HHS Public Access

Author manuscript

Patient Educ Couns. Author manuscript; available in PMC 2017 January 29.

State of the science of health literacy measures: Validity implications for minority populations

Tam H. Nguyen[a,*], **Hyunjeong Park**[b], **Hae-Ra Han**[c], **Kitty S. Chan**[d], **Michael K. Paasche-Orlow**[e], **Jolie Haun**[f,g], and **Miyong T. Kim**[h]

[a]School of Nursing, Boston College, Chestnut Hill, USA

[b]Department of Nursing, Towson University, Towson, USA

[c]School of Nursing, Johns Hopkins University, Baltimore, USA

[d]School of Public Health, Johns Hopkins University, Baltimore, USA

[e]Boston University School of Medicine, Boston Medical Center, Boston, USA

[f]Center of Innovation on Disability and Rehabilitation Research, James A. Haley VA Hospital, Tampa, USA

[g]Department of Community and Family Health, College of Public Health, University of South Florida, Tampa, USA

[h]School of Nursing, University of Texas at Austin, Austin, USA

Abstract

Objectives—To review the evidence supporting the validity of health literacy (HL) measures for ethnic minority populations.

Methods—PubMed, CINAHL, and PsycINFO databases were searched for HL measures between 1965 and 2013.

Results—A total of 109HL measures were identified; 37 were non-English HL measures and 72 were English language measures. Of the 72 English language measures, 17 did not specify the racial/ethnic characteristic of their sample. Of the remaining 55 measures, 10 (18%) did not include blacks, 30 (55%) did not include Hispanics, and 35 (64%) did not include Asians in their validation sample. When Hispanic and Asian Americans were included, they accounted for small percentages in the overall sample. Between 2005–2013, a growing number of REALM and TOFHLA translations were identified, and new HL measures for specific cultural/linguistic groups within and outside the United States were developed.

Conclusions—While there are a growing number of new and translated HL measures for minority populations, many existing HL measures have not been properly validated for minority groups.

[*]Corresponding author at: Boston College, 140 Commonwealth Ave, Chestnut Hill, MA 02467, USA. Fax: +1 617 552 8853., tam.nguyen@bc.edu (T.H. Nguyen).

Conflicts of interest: none.

Practice Implications—HL measures that have not been properly validated for a given population should be piloted before wider use. In addition, improving HL instrument development/validation methods are imperative to increase the validity of these measures for minority populations.

Keywords

Health Literacy; Psychometric; Review; Minority Health

1. Introduction

Health literacy (HL) is most widely defined as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” [1]. HL is necessary in almost all facets of healthcare, and poses a challenge to millions of individuals worldwide [2,3]. However, limited HL is endemic and especially problematic among older adults, ethnic minorities, and those who are less educated [2,4–6].

Ethnic minorities in the United States (US) and people with low English proficiency are particularly vulnerable. For example, in the US 62%, 48%, and 27% of adult immigrants born in Latin America, Asia, and Africa, respectively, are unable to speak English well [7]. In addition, many have low educational attainment, with 56%, 21%, and 14% of immigrants born in Latin America, Asia, and Africa, respectively, having less than a high school education [7]. Furthermore, up to 46–59% of elderly immigrant populations, such as the Hmong, Korean, Vietnamese, and Chinese Americans, are categorized as linguistically isolated [8,9], and have reported high levels of stress as a result of their inability to communicate with health-care providers [10,11]. Similar trends have been reported globally. Notably, in Europe, where there are over 75 million immigrants, minorities also score low on health literacy measures, have poor access to health information, and receive less healthcare services [3].

The high prevalence of limited HL and the associated consequences have been extensively documented [2,12]. Supporting the expansion of this literature, there has been a significant proliferation of tools to measure HL. Several reviews of these measures have been published [13–15]. However, existing reviews do not focus on validation issues affecting appropriate use with ethnic minority populations, and have been more narrowly focused in various ways. For example, Luk and Aslani [14] focused exclusively on written health information from the “document and user perspective,” thereby precluding measures that focused on other domains of HL. Mancuso et al. [13] did not limit their work to a specific domain of HL, but only included measures from 1991 to 2006; the number and linguistic diversity of HL measures has substantially increased in the intervening years. The absence of a critical evaluation for the validity of HL tools for use among ethnic minorities is particularly important given the potential introduction of measurement bias when researchers use instruments that have not been well validated for a particular population. Specifically, measures that have been developed using methods associated with Classical Test Theory (CTT) result in reliability estimates that are highly dependent on the sample, and require

extensive validation before use in other populations [16,17]. Whereas measures developed using methods associated with Item Response Theory (IRT) or Rasch Modeling, have the potential to produce items and scales that are metrically equivalent across different groups [18,19]. To address this gap, we have systematically reviewed the validation data that have been presented for racial and ethnic minority populations for all HL measures.

2. Methods

2.1. Search strategy

To identify HL measures, three electronic databases (PubMed, CINAHL, and PsycINFO) were searched for studies published between January 1966 and September 2013. In PubMed, MeSH terms and synonyms related to “health literacy” AND “measures” were searched, Appendix A. The search strategies used in CINAHL and PsycINFO (Appendix B and C) were built to reflect similar MeSH terms used in the PubMed search, but were more specific to the respective databases. After eliminating duplicate studies, the titles and abstracts of the remaining studies were reviewed independently by two authors (T. Nguyen and H. Park), Fig. 1. A structured form was used to summarize study eligibility, record reasons for rejection, and assess the characteristics of each HL instrument, including reliability, validity, target population, number of items, and HL domain(s) measured (i.e., reading, writing, numeracy, listening, speaking, and cultural/conceptual knowledge, with reading and writing encompassing the domain of print literacy, and listening and speaking encompassing the domain of oral literacy). Rather than summarizing the domain (s) measures as a specific HL related skill, as characterized by Sorensen et al. [20], we simply documented the name of the domain(s) measured as it was reported in the primary studies. A kappa coefficient (K) was computed to assess the inter-rater reliability between the two reviewers ($K = 0.89$). Disagreements were resolved by seeking feedback from the senior author (M. Kim). The first author (T. Nguyen) performed a detailed assessment of the included studies and identified additional studies via cross-referencing.

2.2. Study inclusion, exclusion, and evaluation criteria

Articles considered in this review were papers that presented original data on the validation of a new, translated, or modified HL instrument. Articles that described measures not explicitly related to HL, such as financial literacy, were excluded. In addition, articles that did not include any psychometric evaluation were excluded. The HL measures found in this review were evaluated based on: (i) generalizability of the tool to diverse populations, and (ii) psychometric properties of the measure. They were then organized according to whether the measure was general or content/context-specific. To evaluate the generalizability of the measures to diverse populations, the study sample was assessed and documented. The reported psychometrics of each measure was also documented to provide assessments of its reliability and validity.

3. Results

The electronic search identified 7596 publications, with 5076 remaining after duplicates ($n = 2520$) were removed. Two authors independently reviewed the titles and abstracts of all

5076 publications and came to a consensus to review 84 full-length publications. Of the 84 publications reviewed, 10 were rejected on the basis of the aforementioned exclusion criteria. Hand searches of references yielded 21 additional publications. Therefore, a total of 95 relevant publications were identified in this literature review. From the 95 publications, 109 unique HL measures were found since some publications reporting on multiple measures (i.e., short and long forms of a measure, or measures that were translated and validated across several different languages), Fig. 1.

Tables 1–3 summarizes the general HL measures [4,5,21–67], and Tables 4–6 summarizes content/context-specific measures [68–110]. The tables are further organized according to whether the measures are similar to the Rapid Estimate of Adult Literacy in Medicine (REALM), referred to as "REALM-*like* "; similar to the Test of Functional Health Literacy in Medicine (TOFHLA), referred to as "TOFHLA-*like* "; "or "Unique," HL measures. Furthermore, unless measures are directly related to each other (i.e., a shortened or revised version), the measures are listed in chronological order of publication, from oldest to newest.

Of the 109HL measures found in this review, 58 (53%) assess general HL, and 51 (47%) assess HL within a specific content/ context. Seventeen (16%) of the 109 measures are shortened versions of an original HL measure, 22 (20%) are "REALM-like", and 15 (14%) are "TOFHLA-like". The topics covered in the content/ context-specific HL measures vary widely; ranging from arthritis to vascular surgery, with several focusing on dentistry, diabetes, cancer, nutrition, HIV, and eHealth.

Seventy-two were English language HL measures and 37 were non-English HL measures. Of the 72 English language measures identified, 17 did not specify the racial/ethnic characteristic of their sample. Of the remaining 55 measures, 10 (18%) did not include any blacks, 30 (55%) did not include any Hispanics, and 35 (64%) did not include any Asians in their validation sample. When Hispanic and Asian Americans were included, they accounted for small percentages and numbers in the overall sample; (inter-quartile range = 10%–34% [n = 13–154]) and (inter-quartile range = 3.5%–16% [n = 5–36]), respectively.

Between 2005–2013 a growing number of ethnic specific translations of REALM (n = 5) [21,27,28,58,59] and TOFHLA (n = 10) [31–35,37,66] were identified. However, there were notable failed attempts to validate "REALM-like" measures due to the phoneme–grapheme correspondence of some languages (e.g., Spanish) [27,42,97,111]. Similarly, issues of cultural equivalency was noted as a challenge in translating "TOFHLA-like" measures since some items on the TOFHLA are specific to healthcare in the US. As a result, new HL measures for specific cultural/linguistic groups within and outside the US emerged in the literature over the last few years [41,42,45,52,55,56,61,64,65,67,83,84,87,88,90,92,97,101,102,106,108,110]. Among these 37 non-English language measures, only two specified the racial/ethnic characteristics of their sample beyond simply describing the *general* population in which the measure was being validated [35,64]. For example, Ko et al. [35] specified that the sample used to validate their "Health Literacy Test for Singapore" was 52% Chinese, 22% Malay, 24% Indian, and 10% other. In contrast, most other non-English measures were similar to the

"Hebrew Health Literacy Test," which simply reported that 119 Israeli participants were sampled to validate the scale; it is unclear which ethno-linguistic groups from this highly diverse society were represented. Among the 37 non-English HL measures, there were slightly more general HL measures (n = 20, 54%) as compared to content/context specific measures (n = 17, 46%).

With regards to the psychometric properties of the HL measures, the majority of measures (n = 83, 89%) used analysis consistent with Classical Test Theory (CTT) for reliability testing and/or item reduction, while a small but growing number reported using analysis consistent with newer advanced approaches such as Rasch and Item Response Theory (IRT) either alone or adjunctively(n = 16, 15%) with CTT [4,27,28,39,47,52,54,56,61,62,64,67,78,101,108]. The reported reliability was generally strong across most measures (Cronbach's a or KR-20 = 0.70–0.98), with the exception of a few measures [64,65,81,107]. Eleven measures did not report reliability [22,38,58,69,79,99,100,112–115].

The evidence supporting the validity of these HL measures was weaker than the evidence supporting reliability. Whereas the convergent validity of most measures was positive and significant (r = 0.32–0.98), other concurrent construct and predictive validation tests were either not performed, non-significant or only weakly associated. Moreover, the HL measures that used receiver operating curve (ROC) analysis to assess validity often used as the "gold standard" other HL measures that are not comprehensive and/or do not themselves have strong evidence of predictive validity. Fourteen measures did not report validity testing (beyond content validation) [36,37,39,43,56,60,82,92,98,101–103]. Overall, the trends in psychometric findings were similar between the general and content/context-specific HL measures. However, more non-significant associations were reported in the validity testing of content/context-specific HL measures.

4. Discussion and conclusions

4.1. Discussion

In response to numerous scientific calls and proposals for additional measures of HL in recent years [116,117], the literature has proliferated. However, there has been a significant methodological gap that could undermine the accuracy of a large amount of the research in the field. This finding is based on a systematic review of 109 measures of HL, 58 more than the most recent review of HL measures [15]; and examines aspects of validity prior reviews have not [13,14].

Based on this study, several conclusions can be drawn. First, there is a growing body of cultural/linguistic specific HL measures that have been previously tested for use with various minority groups. This is a valuable contribution to the literature. The summary tables developed in this review can help researchers and clinicians identify appropriate HL measure(s) for their specific population, clinical setting, or program of research. However, almost none of the non-English HL measures provide details regarding the ethnic/cultural characteristic of their validation sample beyond a basic general description. The ethno-linguistic and cultural diversity of a given population will influence the extent to which this

may pose a problem for any particular project. For example, this issue is less relevant for Korean language HL measures since the population is relatively linguistically homogenous. Whereas, it may be a greater concern for the use of Hebrew or Hindi language HL measures, as these populations are more linguistically and culturally diverse. Future validations of non-English HL measures should keep the cultural diversity of their target population in mind and include such details in reports. Furthermore, sampling strategies should reflect the needs of high risk groups and be described in sufficient detail in validation reports, as this information has implications for a measure's generalizability.

