

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

Melanie Hawkins ^{1,2}, Gerald R Elsworth,^{1,2} Elizabeth Hoban,¹ Richard H Osborne²

To cite: Hawkins M, Elsworth GR, Hoban E, *et al*. Questionnaire validation practice within a theoretical framework: a systematic descriptive literature review of health literacy assessments. *BMJ Open* 2020;**10**:e035974. doi:10.1136/bmjopen-2019-035974

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-035974>).

Received 28 November 2019
Revised 09 March 2020
Accepted 27 March 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Faculty of Health, Deakin University School of Health and Social Development, Burwood, Victoria, Australia

²Centre for Global Health and Equity, Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, Victoria, Australia

Correspondence to

Dr Melanie Hawkins;
melaniehawkins@swin.edu.au

ABSTRACT

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.

BACKGROUND

It has been argued that the health sector is lacking a theoretically-driven framework

Strengths and limitations of this study

- This is the first time a theoretical validity testing framework, the five sources of evidence from the *Standards for Educational and Psychological Testing*, has been applied to the examination of validity evidence for health literacy assessments.
- A strength of this study is that validity is clearly defined, in accordance with the authoritative validity testing literature, as the extent to which theory and evidence (quantitative and qualitative) support score interpretation and use.
- A limitation was the restriction of the search to studies and health literacy assessments published or administered in English, which may introduce an English language and culture bias to the sample.
- A further limitation was the lack of clarity in some papers about the methods used and results obtained, leading to difficulties in coding validity evidence and may have led to some misclassification of reported evidence for some papers.

of validation practice for the development, testing and use of health assessments.^{1–6} Such a framework could guide and strengthen validation planning for the interpretation and use of health assessment data.^{2,3,7} 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} To ensure that decisions based on data from all health assessments are justified, and lead to equitable outcomes, validation practice must generate information about the degree to which the intended interpretations and use of data are supported by evidence and the theory of the construct being measured.^{10–19} Validation research is complex^{7,20} and a theoretical framework would facilitate an evaluation of a range of

evidence to determine valid interpretation and use of health assessment data.^{2 4 18 20 21}

Health literacy

Health literacy is a relatively new field of research with a range of definitions for different settings^{22–25} and advances in the approaches to its measurement.^{26–32} Some health literacy assessments measure an observer's (eg, clinician's or researcher's) observations of a person's health literacy, which often consists of testing a person's health-related numeracy, reading and comprehension.^{33 34} Objective measurement can support a clinician to provide health information in formats and at reading levels that are suited to individual patients but usually these measures do not assess other important dimensions of the health literacy construct.³⁵ Self-report measures of health literacy have become useful with the rise of the patient-centred healthcare movement, and these typically provide individuals' perspectives of a range of aspects of their health and health contexts.^{23 36} This type of measurement can capture the multidimensional aspects of the health literacy construct to look at broader implications of treatment, care and intervention outcomes.³⁷ Assessments could also combine both objective and self-report measurement of health literacy. Data from health literacy assessments have been used to inform health literacy interventions^{8 19 38–41} and, increasingly, health policies.^{42–46} 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, which is an incorrect assumption.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.

Contemporary validity testing theory

The validity testing framework of the 2014 *Standards for Educational and Psychological Testing* (the *Standards*) is the authoritative text for contemporary validity testing theory.⁵ It results from about 100 years of the evolution of validity theory.^{47 48} The *Standards* defines validity as 'the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests' (p.11) and validation as the process of '...accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations' (p.11). The framework describes five types of validity evidence that can be evaluated to justify test score interpretation and use: (1) *test content*, (2) *response processes* of respondents and users, (3) *internal structure* of the assessment test, (4) *relations to other variables* and (5) *consequences* of testing, as related to validity (table 1).^{5 6 49 50} Evidence from each of these sources may be needed to verify data interpretation and use.

The expectation of the *Standards* and leading validity theorists is that the validation process consists of an evaluative integration of different types of validity evidence (not types of validity) to support score meaning for a specific use.^{2 4 5 13–15 51–57} Integral to this framework are quantitative methods to evaluate an assessment's statistical properties, but also important is validity evidence based on qualitative research methods.^{4 58–65} Qualitative methods are used to ensure technical evidence for *test content* and *response processes*, and to investigate validity-related *consequences* of testing.^{7 12 52 63–69} There are guides to assess quantitative measurement properties^{70–72} but still needed are reviews that include qualitative validity evidence, and that place validity evidence for health assessments within a validity testing framework such as the *Standards*.^{2 4 6 49}

Rationale

As a guide to inform and improve the processes used to develop and test health assessments, this review will examine validation practice for health literacy assessments. Health literacy is a relatively new area of research

Table 1 The five sources of validity evidence^{5 49}

(1) Evidence based on test content.	The relationship of the item themes, wording and format with the intended construct, including administration process.
(2) Evidence based on response processes.	The cognitive processes and interpretation of items by respondents and users, as measured against the intended construct.
(3) Evidence based on internal structure.	The extent to which item interrelationships conform to the intended construct.
(4) Evidence based on external variables.	The pattern of relationships of test scores to external variables as predicted by the intended construct.
(5) Evidence based on the consequences of testing.	Intended and unintended consequences, as can be traced to a source of invalidity such as construct under-representation or construct-irrelevant variance.

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. Thus, an assumption underlying this review is that the field of health is not applying contemporary validity testing theory to guide validation practice, and that the focus of validation studies remains on the general psychometric properties of a health assessment rather than on the interpretation and use of scores. This study will provide an example of the application of the *Standards’* theoretical validity testing framework through the review of sources of validity evidence (generated through quantitative and qualitative methods) reported for health literacy assessments.

The aim of this systematic descriptive literature review was to use the validity testing framework of the *Standards* to categorise and count the sources of validity evidence reported for health literacy assessments and to identify studies that used or made reference to a theoretical validity testing framework. Specifically, the review addressed the following questions:

1. What is being reported as validity evidence for health literacy assessment data?
2. Is the validity evidence currently provided for health literacy assessments placed within a validity testing framework, such as that offered by the *Standards*?

METHODS

King and He situate systematic descriptive literature reviews toward the qualitative end of a continuum of review techniques.⁷³ Nevertheless, this type of review employs a frequency analysis to categorise qualitative and quantitative research data to reveal interpretable patterns.^{32 73–78} This review will appraise validation practice for health

literacy assessments using the *Standards’* framework of five evidence sources. It will not critique nor assess the quality of individual health literacy assessments or studies.

Inclusion and exclusion criteria, information sources and search strategy

The method for this review was previously reported in a protocol paper.⁴⁹ The eligibility and exclusion criteria, information sources and search terms are summarised in [table 2](#). Peer reviewed full articles and examined theses were included in the search. Online supplementary file 1 shows the MEDLINE database search strategy, and this was modified for the other databases. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.⁷⁹ See online supplementary file 2 for the PRISMA checklist.

