Author’s response to reviews

Title: Validation of the 5-item Doctor-Patient Communication Competency instrument for Medical Students (DPCC-MS) using two years of assessment data

Authors:

Jean-Sebastien Renaud (jean-sebastien.renaud@fmed.ulaval.ca)

Luc Cote (luc.cote@fmed.ulaval.ca)

Version: 1 Date: 06 Mar 2017

Author’s response to reviews:

Dr. Julie K. Tilson

Associate Editor

BMC Medical Education

Revised version of the manuscript entitled “Validation of the 5-item Doctor-Patient Communication Competency instrument for Medical Clerks (DPCC-MC) using two years of assessment data” (Ref: MEED-D-16-00467)

Dear Dr. Tilson,

Thank you for the opportunity to revise and resubmit the manuscript entitled “Validation of the 5-item Doctor-Patient Communication Competency instrument for Medical Clerks (DPCC-MC) using two years of assessment data”. We thank both reviewers for their thoughtful comments that helped us improve the manuscript. Our responses to each of their comments are below.

Responses to comments from reviewer #1:
1. My primary concern with the paper is the fact that the generally poor inter-rater reliability results, for both the assessment under study and the primary assessment that it is compared to, is reported in the results but not addressed in the discussion in any way. In fact, it should be addressed as an area of primary concern for this mode of assessment in general. ICC values are often interpreted as: excellent reliability ≥0.8, moderate reliability =0.60-0.79, and questionable reliability <0.60. Thus two of the pairs would be considered questionable, one moderate, and one excellent for the DPCC-MC. The same is the case for the Côté et al. as well. Notably, one of the pairs appears to be no more in agreement than chance alone (0.03). This is obviously a substantial problem for these measures. If they aren't reliable between raters it seems inappropriate to recommend them (or in this case the DPCC-MC) for in an academic setting.

Response to comment 1: This is indeed a preoccupation that needs to be addressed, thank you for pointing this out. We did so by revising the manuscript in the following ways:

- In the methods section describing the second phase of the study, we specified that the intraclass correlation coefficient (ICC) used to estimate inter-rater reliability was the two-way, single-measures, consistency variant, which is always lower than the average-measures variant (Hallgreen, 2012, see manuscript for complete reference). Also, consistent with Hallgreen (2012), we specified that our interpretation of the ICC values is based on Cicchetti’s (1994) guidelines: below .40 is poor, between .40 and .59 is fair, between .60 and .74 is good, and between .75 and 1.00 is excellent.

- In the results section for phase 2, in the paragraph about inter-rater reliability, we improved our presentation of the results. For example, using Cicchetti’s (1994) guidelines we explain that, on average (i.e., mean and median ICC), the DPCC-MC has a good inter-rater reliability and that Côté et al. scale has a fair inter-rater reliability. We also specify that one pair of raters has poor inter-rater reliability using the DPCC-MC and that two pairs have poor inter-rater reliability using Côté et al. scale. In addition, we improved table 4 by adding the interpretation of the ICC values according to Cicchetti’s (1994) guidelines.

- In the discussion section, we added a paragraph about the inter-rater reliability of the DPCC-MC. In it, we acknowledge the fact that having two out of four pairs of raters with only poor to fair inter-rater reliability is preoccupying. We also explain that because two raters pairs reached good to excellent reliability, factors related to the raters and the study setting might explain the variability in the observed inter-rater reliability coefficients. Therefore, we suggest that the factors that could affect the DPCC-MC’s inter-rater reliability should be studied.

2. Furthermore, the methodology used also requires more attention. What type of ICC analysis was used? Was a power analysis done to determine the sample size?
Response to comment 2: As mentioned in our response to comment 1, we now specify in the methods section that we used a two-way, single-measures, consistency ICC, and its values are now interpreted using Cicchetti’s (1994) guidelines. No power analysis was done to determine sample size. Indeed, our small sample size limits the precision of our ICC estimates. We therefore added this information to the discussion in the paragraph on study limits.

3. I also strongly recommend that the authors reconsider the use of 'medical clerks' throughout the paper. I suspect that this term may not be recognized internationally. In the US a medical file clerk (often shortened to medical clerk) refers to a technician trained to assist with routine administrative activities in a physician's practice. Perhaps 'medical student' would provide better recognition of your work by the intended audience.

