Author's response to reviews

Title: Assessment and application of the Chinese Resident Health Literacy Scale using item response theory in a population-based sample in south China

Authors:

Minxue Shen (shenmx1988@gmail.com)
Ming Hu (xysm2011@yeah.net)
Siyun Liu (526909968@qq.com)
Yan Chang (1434913057@qq.com)
Zhenqiu Sun (szg@xysm.net)

Version: 3 Date: 19 May 2015

Author's response to reviews: see over
Dear Ms. Zapanta,

Thank you for the opportunity to receive the reviewers’ comments and we appreciate the suggestions which help improve the manuscript greatly.

We made point-by-point responses to all comments as the followings. Moreover, we revised the manuscript in a traceable way.

We look forward to receive your further advices and decision. Thank you very much.

Minxue Shen,
On the behalf of all authors
Reviewer: Melody S Goodman

Reviewer's report:

**Minor revisions**

1. The abstract conclusion needs to be re-written for grammar and clarity for the reader.

**Reply:** Thank you for the suggestion. We have re-written the abstract conclusion for grammar and clarity consideration. The conclusion is now presented as “The instrument used in 2012 national health literacy has been validated and 64 items are selected based on CTT and IRT. The revised version of the scale has strong psychometric properties and measurement invariance is slight.”

2. Are the SBSQ the same as Single Item Literacy Screening (SILS) items that comprise the Brief Health Literacy Screen (Chew et al)? I have never heard this terminology before and suggest the authors use standard terminology when referring to existing health literacy instruments.

**Reply:** The SBSQ was developed to detect inadequate or marginal health literacy in a clinical setting ([http://www.stfm.org/fmhub/fm2004/September/Lisa588.pdf](http://www.stfm.org/fmhub/fm2004/September/Lisa588.pdf)). The content of the 16 questions was based on five domains identified in a qualitative study of patients with limited health literacy. Based on the comparison with the S-TOFHLA, three questions were selected for the final instrument based on their greater sensitivity and specificity for detecting individuals with inadequate health literacy. The Single Item Literacy Screening (SILS) is a derivative of the SBSQ, which was also developed by Lisa D Chew and her colleagues ([http://www.biomedcentral.com/1471-2296/7/21](http://www.biomedcentral.com/1471-2296/7/21)).

3. On page 5 line 5 I don’t think gregarious is the correct way to define these populations.

**Reply:** Thank you for pointing out the inaccurate word. All human are gregarious, and it is inappropriate to use the word “gregarious” to describe certain people who reside in hospitals, nursing homes, dormitory, etc. We removed the word “gregarious” in the manuscript (Page 5 line 5).
4. What is geracomium?

Reply: Thank you for pointing out the misuse. We referred to the nursing home that provides residence for people who require continual nursing care and have significant difficulty coping with the required activities of daily living. We replaced “geracomium” with “nursing home” (Page 5 line 6).

5. What is a single choice questions are authors referring to the two types of multiple choice questions? Single answer (respondent can choose only one answer) or multiple answer (where respondent can select all the options that apply)?

Reply: Thank you for pointing out the vague expression. The “single choice question” refers to single-answer question (respondent can choose only one answer); and “multiple choice question” refers to multiple-answer question (respondent can select all the options that apply). We revised the expression in the manuscript as suggested by the reviewer (Page 5 line 28).

6. On page 9 line 9 stronger health literacy skills or higher/more likely to have adequate health literacy.

Reply: Sorry, we do not clearly understand this question. We consider this as a suggested phrase and revised the sentence as “Participants in higher socio-economic status (higher educational level and more income) were more likely to have adequate health literacy” (Page 9 line 16).

7. Authors state they use a multivariate model but the model is multivariable.

Reply: Thank you for pointing out the misuse of the word. Multivariate analysis involves observation and analysis of more than one statistical outcome variable at a time. In our study, we have only one outcome, namely the health literacy. We replaced the word “multivariate analysis” with “regression” in the discussion (Page 11 line 8).