Article selection, and data extraction, analysis and synthesis

Duplicates were removed and a title and abstract screening of identified articles was performed in EndNote Reference Manager X9 by one author (MH). 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 with GRE when a query arose about inclusion of an article in the review.

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). Both authors then corroborated to achieve categorisation consistency. General characteristics for each study were extracted but of primary interest were the sources of validity evidence reported, as were statements about or references to a theoretical validity testing

Table 2 Summary of inclusion and exclusion criteria, information sources and search terms

Inclusion criteria	Exclusion criteria
Not limited by start date: end date March 2019	Systematic reviews and other types of reviews, grey literature (ie, any studies or reports not published in a peer-reviewed journal)
Development, application and validity testing studies and examined theses about health literacy assessments	Health literacy assessments designed for specific demographic groups (eg, children) or health conditions (eg, kidney disease)
All definitions of health literacy; and objective, subjective, unidimensional and multidimensional health literacy assessments	Predictive, association or other comparative studies that do not claim in the abstract to contribute validity evidence
Studies published and health literacy assessments developed and administered in the English language	Health literacy assessments developed or administered in languages other than English*
Qualitative and quantitative research methods	Translation studies
Information sources: EBSCOhost (MEDLINE Complete, Global Health, CINAHL Complete, PsycINFO, Academic Search Complete); Embase; Open Access Theses and Dissertations; ProQuest Dissertations; references of relevant systematic reviews; authors’ reference lists	
Search terms: Medical subject headings (MeSH) and text words - valid*, verif*, ‘patient reported outcome*’, questionnaire*, survey*, ‘self report*’, ‘self rated’, assess*, test*, tool*, ‘health literacy’, measure*, psychometric*, interview*, ‘think aloud’, ‘focus group*’, ‘validation studies’, ‘test validity’	

*See *Results* for exceptions.

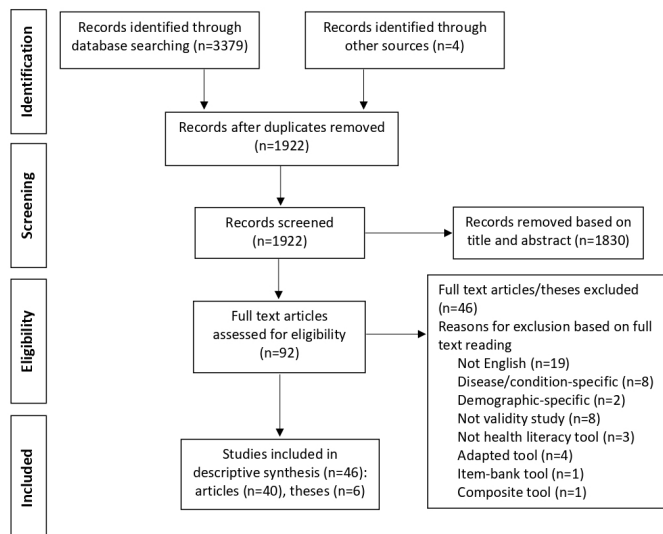


Figure 1 Flow diagram for Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

framework. The validity evidence reported in each article was categorised according to the five sources of validity evidence in the *Standards*, whether or not the authors of the articles reported it that way. When the methods were unclear, the results were interpreted to determine the type of evidence generated by the study. A study was categorised as using or referencing a theoretical validity testing framework if the authors made a statement that referred to a framework and directly cited the framework document or if there was a clear citation path to the framework document.

Descriptive and frequency analyses of the extracted data were conducted to identify patterns in the sources of validity evidence being reported, and for the number of studies that made reference to a validity testing framework.

Patient and public involvement

Patients and the public were not involved in the development or design of this literature review.

RESULTS

Overall, 46 articles were identified for the review. The PRISMA flow diagram in [figure 1](#) summarises the results of the search.⁷⁹ There were 3379 records identified through database searches with four articles identified through other sources. There were 1922 records remaining after 1457 duplicates were removed. After applying the exclusion and inclusion criteria to all abstracts, with full text screening of 92 articles and theses, 40 articles and 6 theses were included in the review (n=46). Reasons for exclusion were that the health literacy assessment was developed in or administered in a language other than English (n=19); the assessment was specific to a disease or condition (n=8) or to a demographic group (n=2); the article was not a validity study (n=8); the study was not using a health literacy assessment (n=3) or used an adapted assessment

(n=4); the assessment was based on an item-bank, which required a different approach to validity testing (n=1), or was a composite assessment where health literacy data were collected and analysed with another type of data (n=1).

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 Neither of these papers were detected by the systematic review because Davis *et al* do not claim these to be measures of health literacy but of literacy in medicine. They state that both versions of the REALM are designed to be used by physicians in public health and primary care settings to identify patients with low reading levels.^{33 82–84} Nevertheless, we included these papers because the REALM and the shortened REALM have been used by clinicians and researchers as measures of health literacy, and are used either as the primary assessment or a comparator assessment in many studies.

Three papers identified in the database search were included in this review even though data were collected using translations of assessments originally developed in English. These studies were included because of the frequency of use of these assessments in the field of health literacy measurement, and because at least part of the data were based on English language research. The Test of Functional Health Literacy in Adults (TOFHLA)⁸⁵ and the Newest Vital Sign (NVS)³⁴ both collected data in English and Spanish. The analyses for the European Health Literacy Survey (HLS-EU) study²³ used data from the English (Ireland), as well as Dutch and Greek versions of the HLS-EU.

Of the 46 studies, 34 were conducted in the USA, 8 in Australia, 2 in Singapore and 1 each in Canada and the Netherlands. There were 4 studies published in the decade between 1990 and 1999, 8 studies between 2000 and 2009 and 34 between 2010 and 2019.

Reports of reliability evidence were provided in 33 studies (72%). This resulted in 44 instances of reliability evidence, of which 29 (66% of all instances) were calculated using Cronbach's alpha for internal consistency, 4 (9% of all instances) using test-retest, 4 (9%) using inter-rater reliability calculations and 7 (16%) using other methods. See [table 3](#) for country and year of publication, and reliability evidence.

Validity evidence for health literacy assessment data

The data extraction framework (online supplementary file 3) was adapted from Hawkins *et al* (p.1702)⁶ and Cox and Owen (p.254).⁵⁸ More detailed sub-coding of the five *Standards'* categories was done and will be drawn on selectively to describe aspects of the results (online supplementary file 4).

Table 3 Country and year of publication, and reliability evidence

	N	%
Country of study		
USA	34	74
Australia	8	17
Singapore	2	4
Canada	1	2
Netherlands	1	2
Year of publication by decade		
1990–1999	4	9
2000–2009	8	17
2010–2019	34	74
Reliability		
Cronbach's alpha	29	66
Test-retest	4	9
Inter-rater	4	9
Other methods	7	16
Total instances of reliability	44	100

Data analysis consisted of coding instances of validity evidence into the five sources of validity evidence of the *Standards*. The results of the review are presented as: (1) the total number of instances of validity evidence for each evidence source reported across all studies, (2) the number of instances reported for objective, subjective and mixed methods health literacy assessments and (3) the number of instances of evidence within each of the *Standards'* five sources, and a breakdown of the methods used to generate evidence.

