

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Questionnaire validation practice within a theoretical framework: a systematic descriptive literature review of health literacy assessments
<b>AUTHORS</b>	Hawkins, Melanie; Elsworth, Gerald; Hoban, Elizabeth; Osborne, Richard

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Rocco Palumbo University Rome "Tor Vergata", Italy
<b>REVIEW RETURNED</b>	18-Dec-2019

<b>GENERAL COMMENTS</b>	<p>Dear Authors,</p> <p>Thank you very much for the opportunity to read this interesting contribution. I agree with your point that theoretically-driven frameworks of validation practice for the development, testing and use of health assessments are missing in the scientific and professional literature. Therefore, your article makes a significant contribution to the advancement of scientific knowledge.</p> <p>Whilst I commend the publication of your paper, I also found some minor spaces for improvement, that are reported below.</p> <p>Firstly, I think that the current title does not effectively stress the originality and the added-value of your study. Since you are aiming to propose a theoretical validity testing framework to inform and improve the processes used to develop and test health assessments, this purpose should appear as the key point in your title.</p> <p>Secondly, I found that the introduction is not able to set the hook for potential readers. In line with the main focus of your research, consider to kick-off your article with a brief, but effective presentation of "health literacy", stressing that validation practices of health literacy assessment tools are negatively affected by the lack of theoretical validity testing frameworks. Moreover, the introduction falls short in emphasizing the originality of your research: to improve it, please consider to contextualize the research questions to the state of the art of scientific literature in the field of health literacy.</p> <p>Thirdly, I think that the first part of the results section (from page 7, line 50 to page 10, line 36) would perform better if included in an autonomous section (or subsection), that should be intended to provide an overview of records involved in the systematic literature review.</p> <p>Fourthly, in my own opinion, you should avoid referring the research questions in labeling the sub-sections of the findings section: please, consider using more appealing and provocative labels that - of course - should address the two research questions that triggered your research.</p>
-------------------------	---

	<p>Fifthly - and lastly - whilst the discussion section is clear and effective, conclusions are too short and do not provide adequate insights into the (conceptual and practical) implications of your research; also, they do not emphasize the main take-aways of your research. Please, consider improving the conclusion section, including more insights into the overall contribution of your article. Once again, thank you very much for the opportunity to read this interesting piece of literature.</p> <p>Yours sincerely, The reviewer</p>
--	--

<b>REVIEWER</b>	Mhairi Campbell University of Glasgow, UK
<b>REVIEW RETURNED</b>	23-Jan-2020

<b>GENERAL COMMENTS</b>	<p>Thank you for the opportunity to review this paper, which I read with interest. The paper is clear and well written. Please see some points of clarification comments below.</p> <p>Abstract – conclusions line 48 “The validity testing framework of the Standards facilitates examination of evidence based on five sources to determine the validity of inferences derived from health assessment data.” Is this a conclusion of this review? Would the space be better used to refer to the need for, or benefit of, adherence to the standards framework?</p> <p>Methods</p> <p>Page 6 Line 35 – (minor point) Suggest changing ‘conducted’ to ‘reported in accordance with PRISMA’, as PRISMA is a reporting tool, and not for guiding conduct of a review.</p> <p>Page 6 Line 40 – (minor point) In table 2, was grey literature, (any reports not published in a peer reviewed journal), included or excluded? This should be reported.</p> <p>Page 7 Line 20 – It’s unclear what proportion of articles were data extracted by a second author, please report the proportion.</p> <p>Page 7 Line 54 (minor point) ‘There were 1,922 records when duplicates were removed.’ Perhaps this could be rephrased, as some readers may be confused, and think an unknown number of duplicates were removed from the 1922 records. Something like ‘1922 articles remained after xxx duplicates were removed’, or ‘There were 1,922 records after duplicates were removed’.</p> <p>Results</p> <p>Page 8 Line 15 – Since attention is drawn to the four papers, it would be helpful to note where these were found, was this through purposive searching or from the review authors’ personal collection of articles – papers you knew about due to your expertise?</p> <p>If these four papers exist, there could be more like this. I’m not suggesting further searching is required. However, it would be useful to note in the limitations section that other such papers (where the authors do not claim/identify an assessment, but you recognise that it is) could exist that were not identified (could not be identified?) in the systematic searching. This would explain why any similar papers were not included in your review.</p>
-------------------------	---

	<p>Discussion</p> <p>Page 13 line 8 – Discussion section – the first paragraph explaining the results may be strengthened by a summary table being included in the results section that shows how few studies referred to a framework. Appreciate that with 46 studies, the included table would be large, however a simplified version of the supplementary file 3 table (possibly using symbols such as ticks and crosses) noting which studies did and did not refer to a framework or the key components of the Standards framework, would let the review reader see the lack of use. It is good practice to present the key characteristics of the included studies. It would also be helpful for the studies in the table to be ordered by what the synthesis focussed on, rather than alphabetically (perhaps listing studies that directly referred to a framework, then those that indirectly referred, then all those that did not refer to a framework?)</p> <p>Page 15 – Most of the paragraphs of the description of how the studies were coded seems to be more suited to the methods or results section rather than the discussion. And then any implications of the coding process considered in the discussion.</p>
--	--