8. Several grammatical errors needs a good proof read.

Reply: Thanks for the advice. We edited the manuscript carefully, corrected the grammatical mistakes and typos, and abridged some long sentences.
9. What is ability? How is it measured? It is unclear to this reader what figure 1 is presenting

Reply: A priority of IRT is to measure the latent variable (the ability parameter). A test is administered to examinees to locate them on the ability scale, and perform comparisons among them. Estimation of ability by IRT practically but not exclusively ranges from -3 to +3. Under classical test theory, the raw score would be the sum of the scores received on all items. But under IRT, the primary interest is in whether an examinee got each individual item correct or not, rather than in the raw test score. Under IRT, maximum likelihood procedures are used to estimate an examinee’s ability. Figure 1 showed that the abilities of our participants are basically normally distributed, and most participants only have moderate and limited health literacy. We added some explanations on ability in the figure legend.

10. Line 15 on page 12 should be “research” not “researches”

Reply: Thank you for indicating the grammatical error. We replaced “researches” with “research” in the text (Page 12 line 8).

11. On line 13 of page 12 consider “represented in” instead of “in the contest of”.

Reply: Thank you for indicating the misuse. We replaced “in the contest of” with “represented in” in the manuscript (Page 12 line 6).

Discretionary revisions

12. Why did the authors choose to use so many different statistical software programs for their analysis and not just use SAS?

Reply: Item response model is not available in SAS currently. Confirmatory factor analysis is very convenient in AMOS (nested in SPSS) due to the visual programming. 2-level linear model can be replaced by the mixed model in SAS, but 3-level (or higher levels) model is not yet available in SAS, hence we are unable to calculate and compare the intra-cluster correlations within each strata of sampling. MLwiN is a statistical software package for fitting multilevel models which uses both maximum likelihood estimation and Markov Chain Monte Carlo methods. The estimation of multilevel parameters in MLwiN is stable and reliable.
Major revisions

13. When the authors state they do not have reliability >0.9 are they referring to Cronbach’s alpha? While being >0.9 is excellent values >0.8 are often considered quite good. I wanted more information on the shorter version of the measure similar to that presented for the long measure.

Reply: Thank you for the suggestion, and we are referring to the Cronbach’s alpha. The value > 0.8 is indeed good. The shorter version includes 19 items: 8 items in knowledge and attitude dimension, 5 items in behavior and lifestyle dimension, and 6 items in health-related skill dimension. The overall Cronbach’s alpha is 0.88, and alpha of three dimensions are 0.76, 0.64 and 0.77 respectively (the dimension Cronbach’s alpha are not good). The split-half coefficient is 0.87. The Pearson correlation coefficient, factor loading, and IRT parameters of the 19-item scale are presented in the following table. The Pearson correlation coefficients and factor loadings of all items are above 0.4 (and most of them are above 0.5). Discrimination parameters of all items are between 0.5 and 2.0 (and most of them are between 1.0 and 2.0), and difficulty parameters are appropriate. We added more details about the shorter version in revised manuscript.

Appendix Table. Psychometric parameters of the short 19-item scale

<table>
<thead>
<tr>
<th>Items</th>
<th>Pearson correlation coefficient</th>
<th>Factor loading</th>
<th>IRT parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Discrimination</td>
</tr>
<tr>
<td>Dimension: Knowledge and attitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B01</td>
<td>0.66</td>
<td>0.55</td>
<td>1.26</td>
</tr>
<tr>
<td>B02</td>
<td>0.62</td>
<td>0.50</td>
<td>1.21</td>
</tr>
<tr>
<td>B07</td>
<td>0.64</td>
<td>0.56</td>
<td>0.95</td>
</tr>
<tr>
<td>B26</td>
<td>0.64</td>
<td>0.56</td>
<td>1.04</td>
</tr>
<tr>
<td>B38</td>
<td>0.50</td>
<td>0.43</td>
<td>0.94</td>
</tr>
<tr>
<td>C04</td>
<td>0.67</td>
<td>0.59</td>
<td>1.05</td>
</tr>
<tr>
<td>C05</td>
<td>0.43</td>
<td>0.41</td>
<td>1.17</td>
</tr>
<tr>
<td>C07</td>
<td>0.66</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>Dimension: Behavior and lifestyle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B16</td>
<td>0.54</td>
<td>0.54</td>
<td>0.94</td>
</tr>
<tr>
<td>B27</td>
<td>0.62</td>
<td>0.54</td>
<td>0.76</td>
</tr>
<tr>
<td>C10</td>
<td>0.67</td>
<td>0.47</td>
<td>1.54</td>
</tr>
<tr>
<td>C11</td>
<td>0.73</td>
<td>0.57</td>
<td>1.46</td>
</tr>
<tr>
<td>C15</td>
<td>0.64</td>
<td>0.44</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Dimension: Skills