Table 4 displays the overall results of the review. For the 46 studies that reported validity evidence for health literacy assessments, we identified 195 instances of validity evidence across the five sources: *test content* (n=52), *response processes* (n=7), *internal structure* (n=28), *relations to other variables* (n=107) and *consequences of testing* (n=1). Across types of health literacy assessments, there were 102 instances of validity evidence reported for health literacy assessments with an objective measurement approach (n=23 studies); 78 instances reported for assessments with a subjective measurement approach (n=20 studies) and 15 instances for assessments with a mixed methods approach or when multiple types of health literacy assessments were under investigation (n=3 studies).

Evidence based on test content

Nearly half of all studies (n=22) reported evidence based on *test content*, which resulted in 52 instances of validity evidence (**table 4** and online supplementary table 1). Expert review was the most frequently reported method used to generate evidence (n=14 instances; 27% of all evidence based on *test content*),^{23 33 34 36 82 83 86–93} followed by the use of existing measures of the construct (n=8; 15%).^{34 36 83 90–92 94 95} Analysis of item difficulty was used five times (10%),^{36 86 89 92 96} with literature reviews,^{23 90 93 97} participant feedback processes about items^{23 34 83 89} and construct descriptions^{23 36 91 97} each used four times (8% each). Participant concept mapping^{23 36 88} and examination of administration methods^{36 98 99} were each used three times (6% each), and participant interviews^{88 100} were used twice (4%). Five other methods were each used once in five different studies: item intent descriptions,³⁶ items tested against item intent descriptions,¹⁰¹ item-response theory (IRT) analysis for item selection within domains,⁹⁰ item selection based on hospital

Table 4 Sources of evidence for all studies, total instances of validity evidence and for objective, subjective and multiple/mixed methods health literacy assessments

	Studies (n=46*)	Instances† (n=195)	Objective‡ (n=23 studies; n=102 instances)	Subjective§ (n=20 studies; n=78 instances)	Multiple and mixed methods (n=3 studies; n=15 instances)
	N (%)	N (%)	N (%)	N (%)	N (%)
Test content	22 (48)	52 (27)	27 (26)	22 (28)	3 (20)
Response processes	6 (13)	7 (4)	2 (2)	5 (6)	0 (0)
Internal structure	15 (33)	28 (14)	11 (11)	15 (19)	2 (13)
Relations to other variables	42 (91)	107 (55)	61 (60)	36 (46)	10 (67)
Validity and the consequences of testing	1 (2)	1 (1)	1 (1)	0 (0)	0 (0)

*Most studies reported more than one source of validity evidence.

†Each time validity evidence was reported within a study.

‡Measures an observer's (eg, clinician's) objective observations of a person's health literacy.

§Self-report (subjective) measure of health literacy.

medical texts⁸⁵ and item selection based on a health literacy conceptual model.¹⁰⁰

Evidence based on response processes

Only seven instances based on *response processes* were reported across 6 of the 46 studies (table 4 and online supplementary table 2). The methods used were cognitive interviews with respondents (n=3 instances; 43% of all evidence based on *response processes*)^{36 88 101} and with users (clinicians) (n=1; 14%),¹⁰¹ as well as recording and timing the response times of respondents (n=3; 43%).^{89 98 100}

Evidence based on internal structure

There were 15 studies (33% of all studies) that reported evidence based on the *internal structure* of health literacy assessments resulting in 28 instances (table 4 and online supplementary table 3). The most frequently reported methods were exploratory factor analysis (including principal component analysis) (n=7 instances; 25% of all evidence based on *response processes*)^{88 93 100 102–105} and confirmatory factor analysis (also n=7; 25%).^{91 106 107} Differential item functioning was reported three times (11%),^{88 91 102} and item-remainder correlations twice (7%).^{36 92} There were nine other methods used to generate evidence for *internal structure*, including a variety of specific IRT analyses for fit, item selection and internal consistency. Each method was reported once, with some authors reporting more than one method.^{36 86 89 90 103 106}

Evidence based on relations to other variables

This was the most commonly reported type of validity evidence across studies (n=42 studies; 91%) (table 4 and online supplementary table 4). There were 18 studies that only reported evidence based on *relations to other variables*.^{80 81 104 108–122} Evidence within this category was coded, as per the *Standards*, into convergent evidence (ie, relationships between items and scales of the same or similar structure), discriminant evidence (ie, assessments measuring different constructs determined to be sufficiently uncorrelated), criterion-referenced evidence (ie, how accurately scores predict criterion performance) and evidence for group differences (ie, relationships of scores with background characteristics such as demographic information). The *Standards* also includes evidence for generalisation but states that this relies primarily on studies that conduct research syntheses, and this review excluded studies that conducted meta-analyses. Across all studies, there were 107 instances of validity evidence reported for *relations to other variables*: 57 instances of convergent evidence (53% of all evidence in this category), 3 instances of discriminant evidence (3%), 17 instances of criterion-referenced evidence (16%) and 30 instances of evidence for group differences (28%).

The most frequently-used methods for convergent evidence were Spearman's^{80 85 94 96 99 105 108 110 116 118 122} and Pearson's^{33 34 82 83 90 93 104 112 113 120 123} correlation coefficients (11 instances and 19% each). These were closely followed by the receiver operating characteristic (ROC)

curve and the area under the ROC curve (also n=11 instances; 19%).^{81 97 99 103 110 111 117 120 123} A further eight instances (14%) of correlation calculations with similar measures were reported but the types of calculation they performed were unclear.^{86 87 92 95 103 115 119 121}

Harper, Elsworth *et al* and Osborne *et al*^{86 90 106} were the only three studies to generate discriminant evidence, as defined by the *Standards*. Harper⁹⁰ used the Pearson correlation coefficient to assess the association of components of a new health literacy instrument with the shortened version of the Test of Functional Health Literacy in Adults (S-TOFHLA). Elsworth *et al*¹⁰⁶ compared the average variance extracted and the variance shared between the nine scales of the Health Literacy Questionnaire (HLQ) (discriminant validity evidence between HLQ scales). Similarly, Osborne *et al*⁸⁶ conducted a multi-scale factor analysis to investigate if the nine HLQ scales were conceptually distinct.

Linear regression models were the most common method to generate criterion-referenced evidence (n=6 instances; 35% of all criterion-referenced evidence).^{86 90 107 114 115 121} The χ^2 test of independence was used by three studies (18%),^{87 115 121} with Spearman's correlation coefficient^{110 115} and logistic regression models^{86 115} each used by two studies (12% each).

There were 16 methods used to generate evidence for group differences and these were spread across 19 studies. The most frequently used methods were analysis of variance (n=5 instances; 17%)^{88 92 93 103 121} and linear regression models (n=4; 13%).^{80 83 91 123}

Evidence based on validity and consequences of testing

One study did investigations that led to conclusions about validity and the *consequences* of testing (p.221).⁸³ Elder *et al* found that the REALM under-represented the construct of health literacy when defined as the ability to obtain, interpret and understand basic health information.