Another important finding was that non-English HL tools represent a small, but growing, proportion of existing HL measures. Most existing HL measures were developed and validated in English. Among them, there is a strong bias towards the validation of measures in White and Black populations. Using tools that are not well-validated for a given population can lead to substantive measurement error. This problem is of particular concern because most of the psychometric testing of existing HL measures has been done using methods grounded in CTT, resulting in findings that are highly dependent on the samples used for instrument validation [17].

The problem of limited generalizability partially accounts for the growing number of HL measures, as many non-English HL measures cite the “lack of validated measures,” as one of the When warranted, developing a new HL measure is a logical next step for researchers and clinicians who cannot find a valid and reliable measure for their target minority group. However, carefully considering the need to develop a new measure is important given the large number of HL measures already available. This is critical since considerable resource and effort are required to pilot and develop new measures. Moreover, using many different HL measures makes it challenging to compare findings across studies. A significant opportunity exists for the creation of an item bank to allow for comparisons across cultures [118]. Achieving this goal will necessarily require collaboration across broad-based groups of stakeholders and requires the use of IRT or Rasch modeling techniques [119]. A promising sign of progress towards these efforts is that a small and growing number of HL measures identified in this review incorporated IRT or Rasch modeling in their psychometric assessments. Continued work across stakeholder groups to refine analytic approaches, theoretical frameworks and definitions, and inclusion of appropriately diverse populations in validation activities will support coordinated efforts to improve HL measurement.

While efforts are made toward these goals, pilot testing is advised before using HL measures that have not been validated for a given target population. Several key considerations have been proposed when piloting measures for cross-cultural research [120,121]. An important first step is to ensure content validity. This can be done by building a bi-cultural/bi-lingual team to explore the meanings and language used by the target population to describe the underlying construct. Having a bi-cultural/bi-lingual team who is familiar with the content, language, and cultures being studied will help ensure that items are interpreted the same way across groups [121]. Another way of ensuring content validity is to engage in focus groups with members of the target group to elicit comments on words, phrases, or items that may be misunderstood and suggestions for rephrasing it [120]. This was particularly important for

tools like the TOFHLA, which includes items based on a passage from the patient "rights and responsibilities" section of a Medicaid application form. While relevant for U.S. residents, this section is not particularly relevant for those outside the U.S., and needs to be modified when used in that manner [31,35]. Similarly, researchers who adopted HL measures for ethnic minority populations in the U.S. have called for changes in certain words and phrases for better comprehension and cultural relevancy [37,97]. After content validity is assessed, translation (if necessary) and back translation is an appropriate next step. Once translation and back translation is complete, pre-testing the instrument through cognitive interviews with 15–30 persons from the target group will help identify revisions that may be needed. Specifically, these interviews serve to probe participants on what they thought was meant by each item, and to inquire about items they may have skipped. After revisions are made based on these interviews, psychometric testing (e.g. construct validity, reliability, or item response patterns) in a large sample can be conducted to further evaluate the measure.

In this review, almost half (47%) of the HL measures were content/context-specific. The rise in content/context-specific HL measures may be the result of increasing recognition that one's HL capacity can be driven by specific contexts and/or content [122]. That is, a person may have exhibit high levels of HL when it comes to interacting with healthcare providers about a common condition, but exhibit low HL when confronting an unfamiliar condition in a different setting. While it makes intuitive sense that HL can be influenced by content/context, empirical data on the validity of content/context-specific HL measures are weak. For example, the TOFHLiD was no better at predicting dental health outcomes than was the TOFHLA [74], and the use of the REALM and REALM-Arthritis resulted in similar findings—neither was associated with physical functioning in the context of arthritis, erythrocyte sedimentation rate or pain [79]. These results have led some to suggest that we should increase the use of general HL measures, such as the REALM, rather than developing additional content/context-specific HL measures [79].

Other authors who have developed content/context-specific HL measures argue that having a disease-specific HL measure reduces the shame and stigma associated with limited literacy, since respondents are not necessarily expected to know specialized medical terminology associated with a specific health condition [69,71]. Moreover, they argue that identification of patients' content-specific literacy deficits prior to initiating education can help develop tailor interventions that meet their literacy needs [69,97]. These considerations may be relevant when working with minority populations with low health literacy. Supporting these sentiments, almost half (n = 17) of the 37 non-English HL measures found in this review were content/context specific.

4.2. Limitations

Although a systematic approach was taken to locate and review the HL measures presented here, this study has several limitations. First, as with many reviews, it is always possible that the search terms and methods used were not sensitive enough to capture everything that is available in the literature. To address this concern, we made use of careful cross-referencing to ensure the inclusion of as many relevant HL measures as possible. The rigor of our search

methodology can be supported by the results, which yielded over 50% more HL measures as compared to the largest previous review [15]. A second limitation of this review was that we only included articles that were published in English. Validation studies of Non-English HL measures may be published in international journals that are not written in English. The search strategy used for this review only yielded one article that was not written in English and met our inclusion criteria, so the extent of this limitation is likely minimal. However, given the focus of this review, future reviews in this area should include non-English articles and use databases such as EMBASE, which includes a wider range of international scholarship. A third limitation is that we excluded measures for which no psychometric data were reported. While such measures can potentially be informative, without access to meaningful information about their reliability or validity, it is difficult to draw valid inferences about their application. A fourth potential limitation of this review is that we only assessed primary studies, additional validations of a particular HL measure could have been reported in subsequent validation studies. Tool validation can be thought of as an on-going process using validation data compiled from applications of the measure across time and populations. Collecting these data was beyond the scope of this study. However, HL measures identified from this review can be used to inform a public repository, or toolshed, where ongoing validation data can be updated and shared widely. Collecting validity data overtime using a public repository also allows for the possibility of monitoring measurement bias that may result from linguistic shifts and changes in population diversity that naturally occur.

5. Conclusions

This paper presents a comprehensive review of the literature on the validity of HL measures for minority populations. In particular, we found a large growth in the number HL measures; many of which were validated in relatively homogeneous populations. This is a threat to the generalizability of many tools and to the potential accuracy of the findings of many studies. Pragmatic solutions are needed. Developing a public repository of relevant tools and their associated validation data is an important first step. This will make it easier to find existing measures that are appropriate for a given population and health context, and allow for future validation studies to be collected and shared. A second step would be to collaborate across a broad-based group of stakeholders to create a robust appropriately validated item bank.

5.1. Practice implications

As the science of HL measurement advances, the limitations of current measures need to be overcome. Never the less, the tables developed from this review can help researchers and clinicians make decisions about which current HL measure best fits their needs. When using a HL measure that has not been validated for a given population, some pilot testing and/or validation should be done before wider use.

Acknowledgements

The authors would like to thank Ms. Stella Seal from the Johns Hopkins Welch Medical Library for her advisement in developing the search strategy.

Disclaimers

Funding to support this paper was provided by: The National Institutes of Nursing Research Individual Ruth L. Kirschstein National Research Service Award (F31 NR010992), The Johns Hopkins Center of Excellence for Cardiovascular Research in Vulnerable Populations (P30 NR011409), and The Johns Hopkins Center to Eliminate Cardiovascular Health Disparities Fellowship (P50HL105187).

Appendix A. PubMed search terms

PubMed search terms

(health literacy [mh] OR "health literacy" OR (health [tiab] AND (literate [tiab] OR literacy [tiab)) OR numeracy [tiab])

AND

(measur* [tw] OR language tests [mh] OR psychometrics [mh] OR tool* [tiab] OR survey* [tiab] OR instrument* [tiab] OR questionnaire* [tiab] OR screen* [tw] OR assessment [tiab])

Appendix B. CINAHL search terms

CINAHL search terms

((MH "Information Literacy") OR “health literacy”) or health N5 litera*

AND

((MH “Language Tests+”) OR “language tests” OR (MH “Psychometrics”) OR “psychometrics” OR (MH “Research Mea-surement+”) OR(MH “Measurement Issues and Assessments”) OR (MH “Research Instruments+”)) or (measur* OR tool* OR survey* OR instrument* OR questionnaire*)

Appendix C. PsycINFO search terms

PsycINFO search terms

DE “Health Literacy” OR “health literacy” OR health N5 litera*

AND

DE “Psychometrics” OR DE “Measurement” OR DE “Achievement Measures” OR DE “Aptitude Measures” OR DE “Attitude Measurement” OR DE “Attitude Measures” OR DE “Body Sway Testing” OR DE “Comprehension Tests” OR DE “Creativity Measurement” OR DE “Criterion Referenced Tests” OR DE “Digit Span Testing” OR DE “Employment Tests” OR DE “Group Testing” OR DE “Individual Testing” OR DE “Intelligence Measures” OR DE “Inventories” OR DE “Multidimensional Scaling” OR DE “Needs Assessment” OR DE “Occupational Interest Measures” OR DE “Pain Measurement” OR DE “Perceptual Measures” OR DE “Performance Tests” OR DE “Personality Measures” OR DE “Posttesting” OR DE “Preference Measures” OR DE “Pretesting” OR DE “Professional Examinations” OR DE “Profiles (Measurement)” OR DE “Projective Testing Technique” OR DE “Psychiatric Evaluation” OR DE “Psychological Assessment” OR DE

"Psychometrics" OR DE "Questionnaires" OR DE "Rating Scales" OR DE "Reading Measures" OR DE "Retention Measures" OR DE "Screening" OR DE "Screening Tests" OR DE "Selection Tests" OR DE "Sensorimotor Measures" OR DE "Sociometric Tests" OR DE "Speech and Hearing Measures" OR DE "Standardized Tests" OR DE "Statistical Measurement" OR DE "Subtests" OR DE "Surveys" OR DE "Symptom Checklists" OR DE "Testing" OR DE "Verbal Tests" OR measur* OR language N5 test* OR tool* OR survey* OR instrument* OR questionnaire*.

Appendix D

Acronym key for tables 1.1, 1.2, 1.3, 2.1, 2.2, 2.3

¶ SORT, SORT-R: Slosson Oral Reading Test, Slosson Oral Reading Test-Revised

§ PIAT: Peabody Individual Achievement Test-Revised (recognitiona and comprehensionb subscales)

€ WRAT, WRAT-R, WRAT-A: Wide Range Achievement Test, Wide Range Achievement Test-Revised, Wide Range Achievement Test-Arithmetic

β REALM: Rapid Estimate of Adult Literacy in Medicine

¢ TOFHLA: Test of Functional Health Literacy in Adults

¢S-TOFHLA: Short Test of Functional Health Literacy in Adults

¢TOFHLA-S: Test of Functional Health Literacy in Adults-Spanish Version

¢Ser-STOFHLA: Serbian Short Test of Functional Health Literacy in Adults

¢TOFHLiD: Test of Functional Health Literacy in Dentistry

¥ ROC: Receiver Operating Characteristic AUROC1: Area Under Receiver Operating Characteristic

Δ a = Item discrimination

□ b = Item difficulty

∞ DIF: Differential Item Functioning

ϕ CIBS-R: Comprehensive Inventory of Basic Skills-Revised (readinga, comprehensionb, mathematicsc, written languaged and listening comprehensione domains)

* Reference group: Low Health Literacy

° ONQI: Overall Nutritional Quality Index SILS: Single Item Literacy Screener

Δ PPCLA: Preschool and Primary Chinese Literacy Scale

» MMSE: Mini Mental Status Examination

µSAHLSA: Short Assessment of Health Literacy for Spanish-speaking Americans

± THLS: Taiwan Health Literacy Scale

÷ NVS: Newest Vital Sign

ÿ REALD-30, REALD-99: Rapid Estimate of Adult Literacy in Dentistry

SRDHS: Self-Reported Dental (Oral) Health Status

Þ OHIP-14: Oral Health Impact Profile-14 (assesses 'oral health quality of life')

Ω Pain VAS: Pain Visual Analogue Scale

ψ MDHAQ: Multidimensional Health Assessment Questionnaire (assesses 'physical functioning')

† CRP: C - Reactive Protein

ö ESR: Erythrocyte Sedimentation Rate

ɔ DKT: Diabetes Knowledge Test

à CVI: Content Validity Index

λ ONQI: Overall Nutritional Quality Index

⌂ Early Childhood Oral Health Impact Scale

≠ PPCLA: Preschool and Primary Chinese Literacy Scale

× HKREALD-30: Hong Kong Rapid Estimate of Adult Literacy in Dentistry

£ CFA: Confirmatory Factor Analysis RMSEA[1]: Root Mean Square Error of Approximation, GFI[2]: Goodness of Fit Index, AGFI[3]: Adjusted GFI, SRMR[4]: Standardized Root Means Square Residual, CFI[5]: Comparative Fit Index, NNFI[6]: Nonnormed Fit Index, TLI[7]: Tucker-Lewis Index, WRMR[8]: Weighted Root Mean Square Residual

