Comments on the revision.

The authors' replies go through my comments one by one so here are corresponding replies to that.

.1. I am happy with the revisions made in response to this.

.2 I perhaps didn't express myself very well there. I wasn't trying to argue that there wasn't any need or value in replicating the studies, only that the eventual publication should have a sound justification for it, which means setting it in the context of subsequent developments in that research area. If there is real doubt as to whether the findings would replicate or not, then certainly the replication attempt is justified, and the subsequent literature does seem to justify feeling some doubt about it. I think I just wanted the authors to have a clear idea about what it would mean if the findings were replicated, and also what it would mean if they were not replicated.

.3 The authors' reply is satisfactory - I just wanted to be sure that readers of the eventual publication would get a clear understanding from the paper of why the replication matters.

.4 O.K.

.5 I don't have any idea of a measure of confidence that would be trustworthy but I stand by my original comment that explicit judgments of confidence are prone to response biases. Replicability is not a guide to trustworthiness because it might mean only that the same response biases were operating in both the original and the replication. I'm not saying the authors shouldn't obtain confidence judgments, just that they should perhaps include some nuanced discussion of the results when they get them. What they have said in their reply to my comment is the right sort of thing, in my view.

.6 O.K.

.7 In the original submission it reads "scores supposedly either representing academic achievement, mental concentration, and sense of humour". The two choices are (i) remove "either" and (ii) change "and" to "or". The authors can go for whichever of those they prefer. Apologies for being a bit pedantic about this.

.8 O.K.

.9 O.K., that is very useful. The research I do pretty much has to be done face-to-face so I have never explored online alternatives. The use of Prolific is probably more common in some areas of psychology than others.

.10 It is the "If things fail..." that concerns me. A simple way to deal with the problem would be to analyse separately for each experiment the data from the participants for whom it was the first one they saw - at that point their judgments could not be affected by the other studies because they haven't done them yet. If the results for that sub-sample resemble those for the full sample, then no problem.

.11 O.K.

.12 O.K.

.13 I sympathise with the authors and I think their discussion of the issue is intelligent and appropriate. I agree with the decision to set alpha at .001 for exploratory analyses. For the analyses where .005 is used, I would suggest that, if they get results significant at .01 but not at .005, they could discuss these or at least list them, so that readers could get a feel for whether there is any likelihood of type 2 errors, but I'm happy for the authors to go with their own judgment on this.

.14 O.K.

Overall. I would like the authors to bear my comment on .4 in mind when writing up the results. The grammatical error commented on in .7 should be corrected. The authors should consider the suggestion for further analysis in .10 but I will leave it to them to decide whether to do it or not. They should also consider the suggestion made in .13 but again it's up to the whether they do it or not. I have no further requests for changes.

Peter White