Use of a validity testing framework when reporting validity evidence for health literacy assessments

Few studies referred to a validity testing framework or used a framework to structure or guide their work. Of the 46 studies, 9 directly or indirectly referenced a validity testing framework, and made a statement to support the citation (see online supplementary file 3). The frameworks directly cited by three studies^{87 101 106} were the 2014 *Standards*;⁵ Michael T Kane's argument-based approach to validation;¹⁴ Samuel J Messick's unified theory of validation;^{17 124} and Francis *et al*'s checklist operationalising measurement characteristics of patient-reported outcome measures.¹²⁵ There were six studies^{36 83 93 96 102 107} that indirectly cited Messick, Kane and/or the 1985, 1999 or 2014 versions of the *Standards*^{5 126 127} through other citations. A 10th study⁸⁸ referenced Buchbinder *et al*,¹²⁸ which cites the *Standards*, but there was no clear statement about validity testing to support the citation.

DISCUSSION

This systematic descriptive literature review found that studies in health literacy measurement rarely use or reference a structured theoretical framework for validation planning or testing. Further, this review's use of the *Standards*' framework revealed that validity testing studies for health literacy assessments most frequently, and often only, report evidence based on *relations to other variables*. It is usual and reasonable for a single validity study to not provide comprehensive evidence about a patient-reported outcome measure, and this is why an organising framework for evaluating evidence from a range of studies is so important. The findings from this review show that validation practice for health literacy assessments does not use established validity testing criteria and is yet to embrace the structural framework of contemporary validity testing theory.^{5 6}

In this review, evidence based on *relations to other variables* was the most frequent type of validity evidence reported across the 46 studies. It was reported more than twice as frequently as evidence based on *test content*, which was the second most commonly reported source of validity evidence. Evidence based on *internal structure* was reported in almost half the studies. This is not an unexpected result given the propensity for validity testing studies to almost routinely conduct correlation of an assessment with another variable (eg, a similar or different assessment).¹²⁹ In the early 20th Century, the focus of test validation was primarily on predictive validity practices (eg, prediction of student academic achievement) and so correlation with known criteria was a common validation practice.^{48 130 131} Development of the theory and practice of validation, and the need to use tests in various contexts with different population groups, has required consideration of the meaning of test scores, and that score interpretations usually lead to decisions or actions that can affect people's lives.^{2 3 52 66} As Kane explains, 'ultimately, the need for validation derives from the scientific and social requirement that public claims and decisions be justified' (p.17).¹³ A structured theoretical framework, such as the *Standards*, facilitates validation planning, testing and integration of evidence for decision-making. It can also support new users of a health assessment to judge existing evidence and previous rationales for data interpretation and use, and how these might justify the use of the assessment in a new context.

Reports of evidence based on *response processes* and on *consequences* of testing were negligible in this review. This is the first time this has been observed in the field of health literacy although it has been observed previously in other fields of research.^{50 68 132} Evidence based on the cognitive (response) processes of respondents (and of assessment users^{59 101}) can be essential to understanding the meanings derived from assessment scores for each new testing purpose.⁶⁹ Consequential evidence, although a controversial area of research,^{50 66} can reveal important outcomes for equitable decision-making, such as those discussed by Elder *et al*.⁸³ regarding the use

of the REALM, a word recognition assessment, with non-native speakers of English in a world in which health literacy is understood to be about equitable access to, and understanding and use of health information and services.^{42 133-135} Potential risks for unintended consequences of testing can be lessened through the development of the content of health assessments using comprehensive grounded practices that ensure wide and deep coverage of the lived experiences of intended respondents.^{36 136-138}

The findings of this review are important because institutions and governments around the world are increasingly implementing health literacy as a basis for health policy and practice development and evaluation.^{43-46 139} There needs to be certainty that inferences made from health literacy measurement data are leading to accurate and equitable decision-making about healthcare, interventions and policies, and that these decisions are as fair for the people with the lowest health literacy as for those with the highest.^{11 19 46 52 140-143} Some types of health interventions are known to widen health inequalities.¹⁴³⁻¹⁴⁷ Messick emphasises construct under-representation and construct-irrelevant variance as causes for negative testing consequences, as related to validity.^{124 148} For example, if a health assessment is biased by a specific perspective about causes of health disparities then construct under-representation can be a threat to the validity of inferences and actions taken from the scores. Likewise, if an assessment reflects a particular social perspective (eg, middle class values and language embedded in the items) then there is the threat that the responses to the assessment are perfused with irrelevant variance derived from that perspective. Evidence from a range of sources is required to justify the use of measurement data in specific contexts (eg, socioeconomic, demographic, cultural, language), and to assure decision-makers of the absence of validity threats.^{4 51 54}

This is the first time that a comprehensive review of sources of validity evidence for health literacy assessments has been undertaken within the theoretical validity testing framework of the *Standards*. For some methods, coding into the five sources of validity evidence was not straightforward and, in these cases, the *Standards* were consulted closely for guidance. Coding of studies by Elsworth *et al* and Osborne *et al*.^{36 106} to *relations to other variables* (discriminant evidence) required some deliberation because the evidence in both studies was for discrimination analyses between independent scales *within* a multiscale health literacy assessment, rather than between different health literacy assessments. The developers of the HLQ view the nine scales as measuring distinct, although related, constructs.³⁶ The *Standards* (p.16) explain that 'external variables may include measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesised to measure the same constructs, and tests measuring related or different constructs'.⁵ It was on the basis of the last part of this statement about tests measuring related or different constructs that these two

studies were coded in *relations to other variables* as discriminant evidence.

In a few studies, some assessments seemed to be regarded as proxies for health literacy, which suggested that the researchers were thinking of them as measuring similar constructs to health literacy. In these cases, evidence was coded in *relations to other variables* as convergent evidence (ie, convergence between measures of the same or similar construct) rather than as criterion-referenced evidence (ie, prediction of other criteria). For example, Curtis *et al*⁸⁶ explored correlations between the Comprehensive Health Activities Scale with the Mini Mental Status Exam as well as with the TOFHLA, the REALM and the NVS.⁸⁶ Driessnack *et al*¹⁰⁸ looked at correlations between parents' and children's NVS scores with their self-reports of the number of children's books in the home. Dykhuis *et al*⁸⁷ correlated the Brief Medical Numbers Test with the Montreal Cognitive Assessment as well as with two versions of the REALM.

Further to coding for *relations to other variables* are the distinctions between convergent evidence, criterion-referenced evidence and evidence for group differences. Coding to convergent evidence was based on analyses of assessments of the same or similar construct (eg, typically, comparisons of one health literacy assessment with another health literacy assessment). Coding to criterion-referenced evidence was based on analyses of prediction (eg, a health literacy assessment with a disease knowledge survey). Coding for evidence of group differences was based on analyses of relationships with background characteristics such as demographic information.