References

1. Ratzan, SC.; Parker, RM. Introduction. In: Seldon, CR.; Zorn, M.; Ratzen, SC.; Parker, RM., editors. National Library of Medicine Current Bibliographies in Medicine: Health Literacy. Bethesda, MD: National Insitutes of Health, US Department of Health and Human Services; 2000.
2. Nielsen-Bohlman, L.; Panzer, AM.; Kindig, DA., editors. IOM. Health literacy: A prescription to end confusion. Washington, DC: The National Academies Press; 2004.
3. Kickbusch, I., editor. WHO, et al.. Health literacy: The solid facts. Copenhagen, Denmark: WHO Regional Office for Europe; 2013.
4. ETS. Health Activities Literacy Scale. 2012. Available from: http://www.ets.org/literacy/about/content/health_activities_content

5. Kurtner, M., editor. NECS, et al.. Health Literacy of American Adults: Results from the 2003 National Assessment of Adult Literacy (NALS). Washington, DC: National Center for Education Statisitics, US Department of Education; 2006.
6. Rudd, R.; Kirsch, I.; Yamamoto, K. Princeton, NJ: Literacy and Health in America, Educational Testing Service; 2004.
7. US Census Bureau. United States Foreign-Born Population. 2000
8. Kim MT, et al. Cardiovascular disease risk factors in Korean American elderly. West. J. Nurs. Res. 2001; 23:269–282. [PubMed: 11291431]
9. Young, JJ.; Gu, N. Demographic and Socio-Economic Characteristics of Elderly Asian and Pacific Island Americans. Seattle: National Asian Pacific Center on Aging; 1995.
10. Kim MT, et al. Constructing and testing a self-help intervention program for high blood pressure control in Korean American seniors—a pilot study. J. Cardiovasc. Nurs. 2006; 21:77–84. [PubMed: 16601523]
11. McPhee SJ. Caring for a 70-year-old vietnamese woman. JAMA. 2002; 287:495–504. [PubMed: 11798372]
12. Berkman ND, et al. Low health literacy and health outcomes: an updated systematic review. Ann. Intern. Med. 2011; 155:97–107. [PubMed: 21768583]
13. Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. Nurs. Health Sci. 2009; 11:77–89. [PubMed: 19298313]
14. Luk A, Aslani P. Tools used to evaluate written medicine and health information: document and user perspectives. Health Educ. Behav. 2011; 38:389–403. [PubMed: 21490309]
15. Haun JN, et al. Health literacy measurement: an inventory and descriptive summary of 51 instruments. J. Health Commun. 2014; 19:302–333. [PubMed: 25315600]
16. Netemeyer, RG.; Bearden, WO.; Sharma, S. Scaling Procedures. Thousand Oaks, CA: Sage; 2003.
17. Nunnelly, J.; Bernstein, IH. Psychometric Theory. 3rd ed.. New York: McGraw-Hill; 1994.
18. Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of Item Response theory. Newbury Park, CA: Sage Publications; 1991.
19. Lord FM. The relation of test score to the trait underlying the test. Educ. Psychol. Meas. 1953; 13:517–548.
20. Sorensen K, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012; 12:80. [PubMed: 22276600]
21. Apolinario D, et al. Short assessment of health literacy for portuguese-speaking adults. Rev. Saude. Publica. 2012; 46:702–711. [PubMed: 22782124]
22. Arozullah AM, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med. Care. 2007; 45:1026–1033. [PubMed: 18049342]
23. Bass PF 3rd, Wilson JF, Griffith CH. A shortened instrument for literacy screening. J. Gen. Intern. Med. 2003; 18:1036–1038. [PubMed: 14687263]
24. Davis TC, et al. Rapid assessment of literacy levels of adult primary care patients. Fam. Med. 1991; 23:433–435. [PubMed: 1936717]
25. Davis TC, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam. Med. 1993; 25:391–395. [PubMed: 8349060]
26. Hanson-Divers EC. Developing a medical achievement reading test to evaluate patient literacy skills: a preliminary study. J. Health Care Poor Underserved. 1997; 8:56–69. [PubMed: 9019026]
27. Lee SY, et al. Development of an easy-to-use Spanish Health Literacy test. Health Serv. Res. 2006; 41:1392–1412. [PubMed: 16899014]
28. Lee SY, et al. Short assessment of health literacy-spanish and english: a comparable test of health literacy for spanish and english speakers. Health Serv. Res. 2010; 45:1105–1120. [PubMed: 20500222]
29. Rawson KA, et al. The METER: a brief, self-administered measure of health literacy. J. Gen. Intern. Med. 2010; 25:67–71. [PubMed: 19885705]
30. Baker DW, et al. Development of a brief test to measure functional health literacy. Patient Educ. Couns. 1999; 38:33–42. [PubMed: 14528569]

31. Baron-Epel O, et al. Validation of a Hebrew health literacy test. Patient Educ. Couns. 2007; 67:235–239. [PubMed: 17386994]
32. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese ersion of short-form test of functional health literacy in adolescents. J. Clin. Nurs. 2012; 21:2429–2437. [PubMed: 22784219]
33. Connor M, Mantwill Schulz SPJ. Functional health literacy in Switzerland-validation of a German Italian and French health literacy test. Patient Educ. Couns. 2013; 90:12–17. [PubMed: 23089240]
34. Jovic-Vranes A, et al. Evaluation of a health literacy screening tool in primary care patients: evidence from Serbia. Health Promot. 2013 epub.
35. Ko Y, et al. Development and validation of a general health literacy test in Singapore. Health Promot. 2011 epub.
36. Parker RM, et al. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J. Gen. Intern. Med. 1995; 10:537–541. [PubMed: 8576769]
37. Rivero-Mendez M, et al. Internal consistency of the Spanish health literacy test (TOFHLA-SPR) for Puerto Rico. P. R. Health Sci. J. 2010; 29:49–53. [PubMed: 20222334]
38. Haun J, et al. Testing the BRIEF health literacy screening tool. Federal Practitioner. 2009
39. Bann CM, et al. The Health Literacy Skills Instrument: a 10-item short form. J. Health Commun. 2012; 17:191–202. [PubMed: 23030570]
40. Fagerlin A, et al. Measuring numeracy without a math test: development of the subjective numeracy scale. Med. Decis. Making. 2007; 27:672–680. [PubMed: 17641137]
41. Ishikawa H, et al. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot. 2008; 23:269–274.
42. Lee TW, et al. Testing health literacy skills in older Korean adults. Patient Educ. Couns. 2009; 75:302–307. [PubMed: 19394184]
43. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med. Decis. Making. 2001; 21:37–44. [PubMed: 11206945]
44. McCormack L, et al. Measuring health literacy: a pilot study of a new skills-based instrument. J. Health Commun. 2010; 15:51–71. [PubMed: 20845193]
45. Rowlands G, et al. Development and validation of a measure of health literacy in the UK: the newest vital sign. BMC Public Health. 2013; 13:116. [PubMed: 23391329]
46. Weiss BD, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann. Fam. Med. 2005; 3:514–522. [PubMed: 16338915]
47. Yost KJ, et al. Bilingual health literacy assessment using the Talking Touchscreen/la Pantalla Parlanchina: Development and pilot testing. Patient Educ. Couns. 2009; 75:295–301. [PubMed: 19386462]
48. Chinn D, McCarthy C. All Aspects of Health Literacy Scale (AAHLS): developing a tool to measure functional communicative and critical health literacy in primary healthcare settings. Patient Educ. Couns. 2013; 90:247–253. [PubMed: 23206659]
49. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med. Decis. Making. 2011; 31:444–457. [PubMed: 20671213]
50. Jordan JE, 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; 91:228–235. [PubMed: 23419326]
51. Kaphingst KA, et al. Health Literacy INDEX: development, reliability, and validity of a new tool for evaluating the health literacy demands of health information materials. J. Health Commun. 2012; 17:203–221. [PubMed: 23030571]
52. Nakagami K, et al. Development and validation of a new instrument for testing functional health literacy in Japanese adults. Nurs Health Sci. 2014; 16:201–208. [PubMed: 23991825]
53. Osborn CY, et al. Development and validation of the General Health Numeracy Test (GHNT). Patient Educ. Couns. 2013; 91:350–356. [PubMed: 23433635]
54. Osborne RH, et al. The grounded psychometric development and initial validation of the health literacy questionnaire (HLQ). BMC Public Health. 2013; 13:658. [PubMed: 23855504]
55. Tsai TI, et al. Methodology and validation of health literacy scale development in Taiwan. J. Health Commun. 2011; 16:50–61. [PubMed: 21058141]

56. Wang J, Schmid BD, Thombs MR. The swiss health literacy survey: development and psychometric properties of a multidimensional instrument to assess competencies for health. Health Expect. 2014; 17:396–417. [PubMed: 22390287]
57. Hahn EA, et al. Health literacy assessment using talking touchscreen technology (Health LiTT): a new item response theory-based measure of health literacy. J. Health Commun. 2011; 16:150–162. [PubMed: 21951249]
58. Pan FC. A short form of the Taiwan health literacy scale (THLS) for Chinese-speaking adults. Int. Schol. Sci. Res. Innov. 2010; 4:1054–1059.
59. Pan FC, Su CL, Chen CH. Development of a health literacy scale for Chinese-speaking adults in Taiwan. Int. Schol. Sci. Res. Innov. 2010; 4:1021–1027.
60. Samora J, Saunders Larson LRF. Medical vocabulary knowledge among hospital patients. J. Health Hum. Behav. 1961; 2:83–92.
61. Sauceda JA, et al. Medication literacy in Spanish and English: psychometric evaluation of a new assessment tool. J. Am. Pharm. Assoc. 2012; 52:e231–e240.
62. Schapira MM, et al. The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. Med. Decis. Making. 2012; 32:851–865. [PubMed: 22635285]
63. Schwartz LM, Woloshin Welch SHG. Can patients interpret health information? An assessment of the medical data interpretation test. Med. Decis. Making. 2005; 25:290–300. [PubMed: 15951456]
64. Sorensen K, 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; 13:948. [PubMed: 24112855]
65. Takahashi Y, et al. Measuring the ability to interpret medical information among the Japanese public and the relationship with inappropriate purchasing attitudes of health-related goods. Asia Pac. J. Public Health. 2011; 23:386–398. [PubMed: 19723690]
66. Zhang XH, et al. Development and validation of a functional health literacy test. Patient. 2009; 2:169–178. [PubMed: 22273168]
67. Steckelberg A, et al. How to measure critical health competences: development and validation of the critical health competence test (CHC test). Adv. Health Sci. Educ. Theory Pract. 2009; 14:11–22. [PubMed: 17902033]
68. Davis TC, et al. Development and validation of the rapid estimate of adolescent literacy in medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. Pediatrics. 2006; 118:e1707–e1714. [PubMed: 17142495]
69. Erby LH, et al. The rapid estimate of adult literacy in genetics (REAL-G): a means to assess literacy deficits in the context of genetics. Am. J. Med. Genet. A. 2008; 146A:174–181. [PubMed: 18076116]
70. Glascoe FP. Safety words inventory and literacy screener: standardization and validation. Clin. Pediatr. 2002; 41:697–704.
71. Nath CR, et al. Development and validation of a literacy assessment tool for persons with diabetes. Diabetes Educ. 2001; 27:857–864. [PubMed: 12211925]
72. Atchison KA, et al. Screening for oral health literacy in an urban dental clinic. J. Public Health Dent. 2010; 70:269–275. [PubMed: 20545829]
73. Diamond JJ. Development of a reliable and construct valid measure of nutritional literacy in adults. Nutr. J. 2007; 6:5–8. [PubMed: 17300716]
74. Gong DA, et al. Development and testing of the test of functional health literacy in dentistry (TOFHLiD). J. Public Health Dent. 2007; 67:105–112. [PubMed: 17557682]
75. Lee JY, et al. Development of a word recognition instrument to test health literacy in dentistry: the REALD-30-a brief communication. J. Public Health Dent. 2007; 67:94–98. [PubMed: 17557680]
76. Richman JA, et al. Evaluation of a word recognition instrument to test health literacy in dentistry: the REALD-99. J. Public Health Dent. 2007; 67:99–104. [PubMed: 17557681]
77. Sabbahi DA, et al. Development and evaluation of an oral health literacy instrument for adults. Community Dent. Oral Epidemiol. 2009; 37:451–462. [PubMed: 19740249]

78. Stucky BD, et al. Development of the two-stage rapid estimate of adult literacy in dentistry. Community Dent. Oral Epidemiol. 2011; 39:474–480. [PubMed: 21592170]

79. Swearingen CJ, et al. Screening for low literacy in a rheumatology setting: more than 10% of patients cannot read cartilage, diagnosis, rheumatologist, or symptom. J. Clin. Rheumatol. 2010; 16:359–364. [PubMed: 21085021]

80. Wallace LS, et al. Development and validation of the rapid estimate of adult literacy in vascular surgery (REAL_VS). Ann. Vasc. Surg. 2009; 23:446–452. [PubMed: 19059757]

81. Apter AJ, et al. Asthma numeracy skill and health literacy. J. Asthma. 2006; 43:705–710. [PubMed: 17092853]

82. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J. Med. Internet. Res. 2006; 8:27.

