Author’s response to reviews

Title: Identifying culturally acceptable cognitive tests in remote Northern Australia

Authors:
Deborah Rock (debrock@hotmail.com)
Ian Robert Price (iprice@une.edu.au)

Version: 1 Date: 03 Jun 2019

Author’s response to reviews:

Reviewer reports:
Adam REEVES, PhD (Reviewer 1):

The authors identify a need for cognitive tests, of attention, visuospatial function, memory, problem solving, flexibility, in younger persons, using a game format, rather than verbalization, appropriate for an Aboriginal population in the remote Northern Territory (NT) of Australia.

Method: new tests were solicited by consultation with locals, re-formulated, piloted, then run. Results were compared to the KICA. English fluency was also tested.

General Comment: this paper is carefully written and backed up by suitable references. The stages of the research, consultation, piloting, and running the tests, are appropriate. The statistical analysis seems excellent. The topic is of great interest, and the paper should be published. I have only minor remarks to make about the presentation.

Comment 1: the hypotheses need more explanation.
The first hypothesis is that an exploratory factor analysis will find different factors - sure, this must be so - but why should they agree with the categories the experimenters have chosen? And if they don't, what will the authors do - change the items in the categories, get more data so the analysis is no longer exploratory, or what? Unclear.

We have rephrased the section on hypotheses (highlighted). The term “hypotheses” was perhaps too strong for this exploratory research. We actually had the more general overall research aim of finding a subset of existing cognitive tests that are suitable for Aboriginal people from remote areas of northern Australia in terms of cultural suitability and willingness to engage with the tests. The “hypotheses” are now rephrased as “analytic approaches”. These are all exploratory with no necessarily clear answer as a “hypothesis” might demand. Hence we make no predictions in the typical
hypothesis sense. As described above, we are trying to contribute to the accumulation of information and ideas which will allow better cognitive tests to be developed for this population.

The terms “hypothesis” and “hypotheses” have been removed from the Discussion. Highlighted.

The second is that the test scores will correlate with KICA scores that test the same function; is this because the same things are being tested in the same way, implying a positive correlation (unless the present test is hopelessly random), or is it something to be discovered because the tests only address the same function at an abstract level? Most readers will not know the KICA and will not know whether a correlation with it is meaningful. Actually, I suspect that the population sampled from the supermarket and that sampled to develop the KICA (people approaching dementia) are so different that any correlation should be small or null, in which case, bringing in the KICA distracts from the purpose of the study.

The KICA was included because it is one of the few available cognitive tests that have some acceptance by aboriginal people in northern Australia (including within this research – see line 153). It is reasonably widely used for general population testing even though it was developed as a dementia screening tool. It is therefore of interest to those working in northern Australia to see how the claims made about the test (such as Item 7 assessing “Frontal/executive function”) perform when compared to other recognised executive function tests. The nature of the tests which correlate with the KICA-EF give useful information about what this item is assessing. Again we emphasize the exploratory nature of the present work in guiding the accumulation of information relevant to the development of well-credentialed and well-accepted tests for this population.

We agree that most readers will not know the KICA but it is published in a recognised journal and can therefore be accessed by readers. It is well-known to health workers in northern Australia and to researchers in Australia.

The third hypothesis seems contradictory: the authors state that English proficiency can help, on line 86, but then hypothesize that it will have no effect (line 93). There is no explanation of why. I can see it, if they planned to exclude test items that did correlate - but this would be a matter of test design, not a hypothesis. At any event, more explanation is needed.

We have now rephrased the hypotheses as “analytic approaches”. As mentioned above, “hypotheses” is too strong a term for this exploratory work. We did expect some significant correlations between the literacy test and some of the tests but we did not have enough basis to make predictions. Revised section on aims & hypotheses highlighted.

Comment 2: it is important to identify tests that are widely used, such as card matching, to see if the Aborigine sample did better or worse than non-Aboriginal rural or city dwellers. Some of the tests were adjusted for the local population, but this is surely a good thing, ideally making performances comparable. (Some of the tests were unique, however.)

We had no intention to compare this population to any other population. We suspect that there will be differences between urban Aboriginal groups (better education, better English literacy, more Western contact) and more rural populations. However, the research was exploratory, to see which tests the Aboriginal groups were comfortable with. The longer term aim is to contribute toward the development of tests which will assist in the detection of cognitive deficits such as caused by injury and substance abuse and hence inform treatment strategies.
Comment 3: for a cognitive psychologist, the (poor) result with the sorting test is most informative - in a cautionary way. It seems obvious that the words used to instruct the participants were poorly chosen, as otherwise the example of equivalent groups (sorting 2 large and 2 small items into each group) would never appear when 'distinct' groups were asked for. These are all normal adults and there is no reason to think that 'distinct' is an impossible concept for them. I wonder if a linguist could explain the outcome. Meanwhile, it just shows how apparently innocent words can trip up researchers and make it appear as if the population being studied is defective in some manner. I wish everyone was as careful as the present team.