Response to comment 3: Thank you for this recommendation. We changed the terms “medical clerks” to “medical students”, “clerks” to “students” or “students on clinical rotations” depending on sentence, and “clerkship” to “clinical rotations”. We also changed the name of the instrument to “Doctor-Patient Communication Competency Instrument for Medical Students” (DPCC-MS).

4. Abstract: The first sentence of the abstract makes too many assumptions about the reader's understanding of the context of the study. Indicate that the students are medical students and indicate that among many competencies doctor-patient communication competency should be assessed.

Response to comment 4: The first two sentences of the abstract were rewritten accordingly: “Medical students on clinical rotations have to be assessed on several competencies at the end of each clinical rotation, pointing to the need for short, reliable, and valid assessment instruments of each competency. Doctor patient communication is a central competency targeted by medical schools however, there are no published short (i.e. less than 10 items), reliable and valid instruments to assess doctor-patient communication competency.”

5. Abstract: The reference to 'Côte et al' measure is awkward in the abstract methods and results. Consider referencing it as "a 15-item instrument developed by Côte et al. (published 2001)” and from there "Côte et al. instrument”.

Response to comment 5: We made the change.
6. Background: The DPC abbreviation does not seem necessary (use as few abbreviations as possible to facilitate ease of reading). (Line 74)

Response to comment 6: We changed “DPC” to “doctor-patient communication” in the manuscript.

7. Background: Use 'during rotations' rather than 'during these rotations' (Line 83)

Response to comment 7: We made the change.

8. Background: The sentence on line 83 has other issues as well: 'more comprehensive' than what? Just comprehensive is sufficient. Additionally, I would not contrast short and comprehensive with reliable and valid - I'm referring to the use of the word 'but'.

Response to comment 8: We made these two changes.

9. Background: Is there a reason <10 items was chosen as sufficiently short? It seems a bit arbitrary. Is there a sense of how much time should be spent completing the assessment? (Naturally, an assessment could have a small number of complicated items that takes an extended time to complete - though I recognize that this is not the case with the DPCC-MD.

Response to comment 9: We clarified why we used this 10-item cutoff in the background section (same paragraph): “Additionally, our experience at Laval University shows that it is logistically impractical to have 10 or more items on doctor-patient communication alone, when five to seven competencies are being assessed for each clinical rotation completed by the student.”

10. Background: There is intermittent use of the terms 'medical clerkship' and 'clinical rotations' to describe the same activity. Only one should be used.
Response to comment 10: We now use only “clinical rotations”.

11. Methods: Then sentence that runs from line 149-152 needs clarification. The second element is not clear. Consider two sentences: one dealing with positive correlation and one dealing with assessment for redundancy. The second element currently consists of a double negative that makes it hard to follow.

Response to comment 11: We changed the sentence to “Inter-item correlations were computed to make sure that all items were positively correlated. We also used these correlations to verify if some items were redundant, defined as having a Pearson correlation coefficient of at least .70, meaning that the items shared 50% of variance or more.”

12. Methods: I recommend against the use of the acronyms in this section (CFA, ULS, MMI).

Response to comment 12: We would prefer to keep them. They are explained in the same section and we think it is more coherent with the results section, which uses these abbreviations too.

13. Methods: The explanation from line 182 through 193 is quite difficult to follow. I think it will help to put up front that the total number of videos viewed was 35 (from line 193) from a pool of, presumably, 167. Please add a sentence describing how this selection was made. Then describe how the 5 raters contributed to the data collection.

Response to comment 13: We rewrote the “sample and procedures” section based on these recommendations. We hope it is now easier to follow.

14. Methods: On line 187 there is a numerical reference to the Côte paper but the name is not introduced. This assumes that the reader then checked the reference to understand the reference to the name on line 190. The reader will need more help with this.
Response to comment 14: We added, before the numerical reference, “developed by Côté et al. (published in 2001)”

15. Methods: The reference to the Instruments section is helpful on line 188 but the numerical reference 2.2.2 is not clear.

Response to comment 15: We removed “2.2.2”.

16. Methods: Given the demographics of students and patients it is rather inappropriate that the two items developed for the global assessment of doctor-patient communication skills assume that both the student and the patient are male. If this is not the case, please indicate as such. If it is the case, I recommend changing that for future studies. (lines 202-204)

Response to comment 16: This error happened when translating from French to English. We changed the items to: “If you had to give this student a score for his/her communication with patient skills, where 1=Insufficient and 10=Superior, what would it be?” and “Globally assess the candidate's ability to communicate effectively with the patient and establish a good relationship with him/her.”

