Reviewer's report

Title: Inter- and intra-observer reliability of clinical movement-control tests for marines

Version: 4 Date: 1 November 2012

Reviewer: Gertrud Nilsson

Reviewer's report:

Re-review of the manuscript MS: 7225483486922265
Inter- and intra-observer reliability of clinical movement-control tests for marines

I think the authors have done a great job in revising the manuscript. Most of the questions have been addressed and revisions have mostly been made as suggested. However, I still have remaining questions/comments about sensitivity, specificity and discriminative validity.

Major Compulsory Revision

My question about the variables included in the regression model aimed for discriminative analysis remains. I still wonder how the BKFO-test can be sensitive in combination with the DSLL test when in the BKFO-test 85% of the subjects passed (observer A) and 82% passed (observer B) at the first test. At the second test 94% passed (observer A and B) (Table 4). (I had added the figures from test 1 and 2 and observer A in my last review). The DSLL-test was obviously more difficult to complete with required motor control as only 12% (observer A) and 18% (observer B) passed in the first test and 22% in the second test. The BKFO-test deviated considerably from the rest of the tests regarding pass or fail.

I must repeat that before the regression model is to be set up I think the authors should clearly outline how the tests responded to identify those with back pain and LE pain. To me it is not clear that the BKFO- test had that capability. From Table 4 it cannot be concluded how many of those who failed also had back or LE pain. Furthermore, when a test is not sensitive enough by itself it is hard to understand how it can be sensitive in combination with another test. If in a clinical situation one person passes in the BKFO- test and fails in the DSLL-test, what does then the BKFO- test add? Or how should these figures and findings be interpreted?

Indeed the discriminative validity was a secondary aim, but still it has been concluded that a combination of low-and high threshold tests have discriminative validity for earlier back pain which I think has not been shown in this study due to the comments above.

Kind regards,
Gertrud Nilsson

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.