## VERSION 1 – AUTHOR RESPONSE

### Response to feedback from Reviewer 1

#### 2.1 Overall

*Dear Authors, Thank you very much for the opportunity to read this interesting contribution. I agree with your point that theoretically-driven frameworks of validation practice for the development, testing and use of health assessments are missing in the scientific and professional literature. Therefore, your article makes a significant contribution to the advancement of scientific knowledge. Whilst I commend the publication of your paper, I also found some minor spaces for improvement, that are reported below.*

We thank you for your supportive comments about our paper and the scientific contribution it will make.

#### 2.2 Title

*Firstly, I think that the current title does not effectively stress the originality and the added-value of your study. Since you are aiming to propose a theoretical validity testing framework to inform and improve the processes used to develop and test health assessments, this purpose should appear as the key point in your title.*

We agree with this comment about the title and have revised the title accordingly. It was:

Validation practice for health literacy assessments: a systematic descriptive literature review using a theoretical validity testing framework

It is now:

Questionnaire validation practice within a theoretical framework: a systematic descriptive literature review of health literacy assessments

#### 2.3 Introduction

*Secondly, I found that the introduction is not able to set the hook for potential readers. In line with the main focus of your research, consider to kick-off your article with a brief, but effective presentation of*

*"health literacy", stressing that validation practices of health literacy assessment tools are negatively affected by the lack of theoretical validity testing frameworks.*

We agree with the reviewer and realise that health literacy should be mentioned much earlier in the introduction. In revision, we have:

1. Altered the third sentence in the first paragraph of the Introduction to mention health literacy assessments  

(p.4) Interpretations of scores from health literacy assessments are increasingly being used to make decisions about the design, selection and evaluation of interventions and policies to improve health equity for individuals, communities and populations. [2-4, 8, 9]
2. Moved the whole health literacy section of the Introduction to below the first paragraph in the introduction, with this sentence (below) added to the end of the paragraph to stress that interpretation of data from health literacy assessments for specific clinical or research purposes would be improved by use of a theoretical validity testing framework:  

(p.4) However, despite the different definitions that health literacy assessments are based on (and thus, necessarily, the different score interpretations and uses), the data are often correlated and compared as if the interpretation of the scores have the same meaning. [27] A theoretical validity testing framework would help researchers, clinicians and policy makers to differentiate between the meanings of data from different health literacy assessments, and evaluate existing evidence to support data interpretations, to enable them to choose the assessment that is most appropriate for their intended clinical or research purpose.

*Moreover, the introduction falls short in emphasizing the originality of your research: to improve it, please consider to contextualize the research questions to the state of the art of scientific literature in the field of health literacy.*

Although we appreciate the thinking this feedback caused us, we conclude that it is difficult to respond to it because this literature review *is* assessing the state of art of scientific literature in the field of health literacy. We cannot talk about the state of the art of health literacy assessment until we have done the review. However, we have added the following sentence in the Rationale paragraph in an attempt to address the reviewer's comments:

(p.6) Health literacy is a relatively new area of research that appears to have proceeded with the 'types of validity' paradigm of early validation practice in education, and so it is ideally poised to embrace advancements in validity testing practices.

Also, the second research question was altered slightly to direct it more toward the context of health literacy. The research question was:

Do the studies place the validity evidence within a validity testing framework, such as that offered by the *Standards*?

It now reads:

(p.7) Is the validity evidence currently provided for health literacy assessments placed within a validity testing framework, such as that offered by the *Standards*?

## **2.4 Results**

*Thirdly, I think that the first part of the results section (from page 7, line 50 to page 10, line 36) would perform better if included in an autonomous section (or subsection), that should be intended to provide an overview of records involved in the systematic literature review.*

We are not sure we understand what the requirement is here but we think the reviewer is asking for the details of each study – e.g., details about the participants etc in each study. These details are not relevant to this review because the focus is not on the participants in the studies but on the evidence presented in support of the validity of interpretations and uses of health literacy assessment data. Information about each study relevant to the review is provided in Supplementary File 3 (Data

extraction framework) – authors, reference to a validity testing framework, the country in which the validation study was undertaken, the health literacy assessment and any comparator instruments, and validity evidence as reported within the five sources of evidence. As mentioned in section 3.4 further down in this document, Supplementary File 3 has been sorted to be more clear. We hope this helps to answer the reviewer’s concerns.