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B22</td>
<td>0.70</td>
<td>0.55</td>
<td>1.42</td>
<td>-0.92</td>
</tr>
<tr>
<td>B24</td>
<td>0.72</td>
<td>0.60</td>
<td>1.75</td>
<td>-0.89</td>
</tr>
<tr>
<td>B25</td>
<td>0.74</td>
<td>0.60</td>
<td>2.00</td>
<td>-0.88</td>
</tr>
<tr>
<td>B33</td>
<td>0.62</td>
<td>0.52</td>
<td>1.12</td>
<td>-1.35</td>
</tr>
<tr>
<td>C09</td>
<td>0.69</td>
<td>0.67</td>
<td>1.03</td>
<td>-0.12</td>
</tr>
<tr>
<td>D03</td>
<td>0.65</td>
<td>0.60</td>
<td>0.82</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

In addition, the item characteristic curves of the 19 items are presented below.
Reviewer: Nicola Diviani

Reviewer's report:

Thank you for the opportunity to review the manuscript “Assessment and application of the Chinese Resident Health Literacy Scale using item response theory: a population-based sample in south China”. The manuscript describes the validation of a health literacy scale that had been used in a national health literacy survey in 2012 but had not previously been validated. As validated measures of population health literacy in China are missing, I welcome the authors’ effort to validate the scale. I am sure this will also add to the growing body of evidence on the concept of health literacy and on its measurement (also beyond China). Additionally, being health literacy a set of skills of different difficulty, I find the choice of item response theory in addition to classical test theory particularly sound and relevant in this context. I think the topic addressed is relevant for BMC Public Health. However, I have a number of comments that I feel could contribute to improving the overall quality of the manuscript.

- Major Compulsory Revisions

1. Methods/Participants: the sampling strategy is not very clear. The authors could simplify the description and add a figure or diagram to illustrate the sampling process.

   **Reply:** Thank you for the advice. Sorry for the confusion especially by those sampling units. We simplified the description on sampling strategy (Page 5, Participant part) and added a figure to illustrate the process explicitly (newly added Figure 1 in revised paper).

2. Methods/Study design: it would be interesting to have some examples of the 4 different types of questions and of questions in the six different domains.

   **Reply:** Thank you for this suggestion. We added a table to provide some examples of the four types of questions (T/F, single answer, multiple answers, situation questions) in three dimensions (knowledge and attitudes; behavior and lifestyle; and health-related skills) and six different aspects of public health issues (scientific views of health; infectious disease; chronic disease; safety and first aid; medical care; and health information). Situation questions were paragraphs of health information which examined the participant’s reading and interpretation capability, and included single and multiple answers questions. The details can be found in the
3. Methods/Statistical analyses: References are missing on the different analyses conducted (particularly as regards the thresholds).

Reply: Thank you for the suggestion. References on different methods (Pearson correlation coefficient, factor loading in confirmatory factor analysis, and difficulty & discrimination parameters in item response theory) for item selection criterion (threshold) have been added to the paper (Page 7, line 4, 10 & 16). The details of items selection thresholds are explained in the reply to the review’s comment 5.

4. Results: Report alpha for the three dimensions.

Reply: Thank you pointing out the vague data. The Cronbach’s alpha for the three dimensions were 0.90 (knowledge and attitude, 38 items), 0.83 (behavior and lifestyle, 22 items) and 0.85 (skills, 20 items) respectively. After item selection (16 items were removed), the Cronbach’s alpha for the three dimensions were 0.90 (knowledge and attitude, 30 items), 0.83 (behavior and lifestyle, 16 items) and 0.86 (skills, 18 items) respectively. We have reported these details in the revised manuscript (Page 8, line 9 & 26).

5. Results: More detailed information is needed on the reasons why items were removed from the scale.

Reply: Thank you for the advice. We used 3 methods in item selection procedure: Pearson correlation coefficient, confirmatory factor analysis and item response theory.