Reliability was not coded within the five sources of evidence even though it does contribute to understanding the validity of score interpretations and use, especially for purposes of generalisation.⁵ The *Standards* (p.33) classifies reliability into reliability/precision (ie, consistency of scores across different instances of testing) and reliability/generalisability coefficients (ie, in the way that classical test theory refers to reliability as being correlation between scores on two equivalent forms of a test, with the assumption that there is no effect of the first test instance on the second test instance). The predominant focus in the reviewed papers was on the latter conception of reliability, most often calculated using Cronbach's alpha.

Strengths and limitations

An element of bias is potentially present in this review because of the restriction of the search to studies published and health literacy assessments developed and administered in the English language. Future studies may be improved if other languages were included. The health literacy assessments reviewed are those that are predominant in the field and may well provide a foundation for validity studies of more specifically targeted assessments.

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. Validation practice is complex and there are many groups publishing validity testing studies that may have limited training and experience in the area.¹⁻⁴ There was a lack of clarity in some papers and theses about the methods used and results obtained, which caused difficulties with classifying the evidence within the *Standards* framework, so some misclassification is possible for some papers. Future work in this area would be improved if researchers used clearly defined and structured validity testing frameworks (ie, the five validity evidence sources of the *Standards*) in which to classify evidence.

The main strength of this study was that validity is clearly defined as the extent to which theory and evidence (quantitative and qualitative) support score interpretation and use. This definition is in accordance with leading authorities in the validity testing literature.^{2 5 13 51} A second strength of this study was the use of an established and well-researched theoretical validity testing framework, the *Standards*, to examine sources of evidence for health literacy assessments. Different health literacy assessments have different measurement purposes. Validation planning with a structured framework would help to determine the sources of evidence needed to justify the inferences from data, and to guide potential users. Application of theory to validation practice will provide a scientific basis for the development and testing of health assessments, enable systematic evaluations of validity evidence and help detect possible threats to the validity of the interpretation and use of data in different contexts.^{2 3 15}

CONCLUSIONS

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.

Twitter Melanie Hawkins @4MelanieHawkins and Richard H Osborne @richardosborne4

Acknowledgements The authors acknowledge and thank Rachel West, Deakin University Liaison Librarian, for her expertise in systematic literature reviews and her patient guidance through the detailed process of searching the literature.

Contributors MH and RHO conceptualised the research question and analytical plan. MH led, with all authors contributing to, the development of the search strategy, selection criteria, data extraction criteria and analysis method. MH conducted the literature search with guidance from EH. MH screened the literature, and extracted and analysed the data with the continuous support of and comprehensive checking by GRE. MH drafted the initial manuscript and led subsequent drafts. GRE, RHO and EH read and provided feedback on manuscript iterations, and approved the final manuscript. RHO is the guarantor.

Funding MH was funded by a National Health and Medical Research Council (NHMRC) of Australia Postgraduate Scholarship (APP1150679). RHO was funded in part through a National Health and Medical Research Council (NHMRC) of Australia Principal Research Fellowship (APP1155125).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Melanie Hawkins <http://orcid.org/0000-0001-5704-0490>

REFERENCES

- 1 McClimans L. A theoretical framework for patient-reported outcome measures. *Theor Med Bioeth* 2010;31:225–40.
- 2 Social Indicators Research Series, Zumbo BD, Chan EK, eds. *Validity and validation in social, behavioral, and health sciences*. Switzerland: Springer International Publishing, 2014.
- 3 Sawatzky R, Chan EKH, Zumbo BD, et al. Montreal Accord on patient-reported outcomes (pros) use series-Paper 7: modern perspectives of measurement validation emphasize justification of inferences based on patient reported outcome scores. *J Clin Epidemiol* 2017;89:154–9.
- 4 Kwon JY, Thorne S, Sawatzky R. Interpretation and use of patient-reported outcome measures through a philosophical lens. *Qual Life Res* 2019;28:629–36.
- 5 AERA, APA, NCME. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 2014.
- 6 Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. *Qual Life Res* 2018;27:1695–710.
- 7 O'Leary TM, Hattie JAC, Griffin P. Actual interpretations and use of scores as aspects of validity. *EM: IP* 2017;36:16–23.
- 8 Bakker M et al. Acting together—WHO National health literacy demonstration projects (NHLDPs) address health literacy needs in the European region. *Public Health Panorama* 2019;5:233–43.
- 9 Klinker CD, Aaby A, Ringgaard LW, et al. Health literacy is associated with health behaviors in students from vocational education and training schools: a Danish population-based survey. *Int J Environ Res Public Health* 2020;17:671.
- 10 Chapelle CA. *The TOEFL validity argument. Building a validity argument for the Test of English as a Foreign Language*, 2008: 319–52.
- 11 Elsworth GR, Nolte S, Osborne RH. Factor structure and measurement invariance of the health education impact questionnaire: does the subjectivity of the response perspective threaten the contextual validity of inferences? *SAGE Open Med* 2015;3:2050312115585041
- 12 Shepard LA. The centrality of test use and consequences for test validity. *Em:ip* 1997;16:5–24.
- 13 Kane MT. *Validation*. In: Brennan R, ed. *Educational Measurement*. Rowman & Littlefield Publishers / Amer Council Ac1 (Pre Acq), 2006: 17–64.
- 14 Kane MT. An argument-based approach to validity. *Psychol Bull* 1992;112:527–35.
- 15 Kane MT. Explicating validity. *Assess Educ Princ Pol Pract* 2016;23:198–211.
- 16 Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 1995;50:741–9.
- 17 Messick S. Validity of test interpretation and use. *ETS Research Report Series* 1990;1990:1487–95.
- 18 Moss PA, Girard BJ, Haniford LC. *Validity in educational assessment. Review of research in education*, 2006: 109–62.
- 19 Batterham RW, Hawkins M, Collins PA, et al. Health literacy: applying current concepts to improve health services and reduce health inequalities. *Public Health* 2016;132:3–12.
- 20 Shepard LA. Evaluating test validity: Reprise and progress. *Assess Educ Princ Pol Pract* 2016;23:268–80.
- 21 Hubley AM, Zumbo BD. A dialectic on validity: where we have been and where we are going. *J Gen Psychol* 1996;123:207–15.
- 22 Nutbeam D. The evolving concept of health literacy. *Soc Sci Med* 2008;67:2072–8.
- 23 Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80.
- 24 Sykes S, Wills J, Rowlands G, et al. Understanding critical health literacy: a concept analysis. *BMC Public Health* 2013;13:150.
- 25 Pleasant A, McKinney J, Rikard RV. Health literacy measurement: a proposed research agenda. *J Health Commun* 2011;16 Suppl 3:11–21.
- 26 Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol* 2011;64:366–79.
- 27 Altin SV, Finke I, Kautz-Freimuth S, et al. The evolution of health literacy assessment tools: a systematic review. *BMC Public Health* 2014;14:1207.
- 28 McCormack L, Haun J, Sørensen K, et al. Recommendations for advancing health literacy measurement. *J Health Commun* 2013;18 Suppl 1:9–14.
- 29 Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. *Nurs Health Sci* 2009;11:77–89.
- 30 Haun JN, Valerio MA, McCormack LA, et al. Health literacy measurement: an inventory and descriptive summary of 51 instruments. *J Health Commun* 2014;19 Suppl 2:302–33.
- 31 Guzys D, Kenny A, Dickson-Swift V, et al. A critical review of population health literacy assessment. *BMC Public Health* 2015;15:215.
- 32 Barry AE, Chaney B, Piazza-Gardner AK, et al. Validity and reliability reporting practices in the field of health education and behavior: a review of seven journals. *Health Educ Behav* 2014;41:12–18.
- 33 Davis TC, Crouch MA, Long SW, et al. Rapid assessment of literacy levels of adult primary care patients. *Fam Med* 1991;23:433–5.
- 34 Weiss BD, Mays MZ, Martz W, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med* 2005;3:514–22.
- 35 Jessup RL, Osborne RH, Buchbinder R, et al. Using co-design to develop interventions to address health literacy needs in a hospitalised population. *BMC Health Serv Res* 2018;18:989.
- 36 Osborne RH, Batterham RW, Elsworth GR, et al. The grounded psychometric development and initial validation of the health literacy questionnaire (HLQ). *BMC Public Health* 2013;13:658.
- 37 Jessup RL, Buchbinder R. What if I cannot choose wisely? Addressing suboptimal health literacy in our patients to reduce over-diagnosis and overtreatment. *Intern Med J* 2018;48:1154–7.
- 38 Roberts J. *Local action on health inequalities: Improving health literacy to reduce health inequalities*. London: UCL Institute of Health Equity, 2015.
- 39 Batterham RW, Buchbinder R, Beauchamp A, et al. The optimising health literacy (Ophelia) process: study protocol for using health