83. van der Vaart R, et al. Does the eHealth literacy scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J. Med. Internet Res. 2011; 13:86.

84. Koo M, Norman CD, Chang HM. Psychometric evaluation of a chinese version of the eHealth literacy scale (eHEALS) in school age children. Int. Electron. J. Health Educ. 2012; 15:29–36.

85. Primack BA, et al. Development and validation of a smoking media literacy scale for adolescents. Arch. Pediatr. Adolesc. Med. 2006; 160:369–374. [PubMed: 16585481]

86. Huizinga MM, et al. Development and validation of the diabetes numeracy test (DNT). BMC Health Serv. Res. 2008; 8:96. [PubMed: 18452617]

87. Ishikawa H, Takeuchi Yano TE. Measuring functional communicative and critical health literacy among diabetic patients. Diabetes Care. 2008; 31:874–879. [PubMed: 18299446]

88. Leung AY, et al. Development and validation of Chinese health literacy scale for diabetes. J. Clin. Nurs. 2013; 22:2090–2099. [PubMed: 23186320]

89. Mulvaney SA, et al. Validation of the diabetes numeracy test with adolescents with type 1 diabetes. J. Health Commun. 2013; 18:795–804. [PubMed: 23577642]

90. White 3rd RO, et al. Development and validation of a Spanish diabetes-specific numeracy measure: DNT-15 Latino. Diabetes Technol. Ther. 2011; 13:893–898. [PubMed: 21714674]

91. Kumar D, et al. Parental understanding of infant health information: health literacy, numeracy, and the parental health literacy activities test (PHLAT). Acad. Pediatr. 2010; 10:309–316. [PubMed: 20674532]

92. Rivera-Vasquez O, et al. A community-based approach to translating and testing cancer literacy assessment tools. J. Cancer. Educ. 2009; 24:319–325. [PubMed: 19838892]

93. Williams KP, Templin TN. Bringing the real world to psychometric evaluation of cervical cancer literacy assessments with Black Latina and Arab women in real-world settings. J. Cancer. Educ. 2013; 28:738–743. [PubMed: 24072456]

94. Williams KP, Templin Hines TNRD. Answering the call: a tool that measures functional breast cancer literacy. J. Health Commun. 2013; 18:1310–1325. [PubMed: 23905580]

95. Wu AD, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. Health Promot. 2010; 25:444–452.

96. Diviani N, Schulz PJ. First insights on the validity of the concept of cancer literacy: a test in a sample of Ticino (Switzerland) residents. Patient Educ. Couns. 2012; 87:152–159. [PubMed: 21945564]

97. Kim MT, et al. Development and validation of the high blood pressure-focused health literacy scale. Patient Educ. Couns. 2011; 87:165–170. [PubMed: 22030252]

98. Mazor KM, et al. Health literacy and cancer prevention: two new instruments to assess comprehension. Patient Educ. Couns. 2012; 88:54–60. [PubMed: 22244323]

99. Pendlimari R, et al. Assessment of colon cancer literacy in screening colonoscopy patients: a validation study. J. Surg. Res. 2012; 175:221–226. [PubMed: 21737097]

100. Gibbs H, Chapman-Novakofski K. Establishing content validity for the nutrition literacy assessment instrument. Prev. Chronic. Dis. 2013; 10:109.

101. Guttersrud O, Dalane JO, Pettersen S. Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific ‘critical nutrition literacy’ scales. Public Health Nutr. 2014; 17:877–883. [PubMed: 23472785]

102. Levin-Zamir D, Lemish D, Gofin R. Media Health Literacy (MHL): development and measurement of the concept among adolescents. Health Educ. Res. 2011; 26:323–335. [PubMed: 21422003]
103. Massey P, et al. Findings toward a multidimensional measure of adolescent health literacy. Am. J. Health Behav. 2013; 37:342–350. [PubMed: 23985181]
104. Ownby RL, et al. Development and validation of a brief computer-administered HIV-Related health literacy scale (HIV-HL). AIDS Behav. 2013; 17:710–718. [PubMed: 22961499]
105. Reynolds JS, et al. The validation of a food label literacy questionnaire for elementary school children. J. Nutr. Educ. Behav. 2012; 44:262–266. [PubMed: 22406011]
106. Wong HM, et al. VAlidation of the Hong Kong oral health literacy assessment task for paediatric dentistry (HKOHLAT-P). Int. J. Paediatr. Dent. 2013; 23:366–375. [PubMed: 23947421]
107. Brega AG, et al. Special diabetes program for Indians: reliability and validity of brief measures of print literacy and numeracy. Ethn. Dis. 2012; 22:207–214. [PubMed: 22764644]
108. Leung AY, et al. Development and validation of the Chinese health literacy scale for chronic care. J. Health Commun. 2013; 18:205–222. [PubMed: 24093357]
109. Osborn CY, et al. Health literacy in the context of HIV treatment: introducing the brief estimate of health knowledge and action (BEHKA)-HIV version. AIDS Behav. 2010; 14:181–188. [PubMed: 19023653]
110. Yin HS, et al. Assessment of health literacy and numeracy among Spanish-Speaking parents of young children: validation of the Spanish parental health literacy activities test (PHLAT Spanish). Acad. Pediatr. 2012; 12:68–74. [PubMed: 22056223]
111. Han HR, et al. Measuring health literacy among immigrants with a phonetic primary language: a case of Korean American women. J. Immigr. Minor. Health. 2010; 13:253–259. [PubMed: 20585985]
112. Bennett IM, et al. Screening for low literacy among adult caregivers of pediatric patients. Fam. Med. 2003; 35:585–590. [PubMed: 12947522]
113. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam. Med. 2004; 36:588–594. [PubMed: 15343421]
114. Morris NS, et al. The single item literacy screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam. Pract. 2006; 7:21. [PubMed: 16563164]
115. Sanders LM, et al. Number of children's books in the home: an indicator of parent health literacy. Ambul. Pediatr. 2004; 4:424–428. [PubMed: 15369414]
116. Baker DW. The meaning and the measure of health literacy. J. Gen. Intern. Med. 2006; 21:878–883. [PubMed: 16881951]
117. NIH, AHRQ, and CDC. [[cited 2012 January 12]] Understanding and Promoting Health Literacy (R01). 2004. Available from: http://grants.nih.gov/grants/guide/pa-files/PAR-07-020.html
118. Pleasant A, McKinney J, Rikard RV. Health literacy measurement: a proposed research agenda. J. Health Commun. 2011; 16:11–21. [PubMed: 21951240]
119. Embretson SE. The new rules of measurement. Psychol. Assess. 1996; 8:341–349.
120. Beaton D, et al. Recommendations for cross-cultural adaptation of health measures. American Academy of Orthopaedic Surgeons Institute for Work and Health. 2002 20151-34.
121. Kim, MT.; Han, H. Cultural Considerations in Research Instrumentation Development. In: Frank-Stromborg, M.; Olson, SJ., editors. Instruments for Clinical Health-Care Research. Sudbury, Massachusetts: Jones and Bartlett Publishers; 2004. p. 73-80.
122. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies. J. Public Health. 2009:303–305.

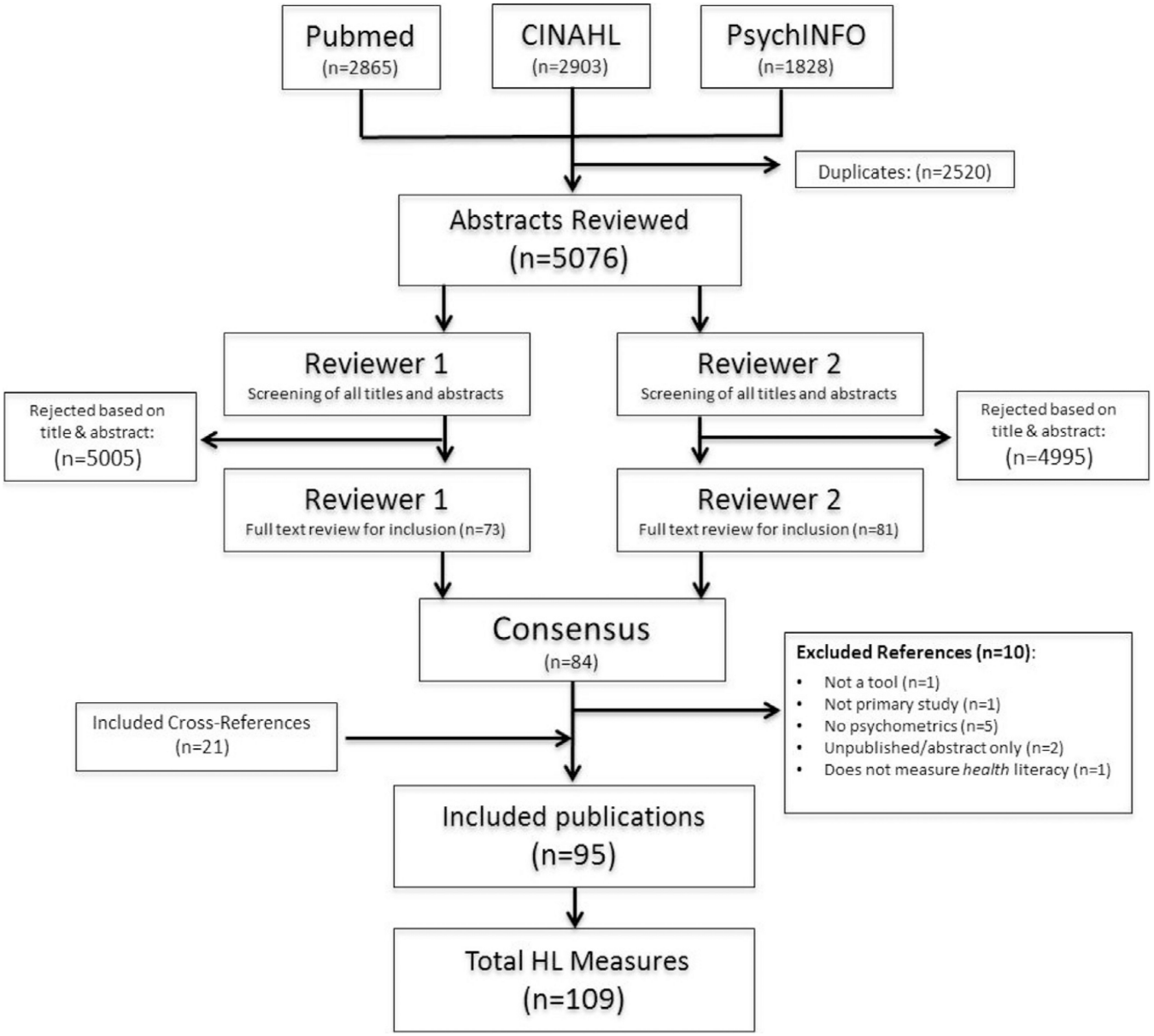

Fig. 1.
Search strategy for the systematic review.

Table 1

General health literacy measures: REALM and REALM 'like'.