1 Pearson correlation coefficient assessed representativeness and independency. An eligible items should be significantly and at least moderately correlated to total score of its domain. We used the value 0.4 as the criteria since a correlation coefficient between 0.4 and 0.7 means moderate correlation (Richard, 1990, p37).

2 Confirmatory factor analysis tests whether the variance-covariance of items conforms to an anticipated or expected scale structure given in a particular research hypothesis. The factor loading 0.4 has been recommended by several researchers (Gerbing & Anderson, 1998, p189; Gorsuch, 1997, p545; Velicer & Fava, 1998, p234).
Item response theory is used to assess the psychometric properties of instruments, and items could be selected on the basis of contents as well as parameter values. In our study, we used a criteria of parameter constraints given by Frank B. Baker (The basics of item response theory, 2nd edition, 2001, p159): difficulty parameter (b) between -3 and 3; discrimination parameter (a) between 0.5 and 2; and guessing parameter (c) between 0 and 0.35. We did not consider the guessing parameter because we set “I do not know” choices for all questions, and participants did not need to guess if they did not know the answer. Items with higher discrimination parameters usually provide more information, but when this parameter gets surprisingly high (i.e. greater than 2.5, commonly seen in clinical and psychological studies), the measured construct is often conceptually narrow (as reviewed by Steven P. Reise, Annu Rev Clin Psychol, 2009, p30). But health literacy is a broad concept with greater content heterogeneity, hence the discrimination should be greater than 0.5 but not exceed 2.0 as indicated by Baker (poor discrimination :0~0.5; moderate: 0.5~1.0; good: 1.0~1.5; excellent: 1.5~2.0). We added these details in the revised paper (Page 7, line 4, 10 & 16).

6. Discussion: The first part of the discussion could be more informative (e.g., clearly repeating the aims of the study and summarizing the main steps undertaken in this context).

Reply: Thank you for the comment. We abridged some of the contents introducing the definition of health literacy, and summarized the aims, methods and main findings of our study in the beginning of the discussion (Page 10 Paragraph 1).

7. Discussion: The part about IRT belongs to the methods section. Here the authors should only summarize it.

Reply: Thanks for the advice. We simplified the description of IRT in the discussion section, and added some details of it in the methods section (Page 10 line 16).

8. Discussion: The authors should discuss the implication of their results (e.g., the fact that some items were dropped) for the results of the 2012 survey. For instance, would having used the validated version of the scale have changed the main conclusions of the survey?
Reply: We consider that there are several implications of the study. Firstly, we found that the original scale is good in reliability and validity. We removed 16 items according to factor analysis and IRT, and the raw scores of the 64-item scale are highly correlated to the scores of the original scale. As a result, the main conclusion of the survey would not change after validation of the scale. Second, we made a shorter version of the scale (19 items) because the original scale is time-consuming. The 19-item version is slightly inferior to the original scale in reliability (Cronbach’s alpha decreased from 0.95 to 0.88) but still useful in population studies. Third, the instruments used in national health literacy survey in 2008 and 2012 are different. In the government report, those longitudinal results (mainly in form of raw scores and percentage of people whose raw scores exceeded a threshold) were compared directly, which is apparently inappropriate. IRT used in our study provided an opportunity for the longitudinal comparison since the ability is estimated according to the participants’ responses to items with different difficulty. Those tests can be linked and ability can be compared using the IRT method. In the end of the discussion, we reported the first and the second implications (Page 11 line 23). But for the third one, we still need to do further work.

- Minor Essential Revisions

9. Language needs careful editing. There are several (too) long sentences and a number of grammatical mistakes (e.g., “the scale have” instead of “has”) and typos (e.g., CAF instead of CFA). I suggest the authors have the manuscript checked edited by an English native speaker.

Reply: Thanks for pointing out these mistakes and typos. We edited the manuscript meticulously and corrected the grammatical mistakes and abridged long sentences.

- Discretionary Revisions

10. Title: I suggest to replace the “:" with “in” for better clarity.

Reply: Thank you for the suggestion and we accepted it.

11. Table 1 and 3: Replace “Illiteracy” with “No formal education”

Reply: Thank you for the advice. “No formal education” is more accurate and we replaced “illiteracy” with the suggested phrase (revised paper Table 2 and 4).