Reviewer's report

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

Version: 1 Date: 22 July 2011

Reviewer: Thomas Forkmann

Reviewer's report:

The authors present a well-written manuscript that reports on their attempts to heighten the efficiency of routine outcome monitoring by self-report through the use of Computer-adaptive Testing (CAT). The manuscript is illustrative and easy to follow. The issue addressed is important and of high interest for modern patient reported outcome assessment. Nevertheless, there are some points that I would like to see addressed by the authors:

Major compulsory revisions

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.

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.

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.

Minor essential revisions

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.

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”?

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

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.

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.

Discretionary revisions

9. Please check the whole manuscript for typos. E.g., in the discussion, fourth and fifth paragraph, it reads “CATs” instead of “CAT”.

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

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

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.