Author, Year	Instrument name	Description (items)	Scoring	Time, mins	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Davis et al., 1991	Rapid Estimate of Adult Literacy in Medicine (REALM)	General list of medical words in descending levels of difficulty (125-items)	Possible range: 0–125 Low HL 0–78: ≤3rd grade 79–103: 4th-6th grade 104–114: 7th-8th grade 115–125: ≥9th grade	~3–5	Test-retest: r = .98 Inter-rater: r = .99	Convergent: SORT[¶], r = .95 PIAT-R[§], r = .94[a],.81[b]	[US] (n = 207) 54% Black 46% White	Print
Davis et al., 1993	REALM (Shortened)	General list of medical words in descending levels of difficulty, shortened (66-items)	Possible range: 0–66 0–18: ≤3rd grade 19–44: 4th-6th grade 45–60: 7th-8th grade 61–66: ≥9th grade	~1–3	Test-retest: r = 0.99	Convergent: SORT-R[¶], r = .96 WRAT-R[€], r = .88 PIAT-R[§], r = .97[a]	[US] (n = 203) 76% Black 24% White	Print
Bass et al., 2003	REALM-R (Revised)	General list of medical words in descending levels of difficulty, shortened (8-items)	Possible range: 0–8 ↑scores = ↑HL	<2	Internal Consistency: α = .91	Convergent: WRAT-R[€], r = .64	[US] (n = 157) Unspecified	Print
Arozullah et al., 2007	REALM-SF (Shorten Form)	General list of medical words in descending levels of difficulty, shortened (7-items)	Possible range: 0–7 0: ≤3rd grade 1–3: 4th-6th grade 4–6: 7th-8th grade 7: ≥9th grade	<1	N/A	Convergent: REALMβ, r = .96 WRAT-R[€], r = .83 CFA[£]: R^2 = .92.1% X^2 = 14.54(p = .41)	[US] (n = 1500) 65% Black 30% White 5% 'Other'	Print
Hanson-Divers, 1997	Medical Achievement Reading Test (MART)	General medical word recognition test modeled after the WRAT[€], with small font size and glossy covering to resemble prescription bottles (42-items)	Possible range: 0–42 "Converted into grade level equivalences similar to the WRAT"	<5	Internal Consistency: α = 0.98	Convergent: WRAT[€], r = 0.98	[US] (n = 405) 38% Black 65% White 6% 'Other'	Print
Lee et al., 2006	Rapid Estimate of Health Literacy for Spanish-speaking Adults (SAHLSA-50)	Spanish general medical word recognition and matching test (50-items)	Possible range: 0–50 ≤37 = inadequate HL	~3–6	Test-retest: r = .86 Internal Consistency: α = .92 IRT: a[Δ] = 2.80–0.68 b[□] = −2.98–0.75	Convergent: TOFHLA[¢], r = .65 REALMβ, r = .76 Concurrent: Physical health status, β = .17	[US] (n = 403) 50% Spanish-speaking 50% English-speaking	Print
Apolinario et al., 2012	Short Assessment of Health Literacy for Portuguese Adults (SAHLPA)	Portuguese general medical word recognition and matching test (18-items)	Possible range: 0–18 ≤14 = inadequate HL	~3–6	Test-retest: ICC = .91 Internal Consistency: α = .90	Concurrent: Edu, r = .65 Self-report literacy, r = .76 MMSE[»], r = 0.63	[Brazil] (n = 226) 100% Portuguese	Print

Author, Year	Instrument name	Description (items)	Scoring	Time, mins	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Lee et al., 2010	Short Assessment of Health Literacy-Spanish and English (SAHL-S&E)	General medical word recognition and matching test, shortened (18-items)	Possible range: 0–18 ≤14 = inadequate HL	~2–3	Internal Consistency: $\alpha = 0.80_{Spanishversion}$ $\alpha = 0.89_{Englishversion}$ IRT: Used DIF analysis	Convergent (SAHL-S): SAHLSA[μ], *r* = .88 TOFHLA-S[¢], *r* = .62 Convergent (SAHL-E): REALMβ, *r* = .94 TOFHLA[¢], *r* = .68	[US] (*n* =403) 50% Spanish-speaking 50% English-speaking	Print
Rawson, et al., 2010	Medical Term Recognition Test (METER)	General medical word recognition test that includes 40 real and 40'fake' words (80-items)	Possible range: 0–40 0–20: low HL 210–34: marginal HL 35–40: functional HL	~2	Internal Consistency: α = .93	Convergent: REALMβ, *r* = .74Concurrent: Trial-Making, *r* = −.42 CDV health, *r* = .21	[US] (*n* = 148) 93% White 7% 'Other'	Print
Pan et al., 2010	Taiwan Health Literacy Scale for Adults (THLS)	General medical word recognition test that reflects health terms used in Taiwan, and includes a subjective 5-point rating scale of respondent's perceived word comprehension (66-items)	Unclear3.0 = acceptable HL2.0 =moderate HL1.7 = poor HL	N/A	Internal Consistency: α = .98	Concurrent: Education, *p* < .05	[Taiwan] (*n* = 681) Taiwanese	Print
Pan, 2010	Shortened Taiwan Health Literacy Scale for Adults (STHLS)	General medical word recognition test that reflects health words used in Taiwan, and includes a subjective rating of word comprehension shortened (15-items)	Unclear	N/A	N/A	Convergent: THLS±, *r* = .98	[Taiwan] (*n* = 839) Taiwanese	Print

*Please see Appendix D for acronym key.

Table 2

General health literacy measures: TOFHLA and TOFHLA 'like'.

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Parker et al., 1995	Test of Functional Health Literacy in Adults (TOFHLA)	General functional HL test using 50 fill in the blank comprehension questions and 17 numeric calculations (67-items)	Possible range: 0–100 ↑scores=↑HL (The numeracy score is multiplied by 2.941 to create a score range of 0–50, which is added to the sum of the comprehension score)	~22	Internal Consistency: α = .98	Convergent: REALM$^{\beta}$ *r* = .84 WRAT$^{€}$ *r* = .74	[US] (*n* = 200) 91% Black 7% White 1% Hispanic	Functional HL and Numeracy
Parker et al., 1995	Test of Functional Health Literacy in Adults, Spanish version (TOFHLA-S)	General functional HL test using fill in the blank comprehension questions and numeric calculations (67-items)	Possible range: 0–100 ↑scores=↑HL (The numeracy score is multiplied by 2.941 to create a score range of 0–50, which is added to the sum of the comprehension score)	~22	Internal Consistency: α = .98	N/A	[US] (*n* = 203) 99% Hispanic	Functional HL and Numeracy
Baker et al., 1999	Short Test of Functional Health Literacy in Adults (S-TOFHLA)	Shortened version of the TOFHLA, with 36 comprehension questions and 4 numeric calculations, shortened (40-items)	Possible range: 0–100 (2pt for each numeracy score, 7pts for each comprehension score) 0–53: inadequate HL 54–66: marginal HL 67–100: adequate HL	~12	Internal Consistency: α = .97	Convergent: REALM$^{\beta}$ *r* = .80	[US] (*n* = 211) 94% Black 6% 'Other'	Functional HL and Numeracy
Baron-Epel, et al., 2007	Hebrew Health Literacy Test (HHLT)	Hebrew translation of the S-TOFHLA, with modifications to accommodate Israeli culture and context (12-items)	Possible range: 0–12 0–2: low HL 3–10: marginal HL 11 –12: high HL	N/A	Internal Consistency: α = .98	Convergent: Subjective reading ability on 10-pt Likert scale, *r* = .80	[Israel] (*n* = 119) 100% Israeli	Functional HL and Numeracy
Zhang et al., 2009	Functional Health Literacy Test (FHLT)	General functional HL test using fill in the blank comprehension questions (i.e. modified cloze) (21-item)	Possible range: 0–21 ↑scores=↑HL	~3min	Internal Consistency: α = .68$_{\text{general public}}$ α = .72$_{\text{rheumatic patients}}$ Test-retest: ICC= .95	Convergent: REALM$^{\beta}$ *r* = .65 $_{\text{public}}$ REALM$^{\beta}$ *r* = .68$_{\text{rheumatic}}$ Divergent: Edu, *r* = .33$_{\text{public}}$ Edu, *r* = .28 rheumatic	[Singapore] (*n* = 223) English speaking Singaporeans	Functional HL and Numeracy
Rivero-	Mendez et al., 2010	Test of Functional Health Literacy in Adults, Spanish version- for Puerto Ricans (TOFHLA-SPR)	Puerto Rican adaptation of the TOFHLA-S (67-items)		Possible range: 0–100 Similar to TOFHLA-S	~20	Internal Consistency: α = .95	N/A

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
[Puerto Rico] (*n* = 30) 100% Puerto Rican	Functional HL and Numeracy							
Ko et al., 2011	Health Literacy Test for Singapore	Singaporean translation of the S-TOFHLA, with modifications to accommodate Singaporean culture and context (40-items)	Possible numeracy: 0–4 ≥3: adequate HL Possible comprehension: 0–36 ≥27: adequate HL	N/A	Internal Consistency: α = .87	Convergent: NVS[÷] *r* = .55	[Singapore] (*n* = 302) 52% Chinese 22% Malay 24% Indian 10% 'Other'	Functional HL and Numeracy
Chang et al., 2012	Chinese Short Form Test of Functional Health Literacy in Adolescents (c-sTOFHLAd)	Chinese translation of the Shortened version of the TOFHLA, using fill in the blank comprehension questions (36-items)	Possible range: 0–36 ↑scores=↑HL	~11	Internal Consistency: α= .85 Test-retest: *r* = .95 (*p* < 0.001)	Convergent: REALM[β] *r* = .74 CFA[£] RMSEA[1] =90% X^2= 2335 (*p* < .001) GFI[2] = .92 AGFI[3] = .90 SRMR[4]= .068 CFI[5] = .81	[Taiwan] (*n* = 300) 100% Chinese	Functional HL
Jovic-Vranes et al., 2013	Test of Functional Health Literacy in Adults, Serbian version (Ser-TOFHLA)	Serbian version of the General functional HL test using 50 fill in the blank comprehension questions and 17 numeric calculations (67-items)	Possible range: 0–100 ↑scores=↑HL (The numeracy score is multiplied by 2.941 to create a score range of 0–50, which is added to the sum of the comprehension score)	~22	Internal Consistency: α = .94	Convergent: Ser-STOFHLA[¢] *r* = .65 ROC[¥] analysis: vs. Ser-STOFHLA[¢] AUROC[¥1] = .79	[Serbia] (*n* = 120) 100% Serbians	Functional HL and Numeracy
Jovic-Vranes et al., 2013	Short Test of Functional Health Literacy in Adults, Serbian version (Ser-STOFHLA)	Serbian version of the Shortened version of the TOFHLA, with 36 comprehension questions and 4 numeric calculations, shortened (40-items)	Possible range: 0–100 (2pt for each numeracy score, 7pts for each comprehension score) 0–53: inadequate HL 54–66: marginal HL 67–100: adequate HL	~12	Internal Consistency: α = .90	Convergent: Ser-TOFHLA[¢] *r* = .65 ROC[¥] analysis: vs. Ser-TOFHLA[¢] AUROC[¥1] = .79	[Serbia] (*n* = 120) 100% Serbians	Functional HL and Numeracy
Connor et al., 2013	Short Test of Functional Health Literacy in Adults, German version (Ger-STOFHLA)	German version of the Shortened version of the TOFHLA, with 36 comprehension questions and 4 numeric calculations, shortened (40-items)	Possible range: 0–100 (2pt for each numeracy score, 7pts for each comprehension score) 0–53: inadequate HL 54–66: marginal HL 67–100: adequate HL	~12	Internal Consistency: α = .73 (comp) α = .33 (numeracy)	Concurrent: Age (*p* < .001) Edu (*p* < .001)	[Switzerland] (*n* = 249) 100% German	Functional HL and Numeracy
Connor et al., 2013	Short Test of Functional Health Literacy in Adults,	Italian version of the Shortened version of the TOFHLA, with	Possible range: 0–100 (2pt for each numeracy score, 7pts for each	~12	Internal Consistency: α = .88 (comp)	Concurrent: Age (*p* < .001) Edu (*p* < .001)	[Switzerland] (*n* = 273) 100% Italian	Functional HL and Numeracy

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
	Italian version (Ity-STOFHLA)	36 comprehension questions and 4 numeric calculations, shortened (40-items)	comprehension score) 0–53: inadequate HL 54–66: marginal HL 67–100: adequate HL		α = .62 (numeracy)			
Connor et al., 2013	Short Test of Functional Health Literacy in Adults, French version (Fren-STOFHLA)	French version of the Shortened version of the TOFHLA, with 36 comprehension questions and 4 numeric calculations, shortened (40-items)	Possible range: 0–100 (2pt for each numeracy score, 7pts for each comprehension score) 0–53: inadequate HL 54–66: marginal HL 67–100: adequate HL	~12	Internal Consistency: α = .61 (comp) α = .80 (numeracy)	Concurrent: Age ($p < .042$) Edu ($p < .335$)	[Switzerland] ($n = 137$) 100% French	Functional HL and Numeracy

*Please see Appendix D for acronym key

Table 3

General health literacy measures: Unique types.

