Author’s response to reviews

Title: Test-retest reliability of adolescents' self-reported physical activity item in two consecutive surveys

Authors:

Kwok Ng (kwok.ng@ul.ie)
Riikka Hämylä (riikka.hamyla@gmail.com)
Jorma Tynjälä (jorma.tynjala@jyu.fi)
Jari Villberg (jari.villberg@jyu.fi)
Tuija Tammelin (tuija.tammelin@likes.fi)
Lasse Kannas (lasse.kannas@jyu.fi)
Sami Kokko (sami.p.kokko@jyu.fi)

Version: 2 Date: 16 Jan 2019

Author’s response to reviews:

Reviewer 1
accept

Reviewer 2

Several concerns remain: The conclusion in the abstract does not match the conclusion in the body. The conclusion in the body seems more appropriate, as only one ICC was excellent. It seems misleading to characterize the results as excellent reliability. In the response, the authors state that the following sentence has been removed, "The reliability of the MVPA item was examined by a correlation matrix." However, it still exists at line 182-183. I remain unclear what is meant by this. Line 189-190: Verb missing in sentence. Line 192: What is meant by "no shift"? Is this exact responses? Line 193: What is meant by greater? Greater than what? What is the purpose for the MVPA 8 categories Kappa analysis and the MVPA Continuous ICC? Are these statistics on the same set of data (i.e. exactly the same set of values, or were they recategorized responses)? If so, why are the interpretations so discordant? Is it expected that categorical responses are less reliable than continuous responses? How should this be interpreted? Which type of ICC was used? The discussion remains somewhat challenging to follow. Further English language editing, especially for the discussion, is recommended.
Respond: Thank you for carrying out the review. We have matched the conclusions in the body and the abstract. We have improved the understanding of the sentences in between lines 182-193. We have removed the Kappa for the 8-answer item. We have stated the type of ICC used (intra-rater). We have also included more discussion points about item ordering as this seems to be an interesting point about future planning of survey questions.

Editor

I agree with many aspects of the review of reviewer 2 in that you should reconsider your use of the kappa statistic for the 8-answer item. It may be most appropriate to remove the kappa results for that assessment and simply report the ICC values. Also, please indicate which model ICC you used.

Other minor considerations:

* The discussion about question ordering, question salience, and context appears most relevant to this paper. Which survey do you believe most accurately represents actual MVPA? Both self-reported data represent second-best approximations, but there is an opportunity here to discuss accuracy. In contract to leading to confusion as you suggest in your discussion, having 3 questions about physical activity and distinguishing low PA may add salience to the MVPA question by surrounding it with other questions about activity. This could increase its accuracy - which could negatively affect reliability. What gives you the confidence that the HBSC survey is the more accurate number? It appears your recommendation is to put the MVPA question first which may increase reliability but in a scenario in which increased salience increases accuracy than it could increase reliability at the sacrifice of accuracy. Of course, this would require testing of question sequence along with objective measures of PA to arrive at the answer, however both possibilities (decreased accuracy through confusion and increased accuracy through salience) are worthwhile discussing in your paper.

Respond: Thank you for taking into consideration of the reviewer’s comments and providing your own perspective. We have addressed the areas that the reviewer and responded appropriately. We have also included some aspects of discussion about the ordering of items.