Reviewer's report

Title: Field assessment of balance in 10 to 14 year old children, reproducibility and validity of the Nintendo Wii board

Version: 2 Date: 6 April 2014

Reviewer: Arnold Huurnink

Reviewer's report:

The authors have done a thorough job to reply on my comments, and, to my opinion, the manuscript has been improved.

However, there remain some issues that need to be addressed, but I think that the authors can come up with a revised manuscript that is suitable for publication in the Journal of ….

My main concern is that the limitation and implication of the current approach to assess validity of NWB is not always adequately addressed throughout the text. It may help to read the following papers, to better understand my point of criticism:


And I think it is essential to be aware of the criticism of Pagnacco et al. on the paper of Clark et al. (not that I agree with all of his statements…):


Secondly, the results section is unnecessarily large, and redundant with Tables en Figures. Readability would be enhanced if only summarized key points of the results are presented in the text, in such a way that it becomes clear how the data should be interpreted.

Minor Essential Revisions

Methods_statistical analyses_paragraph 2:

1) This paragraph is formulated a bit odd, as LOA is actually a quantification method for reproducibility and agreement, and does not visualize it. The Blant-Altman plots will visualize the data points relative to mean difference and LOA. Furthermore, the 95% LOA is not explained. Does the 95% LOA not simply mean that the difference between two sessions (reproducibility), or two methods (validity) will be within the 95% LOA for 95% of all subjects?

Please rephrase this part of the paragraph and explain 95% LOA as well.
2) “MDC is also related to limits of agreement, and only if the change in measure is outside the limits of agreement is the change is statistically significant” Please rephrase.

3) “CCC point estimates were interpreted as poor (0-0.39), modest (0.4 – 0.74), or excellent (0.75-1)” Why do the authors stick to these point estimates? The readability throughout the text will be considerably increased to simply state that CCC (or ICC) > 0.7 is considered sufficient or acceptable. In the discussion CCC values can still be compared with other studies (Clark et al), and the limitation of the point estimate is already discussed.

Major Compulsory Revisions:

Results_Test-retest reproducibility of NWB and AMTI:

1) There exists redundancy between the text and figures/tables. It is sufficient to highlight the key points of the results as it would greatly enhance readability. For example, presenting both SEM and MDC in the text is unnecessary, as both represent absolute reproducibility (and are 100% correlated) and are both presented in the tables.

In my opinion the important results of the reproducibility analyses can be summarized in approximately 5 sentences, for example:

- Out of the eight Bland-Altman plots only one reveals a systematic bias (5%) for the unilateral non-dominant eyes open test on the AMTI.
- For NWB, the MDC varied between 16.9 and 36.9 cm (26.5 – 28.6 % of mean outcome) in the four tests, which is comparable to the results of the AMTI with MDC values between 14.7 and 36.1 cm (26.3 - 28.1%).
- Additionally, all CCC values were above 0.7, ranging from 0.76 to 0.83 for NWB and from 0.79 to 0.86 for AMTI, with all 95% lower bounds being # 0.65. Although confidence intervals of the CCC values largely overlap between NWB and AMTI, the values for NWB are consistently slightly lower.

Results_concurrent validity:

2) “LOA showed larger variation in the unilateral tests than in the bilateral tests and the line of observed agreement indicated that the NWB gave longer measurements in bilateral tests, but shorter measurements in unilateral tests” This information is not really helpful. What the reader want to know is:

- Is there a systematic bias? (you address that correctly in sentence 1).
- Does the 95% LOA between NWB and AMTI differ from the reproducibility results (95% LOA or MDC)? If so, what is the extent of the difference, as this information will give the extent of measurement error due to NWB
- Does the concurrent validity CCC’s differ from the concurrent reproducibility CCC’s, as this information will give the extent of measurement error due to NWB. In the present measurement setup the 95% LOA between NWB and AMTI and
the validity CCC’s per se are of little meaning, unless it is compared with the reproducibility analyses. Therefore it is also impossible to formulate point estimates for validity CCC’s on forehand, when the current validity design is used.

Please rephrase this results section in such a way that the abovementioned points are addressed.

Discussion:

3) Please make it more clear in the first paragraph what the main findings of the paper are, related to the aims formulated in the introduction. Aren’t the main findings that (1) it is possible to measure sway of children in a field setting with sufficient reliability (>0.7), (2) that NWB and AMTI possess equal reproducibility of COPL in children (based on 95% LOA, MDC and CCC), and (3) that a possible measurement error of NWB compared to AMTI is small compared to intra-subject variability, as 95% LOA and CCC of NWB vs AMTI is comparable to 95% LOA and CCC of test-retest comparison.

4) For comparison of your CCC and MDC values of children in bipedal stance with previous studies in adults, the following references may be used:


5) Besides Clark et al., there is another study that has focused on reproducibility and validity of the NWB and AMTI (this should also be added in the introduction):


6) In line with the abovementioned comments for the results section, the discussion should explicitly state that validity of NWB cannot be assessed directly, but that it can be estimated to compare LOA of NWB vs AMTI with the test-retest LOA. If these values are comparable, than the LOA of NWB vs AMTI is mostly due to intra-subject variability. Than it can be concluded that measurement error due to NWB is probably small compared to intra-subject variability.

7) It should be stated in the limitation section that the validity of the NWB was assessed without putting the NWB on top of the AMTI, therefore intra-subject variability did obscure the validity methods. Although this limitation was to some extent compensated by taking the intra-subject variability into account, the current method is probably not able to detect small systematic errors or the assess the exact extent of the possible bias of NWB measurements.
Abstract.
- Please find a way to present actual (ranges of) values of 95% LOA or MDC, so it is clear for the reader that NWB vs AMTI is comparable to test-retest comparisons, instead of only presenting CCC values.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests