Dear Dr. Clark:

Thank you for the opportunity to address the Reviewers’ helpful comments. Please find attached revised manuscript and see below for responses to the comments.

Jan Reinhardt (Reviewer 2):

1. Page 10, line 32: "sample was […] representative". A representative sample is a random sample of the target population. Please write something like "The sample was intended to reflect characteristics …".

Change has been made as Reviewer suggested. See page 6 of revised document.
2. Page 11, 29: That health care is delivered in English is not a good argument for using English language tests. Why do you not state that the tests were available in English only and you wanted to select the best for your RCT which would then be translated into the most common Aboriginal languages?

Change has been made as Reviewer suggested. See page 7 of revised document.

3. Page 15, line 43. If you use ICCs which are more appropriate for your design because of nesting of measurements in patients you do not need to report Pearson correlations at the same time. Easy to solve: just drop Pearson correlations.

References to Pearson’s correlations have been removed. See page 12 of revised document.

4. Discussion: End discussion with some conclusion about which test you will use for your RCT.

Suggestion has been incorporated. See page 32 of revised document.

5. Page 17, line 49: Sentence "To ensure that […]". Move to Methods.

Change has been made according to Reviewer suggestion. See page 12 of revised document.

6. Table 1: Please state statistical method for comparison in heading or legend.

Change has been made according to Reviewer suggestion. See page 41 of revised document. “Effect size comparison of retest study participants’ demographics by cognitive assessment.”

7. Table 4: Drop Pearson correlations.

Change has been made according to Reviewer suggestion. See page 46 of revised document.

8. Table 5: In heading write "Inter-rater reliability analysis based on ICC …".

Change has been made according to Reviewer suggestion. See page 47 of revised document.
1. The aim was "to examine the reliability and acceptability of four cognitive tests for monitoring change over time for Aboriginal people with WKS and alcohol misuse". However, most patients were not diagnosed with Wernicke-Korsakoff’s Syndrome or alcohol use disorders according to the information presented in Table 2. This is very confusing.

In the revised document, the first sentence quoted had been amended to state “The restricted aim of this pilot study was to examine the reliability and acceptability of four cognitive tests for monitoring change over time for Aboriginal people.” The study aimed to pilot selected cognitive tests in participants with similar characteristics to determine whether these would be acceptable to use in patients with Wernicke-Korsakoff Syndrome.

2. Also, more precise diagnostic codes (e.g., F10.2, E51) would provide meaningful information about diagnosis for patients enrolled.

   The table has been amended to include additional diagnostic classification information. See page 42.

3. Also, authors stated in added sentences that this is a pilot study to confirm reliability of the cognitive measures in monitoring cognitive outcomes following treatment for WKS. Current design could not provide direct foundation for reliability of the measurement in WKS patients.

   The focus of the study is to examine the reliability of the cognitive measures in Indigenous Australians. The sentence has been changed to clarify this. See page 3.

4. More precise rationale for the short interval (1-5 days is adequate to confirm reliability of cognitive measures) is needed to support the conclusion--the description added (P11) was not supported with previous studies.

   While longer retest intervals would be desirable, we were primarily interested in alternative ranking of tests, evaluated in terms of short-term test retest reliability because the eventual treatment trial will involve 3-5 days treatment with immediate post-treatment assessment. Given the migratory and remote lifestyles of many of our participants, a longer test-retest interval study while important, may not be feasible.

   We have incorporated additional sentences into the manuscript to address the Reviewer’s concerns. Please see page 11 of the revised manuscript.
5. The authors have not supported the qualitative methods with references. This comment has been addressed on page 12 of the revised document where references have been added.

6. Interruptions or distractions in cognitive testing were noted in substantial proportion of the tests sessions (RUDAS had five [5/19:26%), Corsi had seven [7/19:37%] and CogState had 12 [12/18:67%] interruptions/distractions noted). These results suggest that scores of those patients who were interrupted may be biased. The data obtained from those patients may not be reliable. Please show whether those interruptions affected the results (reliability of the scores). If those data were affected the results, please exclude unreliable data.

Thank you for the observation, however bias in absolute level of scores was not of direct concern to the question under investigation in this study. Interruptions are inevitable in this acute clinical setting so we feel excluding these participants would present a less representative sample of the target population.