17. Methods: The statement on lines 210-211 are confusing. How does the global assessment score produce a score of 14? Based on what is written it seems that there are two questions scored 1-4, resulting in a maximum score of 8.

Response to comment 17: As indicated in the Instruments section (second phase of the study), the first global assessment item is scored on a 10-point scale (1 to 10) and the second on a 4-point scale (1 to 4). Their sum therefore ranges from 2 to 14.

18. Methods: Please explain why junior and senior scores were separated - this comes up in the Results but I don't believe it is addressed in the Methods
Response to comment 18: In the methods section (phase 1 of the student, “analyses” section), we mention in the second sentence that “The internal structure of the scale was assessed separately for junior and senior students to see if it was adequate for both groups”.

19. Results: The comparisons made on lines 281-288 are confusing. I expected to see DPCC-MC compared to the Côte assessment and the Global assessment in succession. There is an unexpected focus on the comparison of the Côte assessment to the Global assessment in the middle.

Response to comment 19: The second specific objective of the study was: “To compare the DPCC-MS psychometric properties to those of a longer scale assessing doctor-patient communication.” Accordingly, in the methods section (second phase of the study, “analyses” section), we mention “Convergent validity of the DPCC-MS and the Côté et al. instrument was assessed by estimating their correlation (Pearson’s coefficient) with the global assessment score. To test if the difference between these two correlation coefficients was statistically significant at an alpha level of .05, we checked if their 95% confidence intervals were overlapping.” Therefore, we correlated Côté et al. instrument to the global assessment scale to compare its convergent validity with the DPCC-MS scale.

20. Results: The same lines are confusing in the use of the word 'both' (lines 282 and 284). Use something to the effect of 'the measures were highly correlated' rather than 'both were highly correlated' - the latter implies that two things were 'both' correlated with the DPPC-MC (which is the case but doesn't appear to be the intent of the sentences).

Response to comment 20: We agree. We changed to “The measures were highly correlated” as suggested by the reviewer.

21. Discussion: The claim that the measure has good psychometric properties (lines 307 and 309 among others) demands qualification regarding inter-rater reliability findings (see comments above). A discussion of why the inter-rater reliability results are so low is also needed.
Response to comment 21: As mentioned in our answer to the first comment of this reviewer, we clarified the type of ICC used (single-measures, which is always lower than average-measures), our interpretation of ICC values using Cicchetti’s (1994) guidelines, and added a paragraph in the discussion about the inter-rater reliability of the DPCC-MS. In addition, as mentioned in our answer to the second comment, we added a sentence in the discussion (paragraph on the limits of the study) about the small sample size used to estimate inter-rater reliability.

22. Tables: Please provide explanations for the abbreviations in Table 2 at the bottom of tables 2 and 3 (all tables would be appropriate for consistency).

Response to comment 22: We agree. We added explanations for the abbreviations at the end of all tables.

Responses to comment from reviewer #2:

1. I was hopeful when I read about the 'Standard Error of Measurement (SEM)' being included in the methods (and presumably the result section). However, as far as I could investigate no evidence regarding the SEM is provided in your paper. The SEM would be SD * srt(1-r) and needs to be incorporated in the decision whether somebody is a good, moderate or poor communicator. You will see, incorporating these values in your table would provide a different perspective of the results. The bad reviewer in me would think that that might be the reason that SEM is addressed in the text but not calculated in the tables (please correct me if I am wrong). […] I hope therefore that your team will consider embedding the SEM (either calculated based on the used classical psychometric analysis) or by using the G-theory analysis (using EDU-G as a suggestion for software). The most simple and straightforward approach would be to use the existing data and calculate the SEM as outlined above (verify formula and associated the references accordingly).

Response to comment 1: We agree with the importance of the standard error of measurement (SEM). As stated in the “analyses” section (in the methods section for the first phase of the study), we calculated the SEM using the formula SD * srt(1-r) and provided a reference (Harvill, 1991). SEM for senior and junior students are provided in the results section (phase 1), at the end of the first paragraph (.12 and .15).
Following the reviewer’s comment, we decided to add the SEM estimates to the analyses of the second phase of the study. This is now announced in the methods and presented in the results, including in table 3.

Thank you for considering the revised manuscript for publication.

Sincerely,

Jean-Sébastien Renaud, PhD
Assistant Professor
Laval University