Reviewer’s report

**Title:** Development and validation of a measure of health literacy in the UK: the Newest Vital Sign

**Version:** 2  **Date:** 9 October 2012

**Reviewer:** Marie-Louise Essink-Bot

**Reviewer’s report:**

The paper describes the development and validation of the UK version of the NVS, a well-known health literacy measure originally from the US. The study is well designed, well conducted and clearly described.

1. Is the question posed by the authors well defined?: YES
2. Are the methods appropriate and well described?: YES (I have few suggestions to improve the description of the data analysis, see below)
3. Are the data sound? YES
4. Does the manuscript adhere to the relevant standards for reporting and data deposition? YES
5. Are the discussion and conclusions well balanced and adequately supported by the data? YES
6. Are limitations of the work clearly stated? YES
7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished? YES
8. Do the title and abstract accurately convey what has been found? YES
9. Is the writing acceptable? YES

**Minor Essential Revisions**

1. Abstract, Methods section: it is stated that 300 participants would be required for statistical significance. My question is: significance of which comparison? Is mentioning statistical significance appropriate in this context? Furthermore I suggest to mention in the Methods section that construct validity was assessed using ROC analysis, analyzing the association between NVS-UK scores and TOFHLA-scores.

2. Methods section, paragraph on Sample and Recruitment: same comment. To me, the basis of the power analysis is unclear.

Methods section, paragraph on Data analysis:

- I think that assessing correlations between NVS and TOFHLA is a useful analysis, but that it is not an assessment of criterion validity. Criterion validity means comparison with a gold standard; I do not think that TOFHLA is a gold standard of health literacy. Correlational analysis may show the analyst the
similarity of concepts being measured.
- I have problems with the word ‘predict’, as in predictive validity, and throughout the manuscript. There is no prediction of some future characteristics using characteristics measured now. This is a cross-sectional study; the researchers investigate associations between two characteristics measured at the same point in time. The association between NVS-scores and education, and the ROC analysis, are to be regarded as analyses of construct validity.

3. Section Results and discussion, subsection Validation:
- please restrict numbers of decimal digits in SD to 1 (1.9 instead of 1.883).
- Next paragraph: please indicate whether a ‘higher’ score on TOFHLA indicates better or worse HL.
- Next paragraph: it is stated that the correlation against TOFHLA was acceptable (r = 0.49). What criterion did the authors use to judge acceptability of the size of this association? I think they may state that r=0.49 indicates about 70% of shared common variance, which seems to me a large proportion, indicating similarity of measured concepts.
- two paragraphs below: I suggest to add some explanation to the figures about sensitivity and specificity. E.g. 40% specificity in this context means (if I understand the matter correctly) that 40% of the subjects with adequate health literacy as defined by TOFHLA also have a good score on the UK-NVS; i.e., 60% of subjects with adequate HL according to the TOFHLA obtain an inadequate score on the NVS. This may indicate 60% false-positives; or it suggests that the NVS is more difficult, i.e., more strict than the TOFHLA.

4. Discussion
- I invite the authors to add their comments on the issue of comparability of UK and US NVS. Can we assume that UK NVS scores are measured on the same scale as US NVS scores; i.e. if a UK population has an average UK NVS score of 3 and a US population has an average US NVS score of 3, can we conclude that the HL levels of these populations are comparable? If this can not yet be assumed, what additional research is necessary to establish international comparability of scores?
- paragraph above ‘Conclusion’ section: The sentence starting with ‘Finally, although ….’ gets lost – some words wrong or a typo somewhere? Please correct.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I have non-financial interests in relation to this paper, because Prof Rowlands and Prof Weiss act as unpaid advisors in one of my research projects.