Since concluding this study, I (DR) have spoken to a clinical neuroscientist working with Aboriginal people in the NT. Although her feedback is anecdotal only, she says she has also found that traditional Aboriginal people (i.e. less acculturated groups) don’t understand sorting tasks. They’re not mentally defective, but the concept doesn’t fit with how they make sense of the world. Neither this study nor her feedback is definitive, but it is worth questioning Western presumptions about the universality of cognitive skills – particularly when those skills are used as the basis of assessments. Which is part of the point of this study.

Comment 4: line 366. The positive correlation with accuracy and negative with (reversely-scored) speed in factor 3 (Table 1) is most interesting. It should be described as 'concentration', as the authors do, but not as 'attention', which implies that both speed and accuracy improve, as in factor 1. Line 376 has dropped the word “concentration” and now only has “attention” (highlighted).

Comment 5: the authors are trying to develop a useful cognitive test for the NT Aboriginal population. Have they succeeded? Hard to know, but what I wonder is, just how well the first 3 factors can describe test results in a western population using well-established tests of cognition. It may be that the current tests fall short, but perhaps they are just as predictive as the best western tests are in the west, in which case, looking for even better tests for the aborigines could be a waste of resources.

This research is exploratory, just to begin the process of developing tools to aid assessment. It is as much about finding tests that tap into cognitive skills that Aboriginal people use as it is about finding tests that Aboriginal people find culturally acceptable. An important component of the research was listening to Aboriginal opinions. Aboriginal people tend to engage poorly in tests they perceive as too westernised, so we believe there is value in developing tests in partnership with them. The higher question of how good cognitive tests are in predicting useful outcomes (in either typical Western societies or remote Aboriginal ones) is indeed an interesting one. The general view throughout the literature seems to be that cognitive tests do provide useful, but modest, information about functional deficits that can inform interventions and health strategies. So, for the moment we aim to contribute to this. But, yes, it may all be a waste of time and resources.

The very first paragraph of the paper has a statement to reflect the role of cognitive testing: “. . . cognitive tests do not provide sufficient information alone, but do provide meaningful information that supports assessment”

Comment 6: the authors make no mention of other aboriginal groups. I have only met those around Alice Springs, so I have no personal knowledge of NT conditions. However I was told by a guide that his entire clan were displaced to western 'boarding schools' in the early twentieth century, breaking up many of the traditional ways of thought and imposing western patterns (which on their later return to the Alice area they were trying to re-establish). Anyway, I wondered if by 'remote' the authors are implying that this never happened to the NT sample they investigated, or whether their sample has also
been influenced by the west (many have some English). This question goes well beyond the current paper, but for outsiders like myself, a broad generalization would be helpful.

The principal researcher (DR) has lived over 10 years in Alice Springs, and also lived several years in Arnhem Land, where the study took place, so is in a position to comment. There are significant differences between the 2 Aboriginal populations. We agree there is significant cultural loss in Alice Springs, due to practices the reviewer mentions, and where Aboriginal people now make up less than 20% of the population. In Maningrida Community, they are around 90% of the population. Very few people in Maningrida are of mixed racial descent, very few speak English as their first language (almost none), and language, ceremonial traditions, cultural beliefs, and traditional hunting practices are still part of day to day life, with some adjustment for improved hunting tools, supermarkets, and mobile phones. The two populations are at different ends of the acculturation scale. Western tools are ok for many acculturated Alice Springs Aboriginal people, but that is not the case for the Arnhem Land people.

Maria Papadakaki (Reviewer 2): Many thanks for the invitation and the opportunity to review this interesting manuscript from Australia. The authors have invested a lot of time and effort to adapt and validate a number of tools to an aboriginal population. The overall aim of the study was "to identify culturally suitable tests". However, I feel that the locality of the study, the small and convenient samples used as well as the use of very few and specific validation tests (e.g. factor analysis) do not allow generating sufficient evidence and reaching a conclusion on the suitability of all those tools. Some more comments are found below:

The general problem of developing culturally suitable, but useful, cognitive tests for Aboriginal people in northern Australia is very difficult to solve in one go. It needs to be an accumulative approach from several researchers. In fact as indicated by the references we cite, this is what is happening. Many authors have contributed small increments to our understanding about what tests are suitable and which ones are not, and how to go about determining that suitability. In particular, BMC Psychology has contributed to the discussion and way forward in this process by publishing articles in this field (e.g., Dingwall). So we are contributing to that discussion.

We are not claiming to have established the validity of these tests, but more the suitability of the tests to this population. Our three stage process to selecting appropriate tests has allowed us to completely eliminate some tests from consideration while supporting the use of some other tests. We have also given ideas about how some tests can be easily adapted to local conditions. This is a very important early step. It is now up to other researchers to work on the validity of the acceptable tests – and what they are exactly measuring and what the scores are actually predicting. We have tried to give some information about validity by our PCA, correlations with the comparable KICA items, and the literacy measure. But we make no claim that the validity of the acceptable tests has been demonstrated.

