Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-3555

Reviewer Comments

The authors are testing an autostereoscopic screen adapted for vision screening with eye-tracked directional backlight technology on a fairly large sample of participants. The visual acuity tumbling E is used on the device and compared with a paper acuity eye chart. They report various dimensions, including subjective likeability, test-retest or correlation between the two acuity tests. They generally find the autostereoscopic test to be a better test than the paper chart. I have significant uncertainty about the manuscript. I will focus on some major points and if those can be addressed, I would bring in the other aspects.

1) It is often difficult to follow how the claims are linked to specific measures. For example, the manuscript report that the autostereoscopic device has better likeability, test-retest, accuracy, efficiency, is faster, user-friendly, multifunctional, subjectively more popular and comfortable, has better inter-visit variability and clinical repeatability, validity, convenience, sanitary and produced less fatigue. While some of those adjectives are clearly linked to actual measures, like in the abstract, others are not. Please provide a clear linkage between the tested qualities and the actual measures, at least in the methods, and remove all synonyms of those qualities, to prevent confusion.

Response : We thank the referee for the valuable suggestions. We agree with him/her that it is more readable and logical to link the claims with the specific measures. In our study, repeatability test, accuracy, likeability, visual fatigue and work efficiency were all evaluated clinically, and the adjectives used were reproducible, accurate, subjectively popular and comfortable and fast, respectively. The related contents have been revised in this manuscript, and the abstract was rewritten, with additional information provided on Page 2, Lines 40 to Page 3,Lines 46. The main body content was also revised, as discussed in our response to question 2. Please kindly refer to the revised manuscript for further details.

2) The authors describe an auto-stereoscopy device and in the introduction, explain the merits of the device, particularly to test binocular vision (stereo, balance) or color vision. In the methods, the procedures that were used for stereo and color testing are even described. However, only visual acuity is reported. Therefore:

a. I would like to know why those results are not reported here. Is it because they are not positive? Could you report whether they were positive or negative, in the response to the reviewer so that I can assess whether they should be reported here or not.

Response : We performed tests for binocular vision (stereo, balance) and colour vision, and all the test results were positive. We thought that the manuscript would be long and tedious if we described each of the test results in detail, as we did for visual acuity, but this obviously caused ambiguity and uncertainty for the reviewer. In this new submission, we clearly pointed out in the abstract and in the main body of the manuscript (Page 2, Lines 40 to Page 3,Lines 46; Page 12, Lines 240 to 246; Page 13, Lines 256 to 262, 268 to 275, and 277; and Page 14, Lines 278

to 280, 282 to 285, 287 to 290 and 292) that the test results were all positive for binocular vision (stereo, balance) and colour vision, an assertion that is supported by the summarized test results described in the results section (Page 13, Lines 256 to 262, 268 to 275, and 277; and Page 14, Lines 278 to 280, 282 to 285, 287 to 290 and 292). Nevertheless, the page limit does not allow us to describe these results with the same level of detail that we used for visual acuity.

b. In the pre-registration, did you state that you would report them? If yes, they should be reported at the same time as the visual acuity results here to prevent salami-slicing.

Responses : We revised the manuscript accordingly, as mentioned in our response to question 2a. We had no intention at all to perform salami slicing; rather, we simply aimed to avoid a long and tedious description. In the manuscript, we tried to present a full description of all the positive test results while keeping the manuscript within a reasonable length.

c. If it is clear that they could be reported in another report, could you please rewrite the paper accordingly – at the moment, from the methods and introduction, it really feels that you are going to report the stereopsis results, at least, given the emphasis put on the interest of autostereoscopy to measure stereopsis.

Responses : Indeed, many new and interesting results have been obtained by applying autostereoscopy for multifunctional and self-assisted visual tests. In this new submission, we have tried to provide the fullest possible description of the test results, as we have mentioned in our responses to the previous two questions.

2) Do you provide a Bland-Altman plot for the test-retest which allows to assess whether there is learning between tests 1 and 2 (with the bias). But you are not reporting whether the bias is significantly different from 0.

Responses The Bland-Altman plot has been added; the bias was -0.02, and the 95% limit of agreement was -0.14 to 0.09. As the bias range included 0, it was not significantly different from 0, which suggests good repeatability.

3) Could you also provide a Bland-Altman plot for the comparison between the two acuity tests: this is helpful to assess the quality of the tests and how they compare.

Responses: We agree with the referee. The Bland-Altman plot for the comparison between the two acuity tests has been provided in the new submission; please see figure 6C and the 2nd paragraph of the results section on Page 13, Line 266 for further details.

4) Why do you use an ICC for the correlation between the tests? If each test measures the acuity on a continuous logMAR scale that has the same range of possible responses, a correlation is enough, no?

Responses: The intraclass correlation coefficient (ICC) was used to evaluate the correlation between visual acuity scores assessed by the autostereoscopy test and traditional methods. We agree with the referee that the correlation analysis should be enough to show the relationship

between the autostereoscopy test and the traditional methods.

5) I do not understand the radar map: what are the measures that are represented?

Responses : The radar map was based on the questionnaire that was administered to the subjects undergoing the test. The data were obtained to compare the accuracy, efficiency, comfort, data management and user experience of the autostereoscopy test with those of the traditional methods. To make the manuscript more readable, the test results have been re-described in figures 6A to 6F based on the clinical data presented in this manuscript. Please see Page 13, Line 256 to Page 14, Line 293 for further details.