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Samora, 1961	Comprehension of 50 medical terms	A list of 50 words, each of which was placed in a simple sentence. Understanding of each word was assessed in context (50-items)	Responses categorized into one of four categories from no understanding to complete understanding	N/A	Inter-rater reliability =96%	N/A	[US] (n = 125) ~50% White ~25% Black ~25% Hispanic	Compre-hension
Lipkus et al., 2001	Numeracy Scale	Numeracy scale that includes 3 general calculation questions and 7 calculation questions in the context of health risks (10-items)	Possible range: 0–10 ↑scores=↑HL	~8min	Internal Consistency: α = .78	N/A	[US] (n = 463) 22% Black 76% White 2% 'Other'	Numeracy
Bennett et al., 2003	Screening Questions for Low Literacy	Screening questions used to identify low HL among adult caregivers of pediatric patients (3-items)	Possible range: 0–3 <1: adequate HL ≥2: low HL	~1–2	N/A	ROC[¥] analysis: vs. REALM[β] AUROC[¥1]= .54 Cut-off score= 2 84% sensitivity, 56% specificity	[US] (n = 98) 91% Black 9% Other	Unspecified
Weiss et al., 2005	Newest Vital Sign (NVS)	General health literacy test using an ice cream nutritional label (6-items) Spanish versionalso available	Possible range: 0–6 0–4: inadequate HL 5–6: unlikely to have low HL	~3min	Internal Consistency: α = .76$_{\text{English version}}$ α = .76$_{\text{Spanish version}}$	Convergent: TOFHLA[¢], r = .56 ROC[¥] analysis: vs. TOFHLA[¢] AUROC = .88$_{\text{English ver.}}$ AUROC = .72$_{\text{Spanish ver.}}$ Cut-off score= 4 10 0% sensitivity 64% specificity	[US] (n = 148) 93% White 7% 'Other'	Print
Rowlands et al., 2013	Newest Vital Sign UK Version (NVS-UK)	General health literacy test using an ice cream nutritional label, using the metric scale (6-items)	Possible range: 0–6 0–4: inadequate HL 5–6: unlikely to have low HL	~3min	Internal Consistency: α = .76$_{\text{English}}$	Convergent: TOFHLA[¢], r = .49 ROC[¥] analysis: vs. TOFHLA[¢]: AUROC[¥1]= .81	[UK] (n = 337) unspecified	Print
NCES, 2006	National Assessment of Adult Literacy (NAALS)	Task based HL scale that includes12 prose, 12 document and 4 quantitative items organized around	Result reported as 'Performance Levels: Below Basic Basic Intermediate Proficient	~1h	Test-retest: r = .98 Inter-rater: Kappa>.80 Internal Consistency: $r_b \geq$ .20	Extensive validity data published	[US] (n > 35,000) Representative of US population	Prose Document Quantitative

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
		3 domains of health- clinical, prevention and navigation (28-items)	*(Please see citation for specifics)		IRT analysis: α^{Δ}= .41–4.60 $b^{\square}$ = −6.34–1.77			
Fagerlin et al., 2007	Subjective Numeracy Scale (SNS)	General numeracy scale measured by subjective belief in ones' ability and preference regarding presentation of numeric information (8-items)	Possible range: 0–8 ↑scores=↑HL	~5min	Internal Consistency: α = .82	Convergent: Numeracy Scale, *r* = .68	[US] (*n* = 287) 72% White 28% 'Other'	Numeracy
Sanders et al., 2004	Number of Children's Book	How many children's books do you have in your home?	Open ended <9 books: low HL ≥10 books: adequate HL	~1–2	N/A	ROC¥ analysis: vs. S-TOFHLA¢ AUROC¥1= .54 Cut-off score = 10 books PPV= .91 NPV= .24	[US] (*n* = 163) 36% Black 20% White 35% Spanish 9% Haitian	Unspecified
Chew et al., 2004	HL Literacy Screening Items (HLSI)	Screening questions used to identify people with inadequate or marginal HL (3-items)	5-point Likert	~1–2	N/A	ROC¥ analysis: vs. S-TOFHLA¢ AUROC¥1 = .87$_{\text{Help Read}}$ AUROC¥1 = .80$_{\text{Confident w/forms}}$ AUROC[1]= .76$_{\text{Problems}}$	[US] (*n* = 332) 10% Black 81% White 9% 'Other'	Unspecified
Morris et al., 2006		Shortened HLSI: "How often do you	5-Pt Likert, >2=limited HL	<1	N/A	learning ROC¥ analysis: vs. S-TOFHLA¢		Unspecified
	Single Item Literacy Screener (SILS)	need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy?" (1-item)	1: never 2: rarely 3: sometimes 4: often 5: always			AUROC¥1 = .73 Cut-off score= 2 54% sensitivity 83% specificity	[US] (*n* = 969) 97% White 3% 'Other'	
Ishikawa et al., 2008	Communicative and Critical Health Literacy scale (CCHL)	Shortened and generalized HL measure previously developed by Ishikawa et al. (2008) that assesses how well participants can	Each item measured on a 5-point Likert (1 = strongly disagree, 5 = strongly agree) All items are	N/A	Internal Consistency: α = .86	Concurrent: Regular diet, OR* = 2.0 Exercise weekly, OR*=2.2 ↑Active coping style, *p* < .05 ↓Somatic symptoms,	[Japan] (*n* = 229) 10 0% Japanese men	Communicative Critical HL

Patient Educ Couns. Author manuscript; available in PMC 2017 January 29.

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
		extract, communicate and use information (5-items)	summed and averaged, with possible range: 1–5 <4: low HL ≥4: adequate HL			$p < .05$		
Haun et al., 2009	BRIEF Health Literacy Screening tool	3 screening items from Chew et al, with a 4th item added that assesses understanding of what was told to individual about his/her health	Scores range from 2 to 20 2–12 =inadequate HL 13–16 =marginal HL 17 –20=adequate HL	~2	N/A	Convergent: STOFHLA¢, $r = .42$ REALMβ, $r = .40$ ROC¥ analysis: vs. TOFHLA¢: AUROC¥1 = .68 vs. REALMβ: AUROC¥1 = .68 Cut-off score= 12 79% sensitivity	[US] (n = 378) 74% White 18 % Black 3% Latino 3% Native A 1% Other	Unspecified
Hahn et al., 2011	Health Literacy Talking Touchscreen (Health LiTT)	Computer adaptive test for health literacy (90-items in test bank), participants answer approximately 30 items	Multiple choice, correct/ incorrect Pilot of computer adoptive testing, min. and max. stop values unclear	N/A	IRT : b□ = .22–.96 (SD= .16) Internal Consistency: α= .72–.91	Concurrent: Race/ Ethnicity, $p < .05$ Education, $p < .05$ Income, p<.05 Computer use, $p < .05$	[US] (n =608) 67% Black 16% White 13%Hispanic 4% 'Other'	Prose Document Quantitative
Lee et al., 2009	Korean Health Literacy Scale (KHLS)	Korean general HL scale with short passages, pictures and graphs (24-items)	Possible range: 0–24 ↑scores=↑HL	~15–20	Internal Consistency: α = .89	CFA£: RMSEA1 = .039	[Korea] (n = 411) 100% Koreans	Print Numeracy
Mc-Cormack et al., 2010	Health Literacy Skills Instrument (HLSI)	Computer based measure that assesses 5 subscales: prose, document, quantitative, oral and internet seeking skills (25-items)	Possible range: 0–25 Multiple choice, unclear scoring interpretation	~30	Internal Consistency: α = .86	Convergent: TOFHLA¢ $r = .47$	[US] (n =332) 13% Black 64% White 17 % Hispanic 6% 'Other'	Prose, Document, Quantitative, Oral, Internet Seeking Skills
Bann et al., 2012	Health Literacy Skills Instrument, Short form (HLSI-SF)	Computer based measure that assesses 5 subscales: prose, document, quantitative, oral and internet seeking skills (10-items)	Possible range: 0–10 ↑scores=↑HL	~5–10	Internal Consistency: α = .70 IRT a = .64 to 2.46 b =1.88 to 2.06	NA	[US] (n =889) 13% Black 64% White 17 % Hispanic 6% 'Other'	Prose, Document, Quantitative, Oral, Internet Seeking Skills

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Galesic et al., 2011	Graph Literacy Scale	General graph literacy scale that assesses understanding of risk in health domain (13-items)	Possible range: 0–13 ↑scores=↑GL	~10	Internal Consistency: $\alpha = .74_{\text{German Sample}}$ $\alpha = .79_{\text{US Sample}}$	Convergent: IALS$^{\Omega}$+ NAALS+PISA$^{\phi}$ $r = .32_{\text{Germany}}$, $r = .50_{\text{US}}$ Concurrent: Education, $r = .29_{\text{German}}$ Education, $r = .54_{\text{US}}$ Divergent: S-TOFHLA$^{\text{¢}}$, $r = .19$	[Germany, US] (n =495)$_{\text{German}}$ (n = 492)$_{\text{US}}$ Unspecified race	Graph
Tsai, et al., 2011	Mandarin Health Literacy Scale	Mandarin HL measure that asses: health materials, outpatient dialogues, interpretation of prescription labels and health-related written information (50-items)	Possible range: 0–50 ↑scores=↑HL Multiple choice, correct/ incorrect	~25	Internal Consistency: $\alpha = .97$ Split-half Reliability: $r = .95$	Concurrent: Yrs. Of education, $r = .72$ Reading habit, $r = .34$ Health knowledge, $r = .55$ Health status, $r = .05$ [NS]	[Taiwan] (n = 323) 49% North Taiwanese 51% South Taiwanese	Prose Document Numeracy
ETS, 2012	Health Activities Literacy Scale (HALS)	Task based HL items in 5 health-related areas: health promotion, health protection, disease prevention, health maintenance, and systems navigation (191-item test bank)	Possible range: 0–500, ↑scores=↑HL Level 1: 0–225 Level2: 226–275 Level3: 276–325 Level4: 326–375 Level5: 378–500	~1h full ver. ~30-40 minlocator ver.	IRT analysis: Internal Consistency: Index of Expectation = .935	Concurrent: ↑HL, r/t ↑education ↑HL, r/t ↑newspaper read ↑HL, r/t birth in US ↑HL, r/t full-time employ ↑HL, r/t type of occupation	[US] (n =988) 24% Black 29% White 22% Hispanic 25% 'Other'	Prose Document Quantitative
Wang et al., 2012	Swiss Health Literacy Survey	HL questionnaire that assesses "citizen" and "patient-centered" competencies within 5 main dimensions (36-items, citizen + 31-items, patient)	Unclear	~30	Internal Consistency: $\alpha = .72_{\text{Information}}$ $\alpha = .81_{\text{Cognative}}$ $\alpha = .81_{\text{ICT}}$ $\alpha = .81_{\text{Health activste}}$ DIF$^{\infty}$ 6-items had sig. DIF	N/A	[Switzerland] (n = 1255) 52% German–Swiss 24% French–Swiss 24% Italian–Swiss	Information Cognitive Skills Social Roles Med Management Healthy Lifestyle
Osborne et al., 2012	Health Literacy Questionnaire (HLQ)	Comprehensive measure that assesses HL across individuals and organizations (44-items)	Unclear	N/A	Internal Consistency: $\alpha = .88_1$ $\alpha = .88_2$ $\alpha = .86_3$ $\alpha = .84_4$ $\alpha = .77_5$ $\alpha = .90_6$ $\alpha = .88_7$ $\alpha = .89_8$	CFA$^{\varepsilon}$: $X^2 = p < .05$ RMSEA[1] = .076 CFI[5] = .936 TLI[7] = .930 WRMR[8]= 1.698	[Australia] Initial validation (n = 634) 10 0% Australian Replication sample (n =405) 10 0% Australian unspecified	[1]Understood by provider [2]Have sufficient info to manage health [3]Actively manage health [4]Social support for health [5]Appraisal of health info

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
					$\alpha = .88_9$		race/ethnicity	[6]Actively engage with provider [7]Navigate health system [8]Find health info [9]Understand info enough to know what to do Plain language Clear purpose Supporting graphic User involvement Skill-based learning Audience Approp. User instructions Development details Eval. methods Strength of Evid.
Kaphingst et al., 2012	Health Literacy INDEX	Tool for evaluating the HL demands of health information material (63-indicators, organized into 10 criteria)	Each indicator scored as "Yes" or "No." Overall score is computed by taking proportion of "Ye s " out of 63. ↑scores=↓HL demands of material	N/A	Inter-rater: >80% for most indicators	Convergent: 12 HL experts, *r* = .89	[US] 300 print, websites, and A/V materials	
Osborn et al., 2013	General Health Numeracy Test (GHNT-21)	General numeracy scale that assesses: number hierarchy, performing calc., and estimating probabilities (21-items)	Possible range: 0–21, ↑scores=↑HL	~30	Internal Consistency: KR = .87	Concurrent: ↑HL, r/t ↑med understand, but not med adherence	[US] (*n* =205) 70% White 26% Black 4% Other	Numeracy
Osborn et al., 2013	General Health Numeracy Test (GHNT-6)	Shortened GHNT scale (6-items)	Possible range: 0–6, ↑scores=↑HL	~5–8	Internal Consistency: KR = .77	Concurrent: ↑HL, r/t ↑med understand, but not med adherence	[US] (*n* =205) 70% White 26% Black 4% Other	Numeracy