1. The authors mention that they explored "construct validity" through carrying out factor analysis. I realize that they aimed at validating certain tools in a specific population by examining their internal structure as well as their discriminant validity against external criterions (gold standards)? If yes, this is not explicitly stated and needs to be revised.
We have rephrased this to remove reference to “construct validity” (highlighted – sentence starting "In the third stage . . ."). We explored how the various scores we obtained grouped together – and hence get some idea as to the underlying constructs that were being assessed. But it was very exploratory, and we cannot claim to have established “construct” validity. We did not attempt to assess “discriminant validity”.

2. An ad-hoc English literacy test was used to investigate whether any tests were detecting literacy rather than cognitive skills. Discriminating between literacy and cognition seems to be an important aspect of this study but it is neglected in the introduction. It needs a more extended justification. Besides that, the authors don't state how their sampling design ensured efficient assessment of discriminant validity.

We did have a couple of sentences on the relevance of literacy and education to test scores in indigenous populations, but these were not in the one place. We have now added another sentence and brought the scattered points together to emphasise this a little more. (Highlighted, 3rd paragraph). Regarding how the sampling design addressed “discriminant validity” we can only emphasise again that this was an exploratory research program aimed at assessing suitability of various tests to this population. If Aboriginal people do not engage with the tests in the first place we cannot do anything about “construct” or “discriminant” validity.

The sample was, admittedly, a convenience sample. Working with indigenous people requires understanding and accommodation to their culture and ways of behaving. Finding a large representative sample was beyond the resources of this project.

3. "There were three hypotheses……" These hypotheses refer to the expected outcomes of the statistical tests. But what do these outcomes mean regarding the study objectives? Did the authors try to generate evidence on the validity of these tools? Please make it more clear for the broader readership.

The three “hypotheses” have been rephrased to more clearly indicate an overall “aim” and three “analytic approaches”. The overall aim for the project were more general than “hypotheses” and the “hypotheses” as stated were not clearly answerable in the sense that hypotheses can be specifically answered.

We did provide some evidence for what the test scores were measuring, by examining how they related to each other (Principal Components Analysis), how the test scores related to the corresponding items in the KICA, and the relationship between the scores and the literacy test. These results were meant as indicative only and could not be interpreted as establishing validity in any sense. Our main aim was to determine the suitability of the tests to this population. It will be up to later researchers to work with the tests that we have identified to more formally establish validity.

4. When did this study take place? Please add the year.

2017

This now included (and highlighted)

5. What was the methodology employed for the initial selection of tools?

The Procedure section has been revised and rearranged (all highlighted) to include more information on initial selection of the cognitive tests and to make clearer what the procedure was.

6. "Thirty people were invited to join consultation groups" - please describe the recruitment strategy.
We already have “Thirty working people . . .” and “Participants were Aboriginal teachers, health-workers and well-being workers, two were non-Aboriginal mental-health workers” – all were “multi-lingual”.
This indicates that these people were recruited from health services, mental health (and well-being) services and education services. Aboriginality and clan/language association was described. This section left unchanged. Not highlighted.

7. The characteristics of the study sample are not presented. What was the profile of the participants? Please add relevant tables.
Age range, gender, tribal identity and languages for participants are noted in the text in the methods section for the pilot and main trial. Gender and tribal identity only is noted for the consultation participants.
We do not think further information is needed.

8. "Participants rejected the progressive matrices, Tower of London, and Block Design. These were considered confusing and too reflective of non-Indigenous ways of thinking." What did this evaluation process involve? were there evaluation criteria used to assess tool suitability? if yes, what were these criteria and who set the criteria?

The decisions were based on feedback from the Aboriginal participants. It was not based on feedback from single individuals, but rather if several people gave consistent responses, that a test was confusing, inappropriate, or too reflective of non-indigenous ways of thinking, the feedback was taken seriously. Some of the consultation group were curious to try some of the tests; if competent individuals had difficulty doing even very simple items, this was also seen as negative feedback about the suitability of that item.
The results could be due to the nature of the sample. The feedback is reported as indicative only.
We have left the relevant section for this as it was. We have already emphasised the consultative process we used for the pilot stage. It is clear that we did not use explicit criteria but rather verbal feedback.

9. "One researcher conducted all testing" - what was the profile / background and training of this researcher?
Another sentence added to give the researcher’s credentials. Highlighted.

10. "Overall, feedback to the final set of tests was positive. The use of familiar objects, practice sessions and shorter tasks worked well" - what does "worked well" mean? was this assessment among the objectives of this study?

We believe the statements that “feedback . . . was positive” and “worked well” are already quite clear and are explained further by the example quotes given. “it’s alright, not too much balanda way” indicates the tests worked well because of cultural acceptability. “it helped me to focus and to bring back my memory” illustrates that the tests worked well by recognition of the cognitive requirements of the tests. And “participants noted that they felt happier afterwards, and several suggested using the tests for wellbeing” illustrates that the tests worked well for engagement and enjoyment.
We have left this section as it was. Not highlighted.