- literacy profiling and community engagement to create and implement health reform. *BMC Public Health* 2014;14:694–703.
- 40 Beauchamp A, Batterham RW, Dodson S, et al. Systematic development and implementation of interventions to optimise health literacy and access (Ophelia). *BMC Public Health* 2017;17:230.
 - 41 Barry MM, D'Eath M, Sixsmith J. Interventions for improving population health literacy: insights from a rapid review of the evidence. *J Health Commun* 2013;18:1507–22.
 - 42 Australian Bureau of Statistics. National Health Survey: Health Literacy, 2018, 2019. Available: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4364.0.55.014Main+Features12018?OpenDocument> [Accessed 14 Oct 2019].
 - 43 Trezona A, Rowlands G, Nutbeam D. Progress in implementing national policies and strategies for health Literacy-What have we learned so far? *Int J Environ Res Public Health* 2018;15:1554.
 - 44 WHO Regional Office for Europe, WHO Health Evidence Network synthesis report. *What is the evidence on the methods, frameworks and indicators used to evaluate health literacy policies, programmes and interventions at the regional, National and organizational levels?* Copenhagen, 2019.
 - 45 Putoni S. *Health Literacy in Wales - A Scoping Document for Wales*. Wales: Welsh Assembly Government, 2010.
 - 46 Scottish Government NHS Scotland. *Making it Easier: A Health Literacy Action Plan for Scotland 2017–2025*. Edinburgh, 2017.
 - 47 Kelley TL. *Interpretation of educational measurements*. In: *Measurement and adjustment series*. Yonkers-on-Hudson, NY: World Book, 1927: 1–363.
 - 48 Sireci SG. On the validity of useless tests. *Assess Educ Princ Pol Pract* 2016;23:226–35.
 - 49 Hawkins M, Elsworth GR, Osborne RH. Questionnaire validation practice: a protocol for a systematic descriptive literature review of health literacy assessments. *BMJ Open* 2019;9:e030753.
 - 50 Cizek GJ, Rosenberg SL, Koons HH. Sources of validity evidence for educational and psychological tests. *Educ Psychol Meas* 2008;68:397–412.
 - 51 Messick S, Linn R. *Validity*. In: *Educational Measurement*. New York: American Council on Education/Macmillan publishing company, 1989.
 - 52 Messick S. *Consequences of test interpretation and use: The fusion of validity and values in psychological assessment*. In: *Ets research report series*. , 1998: 1998, i–32.
 - 53 Cronbach LJ, Wainer H. *Five perspectives on validity argument*. In: *Test validity*. New Jersey: Lawrence Erlbaum Associates Inc, 1988: 3–17.
 - 54 House E. *Evaluating with validity*. Beverly Hills, California: Sage Publications, 1980.
 - 55 Shepard LA, Darling-Hammond L. *Validity Etest*. In: *Review of research in education*. American Educational Research Association, 1993: 405–50.
 - 56 Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas* 2013;50:1–73.
 - 57 Kane MT. Validity as the evaluation of the claims based on test scores. *Assess Educ Princ Pol Pract* 2016;23:309–11.
 - 58 Cox DW, Owen JJ, Zumbo BD, Chan EK, eds. *Validity evidence for a perceived social support measure in a population health context, in Validity and validation in social, behavioral, and health sciences*. Switzerland: Springer International Publishing, 2014.
 - 59 Zumbo BD, Hubley AM. *Understanding and investigating response processes in validation research*. In: Michalos AC, ed. *Social indicators research series*. Vol. 69. Switzerland: Springer International Publishing, 2017.
 - 60 Hubley AM, Zumbo BD. *Response processes in the context of validity: Setting the stage*. In: *Understanding and investigating response processes in validation research*. Switzerland: Springer International Publishing, 2017: 1–12.
 - 61 Onwuegbuzie AJ, Leech NL. *Validity and qualitative research: An oxymoron?* Quality and Quantity 2007;41:233–49.
 - 62 Onwuegbuzie AJ, Leech NL. On becoming a pragmatic researcher: the importance of combining quantitative and qualitative research methodologies. *Int J Soc Res Methodol* 2005;8:375–87.
 - 63 Castillo-Díaz M, Padilla J-L. How cognitive interviewing can provide validity evidence of the response processes to scale items. *Soc Indic Res* 2013;114:963–75.
 - 64 Padilla J-L, Benítez I. Validity evidence based on response processes. *Psicothema* 2014;26:136–44.
 - 65 Padilla J-L, Benítez I, Castillo M. Obtaining validity evidence by cognitive interviewing to interpret psychometric results. *Methodology* 2013;9:113–22.
 - 66 Moss PA. The role of consequences in validity theory. *Em: Ip* 1998;17:6–12.
 - 67 Hubley AM, Zumbo BD. Validity and the consequences of test interpretation and use. *Soc Indic Res* 2011;103:219–30.
 - 68 Zumbo BD, Hubley AM. Bringing consequences and side effects of testing and assessment to the foreground. *Assess Educ Princ Pol Pract* 2016;23:299–303.
 - 69 Kane M, Mislevy R. *Validating score interpretations based on response processes*. In: Ercikan K, Pellegrino JW, eds. *Validation of score meaning for the next generation of assessments*. New York: Routledge, 2017: 11–24.
 - 70 Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
 - 71 DeVellis RF. A consumer's guide to finding, evaluating, and reporting on measurement instruments. *Arthritis Care Res* 1996;9:239–45.
 - 72 Aaronson N, Alonso J, Burnam A, Lohr KN, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–205.
 - 73 King WR, He J. Understanding the role and methods of meta-analysis in is research. *CAIS* 2005;16.
 - 74 Yang H, Tate M. A descriptive literature review and classification of cloud computing research. *CAIS* 2012;31.
 - 75 Schlagenhauer C, Amberg M. *A descriptive literature review and classification framework for Gamification in information systems*. Germany: Gartner: European Conference on Information Systems, 2015.
 - 76 Roter DL, Hall JA, Katz NR. Patient-Physician communication: a descriptive summary of the literature. *Patient Educ Couns* 1988;12:99–119.
 - 77 Guzzo RA, Jackson SE, Katzell RA, et al. *Meta-analysis analysis*. In: Cummings LL, ed. *Research in Organizational Behavior*. Greenwich: JAI Press, 1987: 407–42.
 - 78 Paré G, Trudel M-C, Jaana M, et al. Synthesizing information systems knowledge: a typology of literature reviews. *Inf Manage* 2015;52:183–99.
 - 79 Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
 - 80 Barber MN, Staples M, Osborne RH, et al. Up to a quarter of the Australian population may have suboptimal health literacy depending upon the measurement tool: results from a population-based survey. *Health Promot Int* 2009;24:252–61.
 - 81 Wallace LS, Rogers ES, Roskos SE, et al. Brief report: screening items to identify patients with limited health literacy skills. *J Gen Intern Med* 2006;21:874–7.
 - 82 Davis TC, Long SW, Jackson RH, et al. *Rapid estimate of adult literacy in medicine: a shortened screening instrument*. *Fam Med* 1993;25:391–5.
 - 83 Elder C, Barber M, Staples M, et al. Assessing health literacy: a new domain for collaboration between language testers and health professionals. *Lang Assess Q* 2012;9:205–24.
 - 84 Dumenci L, Matsuyama RK, Kuhn L, et al. On the validity of the rapid estimate of adult literacy in medicine (realm) scale as a measure of health literacy. *Commun Methods Meas* 2013;7:134–43.
 - 85 Parker RM, Baker DW, Williams MV, et al. The test of functional health literacy in adults. *J Gen Intern Med* 1995;10:537–41.
 - 86 Curtis LM, Revelle W, Waite K, et al. Development and validation of the comprehensive health activities scale: a new approach to health literacy measurement. *J Health Commun* 2015;20:157–64.
 - 87 Dykhuis KE, Slowik L, Bryce K, et al. A new measure of health Numeracy: brief medical numbers test (BMNT). *Psychosomatics* 2019;60:271–7.
 - 88 Jordan JE, Buchbinder R, Briggs AM, et al. The health literacy management scale (HeLMS): a measure of an individual's capacity to seek, understand and use health information within the healthcare setting. *Patient Educ Couns* 2013;91:228–35.
 - 89 Zhang X-H, Thumboo J, Fong K-Y, et al. Development and validation of a functional health literacy test. *Patient* 2009;2:169–78.
 - 90 Harper R. *Comprehensive health literacy assessment for college students in Department of Journalism and Technical Communication*. Fort Collins, Colorado: Colorado State University, 2013.
 - 91 Bann CM, McCormack LA, Berkman ND, et al. The health literacy skills instrument: a 10-item short form. *J Health Commun* 2012;17 Suppl 3:191–202.
 - 92 McCormack L, Bann C, Squiers L, et al. Measuring health literacy: a pilot study of a new skills-based instrument. *J Health Commun* 2010;15 Suppl 2:51–71.
 - 93 DeBello MC. *The development and psychometric testing of the health literacy knowledge, application, and confidence scale*