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Nakagami et al., 2013	Japanese Functional Health Literacy Test (JFHLT)	General HL measure that assesses numeracy, reading and comprehension, as well as experiences patients might experience in Japanese Hospital settings (16-items)	Possible range: 0–16, 0–10 =inadequate FHL 11 –13 =moderate FHL 14–16 =adequate FHL	~10–15	Internal Consistency: α = .81 IRT $b^{[smbol]}$ = − 1.08 to −3.87	Concurrent: Health knowledge, r = .37	[Japan] (n = 535) 10 0% Japanese	Reading Numeracy Comprehension
Chinn and McCarthy, 2013	All Aspects of Health Literacy Scale (AAHLS)	General measure of HL for use in the primary care setting (14-items)	Unclear	~7	Internal Consistency: $\alpha = .75_{TOT}$ $\alpha = .82_{Functional}$ $\alpha = .69_{Comm}$ $\alpha = .42_{Critial}$	Convergent: Functional vs Communicative, r = .39 Functional vs Critical, r = .59 Communicative vs Critical, r = .19	[US] (n =146) 56% Asian 3% Black 53% White 6% Other	Functional Communicative Critical HL
Jordan et al., 2013	Health Literacy Management Scale (HeLMS)	General measure of HL that assesses individual abilities, and their broader environment, to determine individual's ability to seek, understand, and use health information in health care settings (29-items)	Each item scored on a 5 point Likert scale (range 1–5). Scale scores are calculated as the mean score for each of 8 domains. ↑scores=↑HL	N/A	Internal Consistency: $\alpha = .87_1$ $\alpha = .82_2$ $\alpha = .87_3$ $\alpha = .83_4$ $\alpha = .83_4$ $\alpha = .87_5$ $\alpha = .86_6$ $\alpha = .89_7$ $\alpha = .86_8$ Test-retest: ICC = .73-.96	CFA[£]: RMSEA[1]= .07 SRMR[4] = .05 CFI[5]= .97	[Australia] Construct (n = 333) Replication (n = 350) 10 0% Australian, unspecified race/ethnicity	[1]Pt attitudes about health [2]Understand health info [3]Social support for health [4]Social-economic considerations [5]Accessing service [6]Communicate with provider [7]Being proactive [8]Use health info
Schwartz et al., 2005	Medical Data Interpretation Test	Measures individual's ability to make sense of and compare medical statistics about disease risk and risk reduction	Possible range: 0–10 0, ↑scores=↑HL	N/A	Internal Consistency: α = .71 Test-retest: r = .67	Concurrent: Education, p < .05 Numeracy, p <. 05 Quan Literacy, p < .05 Expertise, p < .05	[US] (n =178) unspecified	Numeracy
Steckelberg et al., 2007	Critical Health Competence Test (CHC)	General measure of health literacy and evidence-based health across 4 scenarios and 4 subareas of competence (72-items)	Unclear	~90 min	Rasch: Mean person parameter Scenario 1: 395 Scenario 2: 497 Scenario 3: 635 Scenario 4: 473	Construct: Cohen's d =4.33 [95% CI 3.51–5.16] WINMIRA ANOVA= .91	[Germany] (n = 300) German students	Critical HL

Author, Year	Instrument name	Description (items)	Scoring	Time, (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Takahashi et al., 2011	Test for Ability to Interpret Medical Information (TAIMI)	General measure of HL that assesses individual's ability to interpret medical information using multiple choice questions (7-items)	Possible range: 0–7, ↑scores=↑HL	N/A	Internal Consistency: $\alpha = .36_{\text{factor 1}}$ $\alpha = .51_{\text{factor 2}}$	Concurrent: Inappropriate purchasing attitudes of health related goods, $p < .05$	[Japan] (n = 6047) Japanese	Critical HL
Sauceda et al., 2012	Medication Literacy Assessment in Spanish and English ($MedLitR_xSE$)	General HL measure that assesses skills needed to manage medication appropriately using 4 "cases" (20-items)	Possible range: 0–20, ↑scores=↑HL	N/A	Internal Consistency: $KR20 = .78_{\text{English}}$ $KR20 = .80_{\text{Spanish}}$	CFA[£] RMSEA[1]= .05 SRMR[4] = .11 CFI[5]= .95 NNFI[6] = .97	[US] (n =181) White, Hispanic, "other" %unclear	Document Numeracy Prose
Sauceda et al., 2012	Shortened Medication Literacy Assessment in Spanish and English (S-$MedLitR_xSE$)	General HL measure that assesses skills needed to manage medication appropriately, shortened version (14-items)	Possible range: 0–14, ↑scores=↑HL	N/A	Internal Consistency: $KR20 = .81_{\text{English}}$ $KR20 = .77_{\text{Spanish}}$ IRT: No DIF in items	Concurrent: S-TOFHLA[¢], $p < .05$	[US] (n =62) White, Hispanic, "other" %unclear	Document Numeracy
Schapira et al., 2012	Numeracy Understanding in Medicine Instrument (NUMi)	20-Item test drawn from an item bank of 110 numeracy questions	Possible range: 0–20, 0–7 = low numeracy 8–12 =low-avg numeracy 12–17 =high-avg numeracy 18–20=high numeracy	N/A	Internal Consistency: α = .86 IRT a = .39–1.98 b = −1.70–1.09 No DIF in items	Concurrent: WRAT-A[€], r = .73 S-TOFHLA[¢], r = .43	[US] (n =1099) 46% White 42% Black 4% Asian 1% Native American 7% "other, or missing" 30% Hispanic	Numeracy
Sorensen et al., 2013	European Health Literacy Survey Questionnaire (HLS-EU-Q)	Comprehensive HL measure used to asses the relation between abilities, system demands, and decision making at the population level (47-items)	Possible range: 0–50, 0–25: inadequate HL 26–33: problematic HL 33–42: sufficient HL 42–50: excellent HL	12–15	Internal Consistency: α= .51–.91	Concurrent: NVS[÷] r = .73	[European Union] (n =99) Irish and Dutch % unclear	Accessing Understanding Appraising Apply

*Please see Appendix D for acronym key.

Table 4

Content/context-specific health literacy measures: REALM 'like'.

Author, Year	Instrument name	Description (items)	Scoring	Time (mins) (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Nath, 2001	Literacy Assessment tool for Diabetes (LAD)	Diabetes specific word recognition test listed in descending levels of difficulty (60-items)	Possible range: 0–60 0–20: ≤4th grade 21–40: 5th–9th grade 41–60: >9th grade	~3–5	Internal Consistency: α = .86 Test re-test: *r* = .86	Convergent: WRAT[€] *r* = .92 REALM[β] *r* = .9	[US] (*n* = 203) unspecified	Print
Glascoe, 2002	Safety Words Inventory and Literacy Screener (SWILS)	Recognition test for children using common safety survival sight words and phrases presented in the context of common logos (29-items)	Possible range: 0–29 Cut-off values varied based on child's age	~3–4	Internal Consistency: Guttman λ = .98	Convergent: CIBS-R[Φ] *r* = .54[a] .47[b], .44[c] .58[d], .38[e] ROC[¥] analysis: vs. CIBS-R[Φ] cut-off varied 73–88% sensitivity 77–87% specificity	[US] (*n* = 934) 12% Black 74% White 10% Hispanic 3% 'Other'	Print
Davis et al., 2006	REALM-teen	General medical word recognition test listed in descending levels of difficulty for teens/ adolescents (66-items)	Possible range: 0–66 0–37: ≤3rd grade 38–47: 4th–5th grade 48–58: 6th–7th grade 59–62: 8th–9th grade 63–66: ≥10th grade	~1–3	Internal Consistency: α = .94	Convergent: WRAT,[€] *r* = .83 SORT-R[¶] *r* = .93 ROC[¥] analysis: vs. SORT-R[¶] AUROC[¥1] = .84	[US] (*n* = 1533) 50% Black 49% White 1% 'Other' Teens	Print
Erby et al., 2008	Rapid Estimate of Adult Literacy in Genetics (REAL-G)	Genetic counseling specific word recognition test listed in descending levels of difficulty (63-items)	Possible range: 0–63 0–21: ≤4th grade 22–50: 4th–5th grade 48–58: 6th–7th grade 59–62: 8th–9th grade 63–66: ≥10th grade	~3–4	N/A	Convergent: REALM[β] *r* = .83 ROC[¥] analysis: vs. REALM[β] AUROC[¥1] = .92 Cutoff = 50 95% Sensitivity 89% Specificity	[US] (*n* = 203) 74% Black 17 % White 9% 'Other'	Print
Erby et al., 2008	Shortened Rapid Estimate of Adult Literacy in Genetics (REAL-G)	Genetic counseling specific word recognition test listed in descending levels of difficulty, shortened (8-items)	Possible range: 0–8 0–3: ≤6th grade 4–8: >6th grade	~3–4	N/A	Convergent: REALM[β] *r* = .80 ROC[¥]	[US] (*n* = 203) 74% Black 17 % White	Print

Author, Year	Instrument name	Description (items)	Scoring	Time (mins) (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
						analysis: vs. REALM[β] AUROC[¥1] = .90 Cutoff = 3 89% Sensitivity 80% Specificity	9% 'Other'	
Wallace et al., 2009	Rapid Estimate of Adult Literacy in Vascular Surgery (REAL_VS)	Vascular surgery specific word recognition test listed in descending levels of difficulty (75-items)	Possible range: 0–75 0–9: ≤4th grade 10–51: 4th–6th grade 52–67: 7th–8th grade 68–75: ≥9th grade	~3–4	Internal Consistency: α = .98	Convergent: REALM[β] *r* = .94 ROC[¥] analysis: vs. REALM[β] AUROC[¥1] = .98 Cutoff = 51 97% Sensitivity 83% Specificity	[US] (*n* = 152) 3% Black 97% White 1% 'Other'	Print
Wallace et al., 2009	Rapid Estimate of Adult Literacy in Vascular Surgery, shortened (REAL_VSs)	Vascular surgery specific word recognition test listed in descending levels of difficulty, shortened (8-items)	Possible range: 0–8 0–2: ≤6th grade 3–8: >6th grade	~1–2	Internal Consistency: α = .86	Convergent: REALM[β] *r* = .82 ROC[¥] analysis: vs. REALM[β] AUROC[¥1] = .91 Cutoff = 2 84% Sensitivity	[US] (*n* = 152) 3% Black 97% White 1% 'Other'	Print
Lee et al., 2007	Rapid Estimate of Adult Literacy in Dentistry (REALD-30)	Dentistry specific word recognition test listed in descending levels of difficulty (30-items)	Possible range: 0–30 ↑scores=↑HL	~2	Internal Consistency: α = .87	83% Specificity Convergent: REALM[β] *r* = .86 TOFHLA[¢] *r* = .64 Concurrent: OHIP-14[Þ] β =− .14 SRDHS, β = .35 (NS)	[US] (*n* = 202) unspecified	Print
Richman et al., 2007	Rapid Estimate of Adult Literacy in Dentistry (REALD-99)	Dentistry specific word recognition test listed in descending levels of difficulty, lengthened (99-items)	Possible range: 0–99 ↑scores=↑HL	~5	Internal Consistency: α = .86	Convergent: REALM[β] *r* = .80 Concurrent: OHIP-14[Þ] β =− .74 SRDHS,	[US] (*n* =102) 64% Black 21% White 5% Asian 11 % 'Other'	Print

Author, Year	Instrument name	Description (items)	Scoring	Time (mins) (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
						β = .05 (NS)		
Stucky et al., 2011	Two-Stage Rapid Estimate of Adult Literacy in Dentistry (TS_REALD)	Revised, two-stage dentistry specific word recognition test. (1st stage:5-items, 2nd stage: 3–6 items)	Possible range: 0–9 (raw score) For interpretation, raw scores are transformed, please see citation for additional detail	~1	IRT: a^{Δ} = 2.91-1.0 mean = 2.12, SD = .44 $b^{\square}$= 1.97 to −2.93 mean = −.02 SD= 1.42	Convergent: REALD-30$^{\ddot{y}}$ r = .96 NVS$^{\div}$ r = .51 Concurrent: OHIP-14$^{\flat}$ β = − .74 SRDHS, β = .10	[US] (n = 1405) 41% Black 42% White 11% A Indian 1% Asian	Print
Atchison et al., 2010	Rapid Estimate of Adult Literacy in Medicine and Dentistry (REALM-D)	Dentistry specific word recognition test, with general medical words listed in descending levels of difficulty (84-items)	Possible range: 0–84 ↑scores=↑HL	N/A	Internal Consistency: α = .96 Test-retest: r = .95	Convergent: REALM$^{\beta}$ r = .99 SILS, p < .05	[US] (n = 200) 11 % Black 57% White 19% Hispanic 9% Asian 3% ‘Other’	Print
Swearingen et al., 2010	Arthritis-Adapted Rapid Estimate of Adult Literacy in Medicine (A-REALM)	Arthritis adapted word recognition test listed in descending levels of difficulty (66-items)	Possible range: 0–66 (0–18: ≤3rd grade 19–44: 4th-6th grade Low) 45–60: 7th-8th grade Literacy 61–66: ≥9th grade	~2–3	N/A	Convergent: REALM$^{\beta}$ r = .94 Concurrent: MDHAQ$^{\Psi}$ (NS) CRP†, (NS) ESR$^{\ddot{o}}$, (NS) Pain VAS$^{\Omega}$ (NS)	[US] (n = 194) 94% White 6% ‘Other’	Print

Table 5

Content/context-specific health literacy measures: TOFHLA ‘like’.

