?

Authors:

Bidyadhar Sa (bidyadhar.sa@sta.uwi.edu)
Chidum Ezenwaka (chidum.ezenwaka@sta.uwi.edu)
Keerti Singh (keerti.singh@cavehill.uwi.edu)
Sehlule Vuma (sehlule.vuma@sta.uwi.edu)
Md. Anwarul Azim Majumder (azim.majumder@cavehill.uwi.edu)

Version: 1 Date: 20 Jan 2019

Author’s response to reviews:

Response to Reviewer Comments for MEED-D-18-00848

Reviewers’ comments

Sr No Comments Response

Grace Wai-Lin Wong, BHM BTrngDev MHlthProm GradCert HPEd PhD (Reviewer 1)

Abstract

1. The abstract is concise. However, there are discrepancies on the number of tutors between the abstract and the manuscript. There is no mentioning of student sample size and significant values of findings in the abstract. Many thanks. We address the recommendations in the abstract and in the main text (Method section).

Literature Review:

2. The authors started the manuscript by describing the advantages of Problem-Base Learning (PBL) and listed a range of assessment methods to examine student’s learning. Many thanks. Based on reviewer’s track changes comments about the assessment of students’ learning (page3) we have rewritten the
3. Are there any problems with these assessment methods?

4. Which assessment methods will give less discrepancies among raters?

5. Which methods are better in maintain objectivity of raters and ensure reliability?

6. The authors described the Hawk-dove effects. You will need to link this with using of a rubric/criteria may eliminate this effect. We changed the sentences as per reviewer comment.

7. The authors need to provide the year of the study in literature review on page 4. Although there are few studies mentioned but the authors failed to identify the gap of knowledge and build up a good justification for this research. We added the years as suggested.

8. The aim of the study was to examine the differences between lenient stringent tutors in assessing PBL tutorials. Is this an assumption? How do you define lenient and stringent in this study? In the statistical analysis section (page 7) we have defined the lenient and stringent tutors. Thank for your observation

Methods:

9. The authors used an appropriate research design. Many thanks.

10. However, there are confusion on number of tutors and number of groups that these tutors are assessing. We have corrected this in the abstract.

11. The authors did not state what type of assessment was carried out? Were they group presentation? Individual presentation? Care plan? We have now clarified in the page 6, first paragraph. The tutors rated the students on PBL process.

12. The authors listed the twelve performance criteria We have already...
used in this assessment, however, no details were given if these are 'yes' or 'no' criteria or if there are different sub-scale in each of these criteria.

mentioned the Likert scale in Page 6:

i. The rating scale consists of 13 items covering 12 performance criteria and one global assessment which are to be rated on a six-point scale (Very Poor (0), Poor (1), Adequate (2), Good (3), Very Good (4) and Excellent (5)).

ii. In the last item, tutors used the six-point rating scale as (Novice (0), Beginning (1), Developing (2), Accomplished (3), Exemplary (4), Master (5) to assess the global performance/competence of the student.

13. The assessment instrument has been used by the school for more than 25 years. Reference 26 is not a publication or reliable source of information. We respect your observation. No publication was attempted. It was an in-house report.

14. How do you justify it is comparable to the other instruments which were used by the other pioneer schools? Did you adapt their instrument? Were the tutors trained to use this instrument? Have all the tutors used this instrument before? We consulted their instruments to develop a comparable and comprehensive rating scale suited for our context.

All tutors were trained to use this scale which has already been mentioned in the methodology section (page 5) and they used the instrument before.

15. It was not clearly stated how the data was collected? Was the tutor leading the group the assessor? Did each tutor assess each group? Were the assessment carried out in one day or in a period of time? Was all the PBL tutorial content the same for this assessment? The assessment was carried out in the period of time. The content is the same for all groups i.e. same PBL case was used in every week (now mentioned in the page 5, methodology section).

16. The authors stated that ethics approval was not necessary as this was part of the school assessment monitoring strategies. However, the results stated certain tutors were lenient or stringent which infringed on individual's privacy and reputation. Thanks for your advice. No personal information or confidential was revealed as the tutors rating scales were coded by assessment.
Results:

17. The first two bar charts are clear. The first figure showed the mean rating of individual group with G8 scored 58 the highest and G9 scored 40 the lowest. Figure 2 showed the mean rating of individual tutor ranges from 31 to 63 which revealed a statistically significant difference among the tutors. However, the rest of the results are confusing.

To bring clarity, the results are now presented in a tabular forms and the presentations in the result section made simpler.

18. How do you classify stringent and lenient tutors? Are the tutors with a higher mean rating scale lenient, with a lower mean rating scale stringent? I would define the tutor with similar rating scale stringent instead. I am not convinced about the necessity of comparing the mean ratings of each tutor. I would be interested in the discrepancies in different criteria. Which criteria were rated consistently or inconsistently among tutors? Using a line chart to plot out the rating by different tutors and the 12 marking criteria would be useful.

We agree with your definitions and in the statistical analysis section (page 7) we have defined the lenient and stringent tutors. We noted your suggestion to look into discrepancies in different criteria and we will consider in our future endeavour.

19. Any results on the years of tutor experience and the mean rating score?

Thanks for recommendation. We added the following: ‘The mean experience of PBL tutors was 12.8 years.’ and ‘The correlation between tutors’ PBL experiences and their mean ratings is found to be moderately significant (r=0.52; p<0.05).’