- (HLKACS), in *College of Education and College of Nursing*. Michigan: Eastern Michigan University, 2016.
- 94 Baker DW, Williams MV, Parker RM, *et al*. Development of a brief test to measure functional health literacy. *Patient Educ Couns* 1999;38:33–42.
 - 95 Shaw TC. Uncovering health literacy: developing a remotely administered questionnaire for determining health literacy levels in health disparate populations. *J Hosp Adm* 2014;3:149.
 - 96 Begoray DL, Kwan B. A Canadian exploratory study to define a measure of health literacy. *Health Promot Int* 2012;27:23–32.
 - 97 Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:12.
 - 98 Chesser AK, Keene Woods N, Wiperman J, *et al*. Health literacy assessment of the STOPHLA: paper versus electronic administration continuation study. *Health Educ Behav* 2014;41:19–24.
 - 99 Dageforde LA, Cavanaugh KL, Moore DE, *et al*. Validation of the written administration of the short literacy survey. *J Health Commun* 2015;20:835–42.
 - 100 Sørensen K, Van den Broecke S, Pelikan JM, *et al*. Measuring health literacy in populations: illuminating the design and development process of the European health literacy survey questionnaire (HLS-EU-Q). *BMC Public Health* 2013;13:1–22.
 - 101 Hawkins M, Gill SD, Batterham R, *et al*. The health literacy questionnaire (HLQ) at the patient-clinician interface: a qualitative study of what patients and clinicians mean by their HLQ scores. *BMC Health Serv Res* 2017;17:309.
 - 102 Morris RL, Soh S-E, Hill KD, *et al*. Measurement properties of the health literacy questionnaire (HLQ) among older adults who present to the emergency department after a fall: a Rasch analysis. *BMC Health Serv Res* 2017;17:605.
 - 103 Sand-Jecklin K, Coyle S. Efficiently assessing patient health literacy: the BHLS instrument. *Clin Nurs Res* 2014;23:581–600.
 - 104 Haun J, Luther S, Dodd V, *et al*. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. *J Health Commun* 2012;17 Suppl 3:141–59.
 - 105 Miller B. *Investigating the Construct of Health Literacy Assessment: A Cross-Validation Approach, in Graduate School, the College of Education and Psychology and the Department of Educational Research & Administration*. Ann Arbor, MI: The University of Southern Mississippi, 2018.
 - 106 Elsworth GR, Beauchamp A, Osborne RH. Measuring health literacy in community agencies: a Bayesian study of the factor structure and measurement invariance of the health literacy questionnaire (HLQ). *BMC Health Serv Res* 2016;16:508.
 - 107 Goodwin BC, March S, Zajdlewicz L, *et al*. Health literacy and the health status of men with prostate cancer. *Psychooncology* 2018;27:2374–81.
 - 108 Driessnack M, Chung S, Perkhounkova E, *et al*. Using the "Newest Vital Sign" to assess health literacy in children. *J Pediatr Health Care* 2014;28:165–71.
 - 109 Goodman MS, Griffey RT, Carpenter CR, *et al*. Do subjective measures improve the ability to identify limited health literacy in a clinical setting? *J Am Board Fam Med* 2015;28:584–94.
 - 110 Cavanaugh KL, Osborn CY, Tentori F, *et al*. Performance of a brief survey to assess health literacy in patients receiving hemodialysis. *Clin Kidney J* 2015;8:462–8.
 - 111 Chew LD, Griffin JM, Partin MR, *et al*. Validation of screening questions for limited health literacy in a large Va outpatient population. *J Gen Intern Med* 2008;23:561–6.
 - 112 Houston AJ, Lowenstein LM, Hoover DS, *et al*. Limitations of the S-TOFHLA in measuring poor numeracy: a cross-sectional study. *BMC Public Health* 2018;18:405.
 - 113 Kirk JK, Grzywacz JG, Arcury TA, *et al*. Performance of health literacy tests among older adults with diabetes. *J Gen Intern Med* 2012;27:534–40.
 - 114 Ko Y, Lee JY-C, Toh MPH, *et al*. Development and validation of a general health literacy test in Singapore. *Health Promot Int* 2012;27:45–51.
 - 115 Kordovski VM, Woods SP, Avci G, *et al*. Is the newest vital sign a useful measure of health literacy in HIV disease? *Journal of the International Association of Providers of AIDS Care* 2017;16): :595–602.
 - 116 McNaughton C, Wallston KA, Rothman RL, *et al*. Short, subjective measures of numeracy and general health literacy in an adult emergency department. *Acad Emerg Med* 2011;18:1148–55.
 - 117 Morris NS, MacLean CD, Chew LD, *et al*. The single item literacy screener: evaluation of a brief instrument to identify limited reading ability. *BMC Fam Pract* 2006;7:21.
 - 118 Quinzanos I, Hirsh JM, Bright C, *et al*. Cross-Sectional correlation of single-item health literacy screening questions with established measures of health literacy in patients with rheumatoid arthritis. *Rheumatol Int* 2015;35:1497–502.
