Author's response to reviews

Title: Item response theory analysis of cognitive tests in people with dementia: a systematic review

Authors:

Sarah McGrory (S.McGrory@sms.ed.ac.uk)
Jason M Doherty (j.m.doherty@sms.ed.ac.uk)
Elizabeth J Austin (Elizabeth.Austin@ed.ac.uk)
John M Starr (jstarr@staffmail.ed.ac.uk)
Susan D Shenkin (susan.shenkin@ed.ac.uk)

Version: 3 Date: 22 November 2013

Author's response to reviews: see over
Dear Dr Lon Schneider,

Thank you for asking us to re-submit a revised manuscript to BMC Psychiatry. We are very grateful for the comments of the two reviewers, and are pleased that the first reviewer is now satisfied and that only a few points were raised by the second reviewer. We have addressed these points and hope that by doing so have improved the manuscript.

We attach a copy of the manuscript with tracked changes, and one without, and provide a point by point response to the reviewer’s comments below. We hope that the article is now suitable for publication in BMC Psychiatry, and look forward to hearing from you.

Reviewer 1:

The authors have properly addressed all our concerns.

We thank the reviewer for this positive comment and are pleased that the reviewer is now satisfied with the review.

Reviewer 2:

1 a) I still find it difficult to accept that non-English articles were excluded.

We acknowledge the decision to exclude non-English articles seems rather restrictive. However while we are only looking at papers where English versions of assessments were used, this has no implications for the validity of cognitive testing in other languages. We have added this as a note (Methods, Exclusion criteria: paragraph 3):

“Non-English language versions of cognitive measures were excluded…….The decision to exclude articles using non-English language assessments has no implications for the validity of cognitive testing in other languages.”

The decision to only include English versions was made as the semantic range of each item is more homogeneous and thus comparable. However following the reviewer’s initial comment on this issue we did review the papers excluded for this reason and determined that of the 21 articles using non-English measures only one additional
paper would have been included if we lifted the restriction on non-English articles (16 did not meet inclusion criteria for the use of non-dementia samples, three written in foreign languages, one for the use of non-dementia sample and analysis of domain specific measure, and one for lack of test/item information). The results of the article otherwise meeting inclusion criteria are discussed (Discussion, Information: paragraph 3).

b) Even after excluding non-English articles they still report that at least 5 MMSE items are different for patients who live in the US compared to patients who live in the UK.

We thank the reviewer for pointing this out. We acknowledge that there are some differences in items between the samples however we felt that as these differences are limited to these items these differences are unlikely to explain the entire difference observed between the two samples (Results: Gibbons et al. paragraph 2):

“There were some differences between items used for the two samples. Orientation to state and county items in US sample were substituted for orientation to county and 2 streets nearby for the UK cohort and the nouns to repeat and remember were also different for the two cohorts. Although these differences limit the direct comparison of difficulty between these items as the differences are limited to these items they are unlikely to explain the entire difference observed between the two samples.”

c) A better approach would be to discuss all relevant papers whether or not these are English or non-English and then to mention the caveat to study language bias using Differential Item Functioning (DIF) in the way this has been done by Teresi and co-workers

We agree with the reviewer and acknowledge that Differential Item Functioning can be used to investigate language bias. We have added some detail on this (Methods, Exclusion criteria: paragraph 3):

“Differential Item Functioning (DIF) can be applied to examine the effect of language bias of items and tests administered in different languages. For example, if patients of equal cognitive ability tested in English and Spanish have unequal probabilities of responding correctly to a particular item on a cognitive assessment, then the item functions differently with respect to language. The effect of different test languages of cognitive assessments has been examined in this way [27; 39; 40; 41; 42; 43]. However these studies did not examine DIF in dementia populations and were therefore not included.”

2. Also of 384 articles, only four studies were selected.

We thank the reviewer for raising this point. Although the number of articles reviewed is small we do reference and discuss several other papers that were not included (Balsis et al.2012, Wouters et al.2008, Ideno et al. 2012, Korner et al. 2012, Benge et al. 2011, Lillo et al. 2012) so that readers are aware of the surrounding literature. However it is in the nature of systematic reviews to provide robust
information with regards to a specific research question which is what we have done. This inevitably leads to studies being excluded for various reasons.

3. Furthermore, at least the CAMCOG should be included.

We are grateful to the reviewer for this suggestion and apologise for not specifying why the analysis of the CAMCOG was not included. While an IRT analysis of the CAMCOG is available, published by Lindeboom et al. in 2004, this article was excluded as 59% of the sample comprised of a normal aging population and for the use of a Dutch version of the CAMCOG.

4. I do not understand the phrase "Examining difficulty hierarchies of the error categories within the items revealed some disordered item categories, for example committing two errors being less difficult than committing one error". Committing two errors being less difficult than committing one error actually reflects neatly ordered categories?

We are grateful to the reviewer for identifying this error. We have removed this sentence and have provided an improved explanation for hierarchical ordering (Discussion, Linearity and the assessment of change in severity: paragraph one):

“Wouters et al. [2] revised the ADAS-cog scoring based on the results of this IRT analysis by weighting the items in accordance with their measurement precision and by collapsing their categories until each category was hierarchically ordered, ensuring the number of errors increase with a decline along the continuum of cognitive ability. Examining difficulty hierarchies of the error categories within the items revealed some disordered item categories.”

Editor’s comments:

*You might spend some ink describing or clarifying your flow diagram, fig 1, in the legend or the first paragraph of the results.*

We agree with the editor’s comment. Figure 1 has been amended for further clarity and is in line with PRISMA guidelines. Additional information has been included in the methods section (Methods, Data Extraction):

“After duplicates were removed the titles and abstracts of 203 articles were screened by two independent researchers. 160 articles were excluded on review of title and/or abstract (excluded for example for non IRT methods, IRT analyses of functional or other non-cognitive assessments).”

*Consider a footnote to or modifying fig 1 and text to indicate the number of the 21 non-English exclusions that were excluded for other reasons as well.*
We agree with this comment and have amended Figure 1 accordingly and have included more detail in text (Discussion, Information: paragraph 3):

“However, all except one were excluded for other reasons also (16 did not meet inclusion criteria for the use of non-dementia samples, three written in foreign languages, one for the use of non-dementia sample and analysis of domain specific measure, and one for lack of test/item information).”