Author’s response to reviews

Title: A proof of principle for using adaptive testing in Routine Outcome Monitoring: the efficiency of the Mood and Anxiety Symptoms Questionnaire -Anhedonic Depression CAT

Authors:

Niels Smits (n.smits@psy.vu.nl)
Frans G Zitman (f.g.zitman@lumc.nl)
Pim Cuijpers (p.cuijpers@psy.vu.nl)
Margien E den Hollander (m.e.den_hollander@lumc.nl)
Ingrid VE Calier (i.v.e.carlier@lumc.nl)

Version: 2 Date: 22 September 2011

Author’s response to reviews: see over
Reply on comments: (MS 2111053855550293)

September 22, 2011

In this file, the original comments are written in a typewriter font, whereas my replies are in normal fonts. To highlight the changes, I gave the changed text a red colour, and a Rev. number in the margin.

Editorial Requests

1. Ethics - Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration (http://www.wma.net/en/30publications/10policies/b3/index.html), and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate.

Reply: In reply to this we added a statement in the methods section on page 3.

2. Authors’ contributions - Please include an Authors’ contributions section before the Acknowledgements and Reference list.

Reply: In reply to this we added the section authors’ contribution on page 11.

Reviewer 1

Please describe response options a bit more

Reply: We gave more details for the rating scale used on page 4.

Reviewer 2

1. Although the issue addressed is interesting and important, in my view, the manuscript would benefit if the research question and the unique contribution to the field would be more precisely stated. As the authors state themselves, there is a growing amount of research already reporting on the advantages of CAT for the assessment of patient-reported outcomes, including substantial item savings. Moreover, there is literature already available describing adaptive versions of already existing static self-report instruments (e.g., Gardner et al., 2004). Thus, I would appreciate if the authors could define in how far their study adds to the knowledge on CAT attributes and advantages in clinical practice.

Reply: In reply this we explicitly introduced two goals on page 3. The first is to show the efficiency of the CAT version of the MASQ-AD, the second is to show that CAT versions of existing scales is an efficient method not only in research settings, but for monitoring patients in clinical settings as well.

2. Linked to a point raised below concerning diagnostic information on the sample is the issue of differential item functioning. DIF between
different diagnostic groups can not be ruled out in general (and is quite probable when assessing constructs related to mental health) so I would appreciate if the authors could add further DIF analyses on this issue to the appendix.

Reply: We performed DIF for the diagnostic variables, and fortunately no DIF was found (page 14).

3. In the appendix, there are two paragraphs dealing with local independence (third and seventh). Is there a reason why these two analyses are reported separately? If not, I would recommend moving paragraph seven directly behind the second paragraph. Further concerning local independence: the authors report that there are signs for specific relationships among items, which means local dependence. LD might be a challenge for item calibration, CAT applications and # estimations so the authors should discuss this issue in greater detail and explain why they decided not to exclude these items from the scale and consequently why they expect that they do not affect measurement.

Reply: We (a) joined the two paragraphs as suggested, and (b) performed extra analyses to study the issue in more detail (which showed the impact of LD was small); see page 12.

4. I would like to see more diagnostic details for the sample in this study. The authors write that according to MINI 4% suffered from a minor and 42% from a major depression. What about the remaining 54%? From my view, more detailed diagnostic information would be helpful because item calibrations and CAT functioning can depend on sample composition so that the reported results could be appreciated in greater depth if more information would be given on sample attributes.

Reply: We added full details on the diagnoses, see page 3.

5. In the results (section characteristics of the CAT) the sentence starting with Evidently, test information did not explain needs some more explanation. What is meant by inconsistent response behavior?

Reply: We added an explanation of inconsistent response behavior on page 7.

6. Next section in the results: It is argued that AUCs were higher than the value commonly used as a lower bound for a large effect size. Please detail the rationale why AUCs should be interpreted in terms of effect sizes.

Reply: We added an explanation (‘AUC is a effect size of clinical significance’) in the methods section on pages 6-7.

7. The authors conclude the discussion with the hope that the reported results would convince diagnosticians to develop their own CAT for their preferred instrument. However, to do that researchers not only need to know how to build an item bank and what measurement assumptions to be evaluated. In my view, they also need software providing a user interface which makes it possible to use the CAT in clinical practice. Maybe the authors would like to add some notes or recommendations for the reader on these more practical issues.

Reply: We added some comments and recommendations on page 10.

8. In the appendix, sixth paragraph, it is reported that the G2 statistic could not be used due to insufficient number of observations in each cell. I would like to see here a short explanation or at least speculation why that happened.
9. Please check the whole manuscript for typos. E.g., in the discussion, fourth and fifth paragraph, it reads CATs instead of CAT.

Reply: We changed these typo’s and checked the whole paper.

10. Please also check reference 8. There are some awkward letters behind the authors names.

Reply: These were typo’s in the reference data base which we removed.

11. I would recommend reducing the number of decimals in the tables to two.

Reply: We disagree that reducing the number of decimals would be an improvement because some information is lost in doing so. For example, the correlation of 0.999 would be rounded to 1.00, but this might cause readers to interpret the outcomes as being identical outcomes, which is not true. Moreover, it is a discretionary comment only. Therefore, we chose to maintain the table as is.