Author's response to reviews

Title: Multiple tutorial-based assessments: a generalizability study

Authors:
Christina St-Onge (Christina.St-Onge@USherbrooke.ca)
Eric Frenette (eric.frenette@fse.ulaval.ca)
Daniel J Côté (Daniel.J.Cote@USherbrooke.ca)
André De Champlain (adechamplain@mcc.ca)

Version: 2 Date: 1 February 2014

Author's response to reviews: see over
Reviewer: William Roberts

1. The first sentence in paragraph 1 of the Methods section is awkward. It is not clear to me a “unit” is a scored component or dimension of the Tutotest-Lite. The reference to Table 1 could be clarified by stating entries are summary statistics.

Response:
We added a definition of “unit” in this context to clarify: “the different contents covered in the curriculum ». We believe that this definition clarifies also the second point requested by the reviewer.

2. I would like more be said about the formation of new groups to each PBL content area? Is PBL standardized among newly formed groups?

Response:
Group formation is as follows: “For each PBL unit (i.e., content), a new group is formed by randomly selecting 8 to 9 students (while maintaining a consistent male to female ratio). » p.4

As stated in the paper, p.5 “All tutors are required to complete a two-day workshop on PBL principles and their application at the medical school where the study was conducted as well as a half-day of content specific training for each unit. Training is provided by the medical education center at this institution. ». To improve clarity, we added : « To ensure standardization of PBL units… [all…]».

3. The sample for this study appears to be from a single medical school. This should be mentioned as a limitation when generalizing findings to the population of all medical students.

Response:
We added: “Finally, generalization of these results should be done cautiously as the analysis was conducted at one medical school.”

4. In paragraph three in the Methods section it is stated “Committee members determine pass/fail standings from the individual scores”. It would improve clarification if the number of members on the committee were given and how the standard for a pass decision is operationalized.

Response:
We made slight modifications to the text as the word standard may have been confusing : “Thus, a student is not promoted if the promotion committee convenes that he or she fails tutorial-based assessment (TBA). To make its decision, the promotion committee reviews students’ TBA scores for each unit of the academic year as well as tutor comments, using a qualitative global assessment perspective. »

5. In the section describing the Tutotest-Lite it is stated in the first paragraph “Clinical domains to be developed within PBL sessions were identified through focus groups composed of faculty members”. If applicable expand discussion to include the extent of overlap between medical practice and domain content identified by committee focus groups.
Response:
We added the following precision: “…which were informed by the competency framework developed by the RCPSC, in order to have an overlap between clinical domains that should be focused on during PBL and those thought to be necessary for good medical practice. »

6. The design notation for the effects G-study design needs to be specified more clearly. For example, should this be conceptualized as a univariate or multivariate G-study design? The definition for “unit” and “clinical domains” is not clear to me. Perhaps the definitions would be clarified if expanded in detail when first mentioned on page 5. As part of the design specification state if this is norm referenced or criterion based examination. If this is a criterion based examination, then the index of dependability is the appropriate measure of reliability.

Response:
p.7 We specified that the design is a univariate generalizability analysis.
p.5 We specified that the four items on the Tutotest-Lite are in fact the four domains discussed later on.

7. In the second paragraph in the Analysis section it is stated “Analysis for Year 1 was conducted using only the three common abilities ...”. This seem to contradict what was stated earlier in the section describing the Tutotest-Lite, “Clinical skills are not taught and therefore not assessed in three of the eight units in Year 1 ...”. Clarify what is meant by these two statements.

Response:
p.7. The following modifications were made to improve clarity. “As previously noted, the form differed for three units of the first year. Thus, to have a balanced design, the analysis for Year 1 was conducted using a reduced form, that is using only the three common abilities in the eight units since systematic missing data cannot be legitimately incorporated in the model proposed.”

8. In the first paragraph within the Overall Reliability section it is not clear what “Additional file 1” is referring to when discussing variability components. Perhaps this is meant to refer to Table 2? The reference “Additional file 1” also shows up in the second paragraph.

Response: "Additional file" were replaced by the correct word (i.e., "Table 2").

9. Include dependability coefficients in Figure 2.

Response: We did as suggested.

Minor Essential Revisions

1. For consistency, in the third paragraph of the Background section numbers and words should not be mixed when referring to quantity “(e.g. three to 31 items)”.

Response: Letters have been replaced by a number for more consistency.
2. The first paragraph in the Data section could be shortened by simply reporting that the sample consists of 384 first-year and 374 second-year students representative of the school in this study.

Response:
We prefer to maintain the wording as is. We did not use samples of students – we used the entire cohort.

3. In the second sentence of the first paragraph in the Results section keep the format consistent in reporting means and standard deviations for each of the two years. Also, clarify if statistics are SEM or SD. In Table 1 these are listed as SD.

Response:
We changed the format when reporting the results in this section as suggested. Values reported are SD.

4. For each year, include test grand means and standard deviations in Table 1.

Response: We did as suggested.

5. Lines in Figure 2 can be better contrasted if one of the lines is segmented.

Response: We changed our Figure 2 so one of the line is segmented

Discretionary Revisions

1. In the second paragraph of the Background section remove “for example” at the end of the first sentence.

Response: We did as suggested.

2. I suggest referring to the scale as a four-point rating scale and dropping the word Likert.

Response: We did as suggested.

3. I suggest “Unit percentage scores means” be rephrased to unit means in percentage metric?

Response: We did as suggested.
Reviewer: Danette McKinley (DM)

This paper is an interesting application of generalizability theory to a problem in measuring competencies developed during the course of problem-based learning. There were aspects of the data that made modeling of certain facets impossible, and resulted in about half of the variance in the model being unaccounted for. That said, I wondered more about the scoring of the instrument, and what facets could account for some of the variance which was left currently unexplained.

1) Unit scoring was clear, but how were clinical domains scored/included in the model? Additional clarification on this component is needed.

Response:
The following clarification was added: Clinical domains, that is the score for each item on the form per unit, was treated as a random facet as elements were selected from a list of 44 items contained in the original Tutotest.

2) Clinical skills were eliminated for Year 1 data, but included for Year 2 since the measurement occurred across all units. Is this a possible reason for the difference in variance components for clinical domain from Year 1 to Year 2 (0% vs. 4%)?

Response:
We ran the analysis for Year with the same three clinical abilities as Year 1 and yes, that seems to explain the 0% vs 4% variability for the clinical abilities facet. We have added a note to Table 2: “The analysis was conducted using only the same three clinical domains included in Year 1. We then observed a 0% variability for the clinical domain facet.” And we added the following precision in the discussion section: “Finally, the difference in percent variability for the facet clinical domain seems to be explained, at least in part, by the exclusion of one item in Year 1.”

3) I am assuming one observation per unit, per domain (3 for Year 1 and 4 for year 2). Is that correct? A sentence or two on scoring would clarify more about the design chosen for the generalizability analysis.

Response: See the clarification made on page 7.

Discretionary Revisions

1) Since no information by tutor was collected, does this mean that data by student group within unit was not collected either? This would need to be mentioned as an additional limitation of the study

Response:
The following precision was added to the limitations: “Similarly, information about group composition was not available to the researchers.”