Discussion:

20. The authors should link the findings to literature which has been discussed in the literature review

Thank you for your comments and
section. As this instrument was comparable to other instruments, could results be compared?

22. Although the results revealed significant discrepancies in rating the assessment, there is no reason to define the high rater as lenient and the low rater as stringent. For the assessment results to be valid, you need to maintain consistency. It would be interesting to find out which are the criteria causing the discrepancies in rating. The authors can use these results to discuss how these can be rectified to ensure reliability of the assessment.

23. The authors have addressed a number of limitations of this study, such as PBL contents. If this tutors were not assessing the groups with the same tutorial content, this study is not valid.

24. Conclusion:

25. The training of tutors should be put in discussion. It is a good suggestion to explore tutor assessment with self and peer assessment of PBL

26. In general, the authors have established a sound research question. The research design is reasonable but not written in a style that is logical and clear to follow. It is obvious that the abstract, the limitations and results were written by a different author as the writing style is different. There are problems with referencing. Some part of the manuscript has a
Judith Strawbridge (Reviewer 2):

27. Many thanks for the opportunity to review your paper on assessment of Problem-Based Learning by Tutors. It is always interesting to explore assessment data and reflect on the challenges of ensuring that assessments are fair and reliable. I have some comments which are intended be helpful and assist in strengthening your paper.

Many thanks for taking time and forwarding constructive feedback. We tried to address from our end.

Abstract

28. It would be really helpful to the reader if the abstract contained sufficient information to orientate the reader to the full paper. The method section of the paper gives no indication of the number of students and tutors, and it is difficult to determine exactly what was done. Without the groups, it reads a little like only three tutors were involved. It may also be worth considering if your conclusion matches your study, in that you conclude that more rigorous training is required, yet in the body of the text you state that all the tutors received the necessary structured training in PBL delivery and assessment (P5 L38-39)

Many thanks. We tried to address the recommendations in the abstract and in the main text (Method section).

Background:

29. P3 L22-26 You might consider the strength of your argument for the superiority of PBL over traditional approaches. Perhaps this might be a little overstated for the evidence provided.

Sentence is rewritten to address the comments

30. P3 L49-54 I'm afraid I am not familiar with tripartite, triple jump assessment and patchwork tests. I'm sure I should be, and must look it up. You might consider if it is worth a line or two about more novel forms of assessment in PBL, particularly if they are of relevance to the way you have assessed in this study or if there are similar challenges?

Please refer the response given to point 2-5 of this table.

31. P4 L5 - 6 You might consider rephrasing re hawks having a tendency to fail most candidates - perhaps it should be "more". Similarly with doves.

Thanks for your comments. Corrected.
P2 2nd paragraph Feedback from students rating tutors is interesting, but you might consider if this is relevant to your study. This is not the focus of our study. We deleted the information.

Methods:

It would be really helpful if this section was clearer, as it is difficult to follow e.g. tutors were assigned to their groups for a period of 6 weeks, during which they rotated to different groups of students. I am not quite sure which groups were taught by which tutors, and this makes reading the results difficult to follow. This has been addressed and made more clear.

P5 L37-38 As mentioned earlier, you state that all tutors received the necessary structured training. I would be very interesting for the reader if this training was described, and given that the experience ranged from 1 - 20 years, is training repeated for tutors at regular intervals? The following paragraph has been added. The structured training covers topic such as, an introduction to the educational philosophy of PBL, systematic approach to PBL, the role of the tutor as a facilitator, encouraging critical thinking and self-directed learning, PBL process assessment, Rubrics: a helpful tool to bring objectivity in assessment, feedback: a helpful tool to enhance learning. The tutors included in the study, their experience ranged from 5 - 25 years not 1-20 years of PBL experience. Yes this training is offered beginning of each academic year, however participation is not mandatory.
Results & discussion

35. I suspect the results would be clearer if I was better able to follow the methods. It seems a shame that there wasn't some detail on the years of experience with PBL of the tutors in the results, particularly as this is a comment in the discussion (P9 L4). Some of the results referred to in the discussion e.g. students getting high scores and then lower scores in subsequent ratings (P8 L56-59), don't seem to be in the results. The final point in the discussion is that robust faculty development may minimise the effect of individual differences in tutor rating - but do we have any context regarding expectations that student performance would change between ratings? Thanks for recommendation. We added the following: ‘The mean experience of PBL tutors was 12.8 years.’ and ‘The correlation between tutors’ PBL experiences and their mean ratings is found to be moderately significant (r=0.52; p<0.05).’

Limitations

36. The limitations are acknowledged. One which might have been possible to address is No 3 - the exploration of PBL experience of each tutor. It might be worth considering if this could be addressed. Perhaps limitation No 5 is overstated - I'm not sure students assessments of tutors was necessary, unless you specifically were to address student perceptions of their tendency to be lenient or otherwise. We have already re-written the limitations as per suggestions of the Reviewer 1 and we added the tutor experiences in our study.

Conclusion

37. As above, a main conclusion is that more rigorous training is required, but from the data do we know this will help given that all tutors were trained? There could be other important conclusions to draw, such as raising the awareness of inevitability of differences in marking and that these need to be taken into account in assessment strategies. Suggestion considered and a statement to this effect has been added to conclusion.

38. Finally, a couple of typographical edits for consideration. P4 33 The explanations supplemented by authors "were" rather than "was". Page 10 Limitation 4 - extra "at" Thanks. We corrected these.