We have included this sentence at the beginning of the results section to help orient the reader:

(p.8) Overall, 46 articles were identified for the review.

*Fourthly, in my own opinion, you should avoid referring the research questions in labeling the sub-sections of the findings section: please, consider using more appealing and provocative labels that - of course - should address the two research questions that triggered your research.*

We agree. This was an oversight. Please see the changes to these labels on p.10 and p.13:

(p.10) Validity evidence for health literacy assessment data

(p.13) Use of a validity testing framework when reporting validity evidence for health literacy assessments

## **2.5 Conclusion**

*Fifthly - and lastly - whilst the discussion section is clear and effective, conclusions are too short and do not provide adequate insights into the (conceptual and practical) implications of your research; also, they do not emphasize the main take-aways of your research. Please, consider improving the conclusion section, including more insights into the overall contribution of your article.*

We have revised the Conclusions section in light of this feedback. It now reads:

(p.17) Arguments for the validity of decisions based on health assessment data must be based on evidence that the data are valid for the decision purpose to ensure the integrity of the consequences of the measurement, yet this is frequently overlooked. This literature review demonstrated the use of the *Standards'* validity testing framework to collate and assess existing evidence and identify gaps in the evidence for health literacy assessments. Potentially, the framework could be used to assess the validity of data interpretation and use of other health assessments in different contexts. Developers of health assessments can use the *Standards'* framework to clearly outline their measurement purpose, and to define the relevant and appropriate validity evidence needed to ensure evidence-based, valid and equitable decision making for health. This view of validity being about score interpretation and use challenges the long-held view that validity is about the properties of the assessment instrument itself. It is also the basis for establishing a sound argument for the authority of decisions based on health assessment data, which is critical to health services research and to the health and health equity of the populations affected by those decisions.

## **3. Response to feedback from Reviewer 2**

*Thank you for the opportunity to review this paper, which I read with interest. The paper is clear and well written. Please see some points of clarification comments below.*

We appreciate the positive feedback from Reviewer 2. Thank you.

### **3.1 Abstract**

*Abstract – conclusions line 48 “The validity testing framework of the Standards facilitates examination of evidence based on five sources to determine the validity of inferences derived from health assessment data.” Is this a conclusion of this review? Would the space be better used to refer to the need for, or benefit of, adherence to the standards framework?*

Thank you, this is useful feedback. We agree the abstract conclusion needs to be refined. We have re-written it with your advice in mind. However, this revised conclusion put us over the 300 word limit for the Abstract so we needed to revise other sentences in the abstract. The abstract now reads:

(p.2-3) **Objective** Validity refers to the extent to which evidence and theory support the adequacy and appropriateness of inferences based on score interpretations. The health sector is lacking a theoretically-driven framework for the development, testing and use of health assessments. This study used the Standards for Educational and Psychological Testing framework of five sources of validity evidence to assess the types of evidence reported for health literacy assessments, and to identify studies that referred to a theoretical validity testing framework.

**Methods** A systematic descriptive literature review investigated methods and results in health literacy assessment development, application and validity testing studies. Electronic searches were conducted in EBSCOhost, EMBASE, Open Access Theses and Dissertations, and ProQuest Dissertations. Data were coded to the Standards' five sources of validity evidence, and for reference to a validity testing framework.

**Results** Coding on 46 studies resulted in 195 instances of validity evidence across the five sources. Only nine studies directly or indirectly referenced a validity testing framework. Evidence based on relations to other variables is most frequently reported.

**Conclusions** The health and health equity of individuals and populations are increasingly dependent on decisions based on data collected through health assessments. An evidence-based theoretical framework provides structure and coherence to existing evidence and stipulates where further evidence is required to evaluate the extent to which data are valid for an intended purpose. This review demonstrates the use of the Standards' theoretical validity testing framework to evaluate sources of evidence reported for health literacy assessments. Findings indicate that theoretical validity testing frameworks are rarely used to collate and evaluate evidence in validation practice for health literacy assessments. Use of the Standards' theoretical validity testing framework would improve evaluation of the evidence for inferences derived from health assessment data on which public health and health equity decisions are based.

### 3.2 Methods

*Page 6 Line 35 – (minor point) Suggest changing 'conducted' to 'reported in accordance with PRISMA', as PRISMA is a reporting tool, and not for guiding conduct of a review.*

Thank you for picking up this detail. It has been corrected as suggested and now the sentence reads:

(p.7) The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

*Page 6 Line 40 – (minor point) In table 2, was grey literature, (any reports not published in a peer reviewed journal), included or excluded? This should be reported.*

Grey literature and reports were excluded from this review. This has been made clear in Table 2 (p.7).