 - 119 Rawson KA, Gunstad J, Hughes J, *et al*. The meter: a brief, self-administered measure of health literacy. *J Gen Intern Med* 2010;25:67–71.
 - 120 Wallston KA, Cawthon C, McNaughton CD, *et al*. Psychometric properties of the brief health literacy screen in clinical practice. *J Gen Intern Med* 2014;29:119–26.
 - 121 Hadden KB, Literacy H. *Pregnancy: Validation of a New Measure and Relationships of Health Literacy to Pregnancy Risk Factors*. Arkansas: University of Arkansas for Medical Sciences, 2012.
 - 122 Soelberg J. *Determining the reliability and validity of the newest vital sign in the inpatient setting*. Chicago, Illinois: Rush University, 2015.
 - 123 Haun JN. *Health Literacy: The Validation of a Short Form Health Literacy Screening Assessment in an Ambulatory Care Setting*. Florida: University of Florida, 2007.
 - 124 Messick S. *Foundations of validity: Meaning and consequences in psychological assessment*. In: *Ets research report series*. , 1993: 1993, i–18.
 - 125 Francis DO, McPheeters ML, Noud M, *et al*. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Syst Rev* 2016;5:129.
 - 126 American Educational Research Association. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 1999.
 - 127 American Psychological Association, National Council on Measurement in Education. *Standards for educational and psychological testing*. American Educational Research Association, 1985.
 - 128 Buchbinder R, Batterham R, Elsworth G, *et al*. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. *Arthritis Res Ther* 2011;13:ar3468.
 - 129 McClimans L. Interpretability, validity, and the minimum important difference. *Theor Med Bioeth* 2011;32:389–401.
 - 130 Kane M, Bridgeman B. *Research on Validity Theory and Practice at ETS, in Advancing Human Assessment*. In: Bennett RE, von Davier M, eds. *The Methodological, Psychological and Policy Contribution of ETS*. Switzerland: Springer Nature: Cham, 2017: 489–552.
 - 131 Landy FJ. Stamp collecting versus science: validation as hypothesis testing. *Am Psychol* 1986;41:1183–92.
 - 132 Spurgeon SL. Evaluating the unintended consequences of assessment practices: construct irrelevance and construct underrepresentation. *MECD* 2017;50:275–81.
 - 133 Nutbeam D. Health promotion glossary. *Health Promot Int* 1998;13:349–64.
 - 134 New Zealand Ministry of Health. *Content Guide 2017/18: New Zealand Health Survey*. Wellington, New Zealand: NZ Ministry of Health, 2018.
 - 135 Bo A, Friis K, Osborne RH, *et al*. National indicators of health literacy: ability to understand health information and to engage actively with healthcare providers - a population-based survey among Danish adults. *BMC Public Health* 2014;14:1095.
 - 136 Busija L, Buchbinder R, Osborne RH. A grounded patient-centered approach generated the personal and societal burden of osteoarthritis model. *J Clin Epidemiol* 2013;66:994–1005.
 - 137 Rosas SR, Ridings JW. The use of concept mapping in measurement development and evaluation: application and future directions. *Eval Program Plann* 2017;60:265–76.
 - 138 Soellner R, Lenartz N, Rudinger G. Concept mapping as an approach for expert-guided model building: the example of health literacy. *Eval Program Plann* 2017;60:245–53.
 - 139 WHO Regional Office for Europe. Health literacy in action, 2019. Available: <http://www.euro.who.int/en/health-topics/disease-prevention/health-literacy/health-literacy-in-action> [Accessed 18 Oct 2019].
 - 140 Nguyen TH, Park H, Han H-R, *et al*. State of the science of health literacy measures: validity implications for minority populations. *Patient Educ Couns* 2015;j.pec.2015.07.013.
 - 141 Marmot M. *Fair society, healthy lives: the Marmot Review: strategic review of health in equalities in England Post-2010*, 2010.
 - 142 Kane M. Validity and Fairness. *Language Testing* 2010;27:177–82.
 - 143 Carey G, Crammond B, De Leeuw E. Towards health equity: a framework for the application of proportionate universalism. *Int J Equity Health* 2015;14:81.



- 144 Beauchamp A, Backholer K, Magliano D, *et al.* The effect of obesity prevention interventions according to socioeconomic position: a systematic review. *Obes Rev* 2014;15:541–54.
- 145 Beeston C *et al.* *Health inequalities policy review for the Scottish Ministerial Task force on health inequalities*. Edinburgh, 2014.
- 146 Addison M *et al.* Equal North: how can we reduce health inequalities in the North of England? A prioritization exercise with researchers, policymakers and practitioners. *Int J Public Health* 2018:1–13.
- 147 Capewell S, Graham H. Will cardiovascular disease prevention widen health inequalities? *PLoS Med* 2010;7:e1000320.
- 148 Messick S. Test validity: a matter of consequence. *Soc Indic Res* 1998;45:35–44.