*Page 7 Line 20 – It's unclear what proportion of articles were data extracted by a second author, please report the proportion.*

The authors MH and GRE were in close collaboration during the selection and data extraction processes. GRE, as second reviewer, read the complete text of 10% or 9 of the 92 articles selected for full text review. We have now included this in the manuscript, with an additional sentence to indicate the close collaboration between the reviewers during screening:

(p.8) Identified full text articles (n=92) were screened for relevance by MH and corroborated with an independent screening of 10% (n=9) of the search results by a second author (GRE). Additionally, MH consulted closely with GRE when a query arose about inclusion of an article in the review.

GRE comprehensively and independently checked the data extraction for all 46 articles. This has been clarified in the text:

(p.8) Data extraction from articles for final inclusion was undertaken by one author (MH) with all data extraction comprehensively and independently checked by a second author (GRE).

*Page 7 Line 54 (minor point) 'There were 1,922 records when duplicates were removed.' Perhaps this could be rephrased, as some readers may be confused, and think an unknown number of duplicates were removed from the 1922 records. Something like '1922 articles remained after xxx duplicates were removed', or 'There were 1,922 records after duplicates were removed'.*

We agree and have altered the text accordingly:

(p.8) There were 1,922 records remaining after 1457 duplicates were removed.

### **3.3 Results**

*Page 8 Line 15 – Since attention is drawn to the four papers, it would be helpful to note where these were found, was this through purposive searching or from the review authors' personal collection of articles – papers you knew about due to your expertise?*

The text has been adjusted to include this information:

(p.9) Four papers were identified from the broader literature. Two papers were identified from the references of previous literature reviews [80, 81]. The other two papers were known to the authors and were in their personal reference lists. These two papers were by Davis and colleagues and describe the development of the Rapid Estimate of Adult Literacy in Medicine (REALM) [33] and the shortened version of the REALM. [82]

*If these four papers exist, there could be more like this. I'm not suggesting further searching is required. However, it would be useful to note in the limitations section that other such papers (where the authors do not claim/identify an assessment, but you recognise that it is) could exist that were not identified (could not be identified?) in the systematic searching. This would explain why any similar papers were not included in your review.*

We agree this could be possible and have added a short paragraph to the limitations:

(p.17) Just as there were two papers known to the authors of an instrument that is frequently used to measure health literacy, and two further papers were identified from published literature reviews, it may be that more papers that would be relevant to this review were not identified. However, since the 1991 publication of the REALM, which was not designed as a health literacy assessment but has since been used as such, we predict that most assessments for the measurement of health literacy will be identified for this purpose, and would thus have been captured by the present search strategy.

### **3.4 Discussion**

*Page 13 line 8 – Discussion section – the first paragraph explaining the results may be strengthened by a summary table being included in the results section that shows how few studies referred to a framework. Appreciate that with 46 studies, the included table would be large, however a simplified version of the supplementary file 3 table (possibly using symbols such as ticks and crosses) noting which studies did and did not refer to a framework or the key components of the Standards framework, would let the review reader see the lack of use. It is good practice to present the key characteristics of the included studies. It would also be helpful for the studies in the table to be ordered by what the synthesis focussed on, rather than alphabetically (perhaps listing studies that directly referred to a framework, then those that indirectly referred, then all those that did not refer to a framework?)*

Unless we are misunderstanding the request of Reviewer 2, we have already included the studies that referred to a validity testing framework in Supplementary File 3. The studies that *did not* refer to a framework are marked with an x. When a study did directly or indirectly refer to a framework, we have cited the framework in that column. The table has now been sorted such that the studies that referenced a validity testing framework (direct then indirect) are at the top of the table. We have also

moved this column to the second column in the table to make it more prominent. We have referred the reader to Supplementary File 3 in the results section:

(p.13) Of the 46 studies, 9 directly or indirectly referenced a validity testing framework, and made a statement to support the citation (see Supplementary File 3).

*Page 15 – Most of the paragraphs of the description of how the studies were coded seems to be more suited to the methods or results section rather than the discussion. And then any implications of the coding process considered in the discussion.*

The authors carefully discussed this comment from Reviewer 2 but we concluded that this section is discussing the methods, not reporting them objectively, as would be done in the Methods section. We prefer to leave these as part of the Discussion.

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Rocco Palumbo University Rome "Tor Vergata", Italy
<b>REVIEW RETURNED</b>	27-Mar-2020

<b>GENERAL COMMENTS</b>	Dear Authors, Thank you very much for this revised manuscript and for your effort to address all the concerns that I had with the original version of your article. I can see that my remarks have been included in the revised articles. I think that your research will make an interesting contribution to the advancement of scientific knowledge and, therefore, I welcome its publication in BMJ OP
-------------------------	---

<b>REVIEWER</b>	Mhairi Campbell University of Glasgow
<b>REVIEW RETURNED</b>	26-Mar-2020

<b>GENERAL COMMENTS</b>	Thank you for your responses and the revisions made in the manuscript, I I have no further comments.
-------------------------	--