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (admin)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Gong et al., 2007	Test of Functional Health Literacy in Dentistry (TOFHLiD)	Dentistry specific HL test using 62 fill in the blank Cloze type comprehension questions and 12 numeric calculations (67-items)	Possible range: 0–100, ↑scores=↑HL (The raw comprehension score is multiplied by .81, and the raw numeracy score is multiplied by 4.17 to create a score range of 0–10 0)	N/A	Internal Consistency: α = .63	Convergent: REALD-99[ÿ] *r* = .82 Discriminate: TOFHLA[¢] *r* = .52 REALM[β] *r* = .53 Concurrent: OHIP-14[Þ] β = −.50 SRDHS, β =−.0021, (NS)	[US] (*n* = 102) 64% Black 21% White 5% Asian 11 % ‘Other’	Functional HL and Numeracy
Sabbahi, 2009	Oral Health Literacy Instrument (OHLI)	Dentistry specific HL test using 38 fill in the blank Cloze type comprehension questions and 19 numeric calculations (57-items)	Possible range: 0–100 (comprehension scorex 1. 31, numeracy scorex2.63) 0–59: inadequate HL 60–74: marginal HL 75–100: adequate HL	~20	Internal Consistency: α = .85	Convergent: TOFHLA[¢] *r* = .61 Discriminate: Oral Knowledge, *r* = .57	[Canada] (*n* =100) unspecified	Functional HL and Numeracy
Diamond, 2007	Nutritional Literacy Scale (NLS)	Nutrition specific HLtest usingfill in the blank, Cloze type comprehension questions (28-items)	Possible range: 0–28, ↑scores=↑HL	N/A	Internal Consistency: α = .84	Convergent: S-TOFHLA[¢] *r* = .61	[US] (*n* = 341) 38% White 62% ‘Other’	Print

[*] Please see Appendix D for acronym key.

Table 6

Content/context-specific health literacy measures: Unique types

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Apter et al., 2006	Asthma Numeracy Questionnaire (ANQ)	Asthma numeracy specific HL test (4-items)	Possible range: 0–4, ↑scores = ↑HL	N/A	Internal consistency with first item: α = .57	Convergent: S-TOFHLA[¢] *r* = .47 REALM[β] *r* = .41	[US] (*n* =73) 58% Black 22% White 21% Hispanic 2% Asian 7% 'Other'	Numeracy
Norman et al., 2006	eHealth Literacy Scale (eHEALS)	Electronic health information specific measure comprised of six domains: traditional literacy, HL, information literacy, scientific literacy, media literacy and computer literacy (8-items)	Each item assessed on 5-point Likert scale Possible range: 8–40 ↑scores = ↑eHL	~75	Internal Consistency: α = .88 Test-retest: *r* = .40–.68 Item-scale: *r* = .51–.76	N/A	[Canada] (*n* = 664) 16% East European 16% East Asian 12% Central Asian 56% Unspecified	Print
Koo et al., 2012	Chinese eHealth Literacy Scale (C-eHEALS)	Chinese (Mandarin) version of the eHealth (8-items)	Each item assessed on 5-point Likert scale Possible range: 8–40 ↑scores = ↑eHL	N/A	Internal Consistency: α = .92	Concurrent: Use processor, *p* = .002 Internet use, *p* = .01 Download file, *p* = .007 Use computer, *p* = .01	[Taiwan] (*n* = 216) 100% Taiwanese 6th grade students	Print
Primack et al., 2006	Smoking Media Literacy (SML)	Smoking media, specific HL measure containing items r/t author/audience, massages/meaning and representation/reality (18-items)	Each item measured on a 5-point Likert scale, with possible raw score range of 0–54, then converted to possible score range of 0–10 bydividing raw score by 5.4 ↑scores = ↑SML	N/A	Internal Consistency: α = .87	Concurrent: Current smoking (*p* = .01) Susceptibility to smoking (*p* < .001) Anti-smoking attitudes (*p* < .001)	[US] (*n* = 1211) 4% Black 92% White 1% 'Other'	Print
Huizinga et al., 2008	Diabetes Numeracy Test (DNT)	Measure of numeracy skills used in diabetes, including topics r/t nutrition, exercise, glucose monitoring, oral meds and insulin use (43-items)	Scores reported as% correct Possible range: 0–100% ↑scores = ↑DN skills	~33	Internal Consistency: KR-20 = .95	DNT: Convergent: REALM[β] *r* = .54 WRAT-3R[€] *r* = .62 DKT[ↄ] *r* = .71 Concurrent: Education, *r* = .52	[US] (*n* = 2/398) 34% Black 63% White 3% 'Other'	Numeracy

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
						Income, *r* = .51 Insulin use, *r* = .04		
Huizinga et al., 2008	Shortened Diabetes Numeracy Test (DNT-15)	Shortened Measure of numeracy skills used in diabetes (15-items)	Scores reported as% correct Possible range: 0–100% ↑scores = ↑DN skills	~10–15	Internal Consistency: KR-20 = .89	DNT-15 Convergent: DNT, *r* = .97 REALM[β] *r* = .52 WRAT-3R[€] *r* = .61 DKT[ɔ] *r* = .67 Concurrent: Education, *r* = .52 Income, *r* = .49 Insulin use, *r* = -.02	[US] (*n* = 2/398) 34% Black 63% White3% 'Other'	Numeracy
White et al., 2011	Spanish Diabetes Numeracy Test (DNT-15 Latino)	Diabetes Numeracy test in Spanish version (15-items)	Scores reported as% correct Possible range:0–10 0% ↑scores = ↑DN skills	~23	Internal Consistency: KR-20 = .78	Convergent: S-TOFHLA[€] *r* = .29 WRAT-4[€] *r* = .46 Concurrent: Acculturation level, *r* = .27 Education level, *r* = .33 Income, *r* = .27	[US] (*n* =144) Hispanic	Numeracy
Mulvaney et al., 2013	Diabetes Numeracy Test for Adolescents with Type 1 DM (DNT-39)	Measure of numeracy skills used in type 1diabetes, including topics r/t nutrition, exercise, glucose monitoring, oral meds and insulin use (39-items)	Scores reported as% correct Possible range: 0–10 0% ↑scores = ↑DN skills	~30	Internal Consistency: KR-20 = .93	DNT-39: Convergent: DNT-14, *r* = .87 WRAT-3R[€] *r* = .40 Concurrent: Parent Education, *r* = .29 HgA1C, *r* = −.22 (N/S)	[US] (*n* =61) 85% White 12% Black 3% Hispanic	Numeracy
Mulvaney et al., 2013	Shortened Diabetes Numeracy Test for Adolescents with Type 1 DM (DNT-14)	Shortened Measure of numeracy skills used in type 1 diabetes (14-items)	Scores reported as% correct Possible range: 0–10 0% ↑scores = ↑DN skills	~10	Internal Consistency: KR-20 = .80	DNT-14: Convergent: DNT-39, *r* = .87 WRAT-3R[€] *r* = .36 Concurrent: Parent Education, *r* = .31 HgA1C, *r* = -.13 (N/S)	[US] (*n* = 133) 88% White 9% Black 2% Hispanic 1% Asian	Numeracy

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Ishikawa et al., 2008	Functional, Communicative and Critical Health Literacy Scale for Diabetes	Combination of functional, communicative, and critical HL in the context of diabetes (14-items)	Each item rated on a 4-point Likert scale, with total scores summed and then divided by 14 Possible range: 1–4, ↑scores = ↑HL	N/A	Internal Consistency: α = .78	Convergent: DKT , *r* = .37 Concurrent: Number of information sources, *r* = .40 Self-efficacy, *r* = .3	[Japan] (*n* = 138) Japanese	Functional Communicative Critical
Leung et al., 2012	Chinese Health Literacy Scale for Diabetes	Chinese (Cantonese) measure of HL of patients with diabetes in terms of decision making on 4 cognitive levels using Bloom's taxonomy (34-item)	Unclear	~7	Internal Consistency: α = .88 Test-retest: *r* = .90	Convergent: Chinese Literacy Scale, *r* = .82 Concurrent: DM knowledge, *r* = .40 DM Self-efficacy, *r* = 26 Chinese Learning, *r* = .303	[Hong Kong] (*n* = 137) 100% Chinese	Remembering Understanding Applying Analyzing
Kumar et al., 2010	Parental Health Literacy Activities Test (PHLAT)	HL measure for parenting, covers 3 clinical domains: nutrition/growth/ development, injury/safety, and medical/ preventative care (20-items)	Scores reported as% correct Possible range: 0–10 0% ↑scores = ↑Parental HL	~21	Internal Consistency: KR-20 = .76	PHLAT: Convergent: S-TOFHLA¢ *r* = .38 WRAT-3 , *r* = .55 Concurrent: Education, *r* = .29	[US] (*n* = 182) 51% Black 37% White 11 % Hispanic 2% 'Other'	Print Numeracy
Kumar et al., 2010	Shortened Parental Health Literacy Activities Test (PHLAT-10)	Shortened HL measure for parenting, includes items r/ t nutrition, growth charts and meds (10-items)	Scores reported as% correct Possible range:0–10 0% ↑scores = ↑Parental HL	N/A	Internal Consistency: KR-20 = .70	PHLAT-10: Convergent: PHLAT, *r* = .91 S-TOFHLA¢ *r* = .36 WRAT-3€ *r* = .53 Concurrent: Education, *r* = .25	[US] (*n* = 182) 51% Black 37% White 11 % Hispanic 2% 'Other'	Print Numeracy
Yin et al, 2012	Spanish Parental Health Literacy Activities Test (PHLAT-10 Spanish)	Spanish version of the PHLAT-10	Possible range: 0–10 0% ↑scores = ↑Parental HL	N/A	Internal Consistency: KR-20 = .61	Convergent: S-TOFHLA¢ *r* = .53 WRAT-3€*r* = .55 Concurrent: Education,	[US] (*n* = 176) 99% Hispanic 1% "Other"	Print Numeracy
Yin et al., 2012	PHLAT-8 Spanish	Shortened version of PHLAT-8 Spanish	Possible range: 0–10 0% ↑scores = ↑Parental HL	N/A	Internal Consistency: KR-20 = .64	Convergent: PHLAT-10 Spanish, *r* = .97	[US] (*n* = 176) 99% Hispanic 1% "Other"	Print Numeracy

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Wu et al., 2010	Health Literacy Measure for High School Students	HL measure that assesses how well high school students 'understand' and 'evaluate' health information (47-items)	Range: 0–107 30 'understand' items, 2pts/each 17 'evaluate' items, ranging from 1 to 4 pts/ each	N/A	Internal Consistency: α = .92	Convergent: Age, *r* = -.17 Male gender, *r* = -.18 Age came to Canada, *r* = -.22 Non-English speaker, *r* = -.15 Mother's edu, *r* = .19 Father's edu, *r* = .22 GPA, *r* = .48 Time reading/ study, *r* = .40 'fair'	[Canada] (*n* = 275) 100% Canadian teenagers 69% spoke language other than English at home	Prose Document Numeracy
Rivera-Vasquez et al., 2009	Spanish Cervical and Breast Cancer Literacy Assessment Tool (Sp-CB-LAT)	Spanish version of Breast and cervical Cancer focused HL measure, administered orally to participants (28-items)	Range: 0–28 ↑scores = ↑HBP-HL	~10	Internal Consistency: α = .99	N/A	[US] (*n* =19) 100% Latina	Functional
Rivera-Vasquez et al., 2009	Arabic Cervical and Breast Cancer Literacy Assessment Tool (Ar-CB-LAT)	Arabic version of Breast and cervical Cancer focused HL measure, administered orally to participants (28-items)	Range: 0–28 ↑scores = ↑HBP-HL	~10	Internal Consistency: α = .81	N/A	[US] (*n* = 56) 100% Arab	Functional
Williams et al., 2013	Breast Cancer Literacy Assessment Tool (B-CLAT)	Breast cancer focused HL measure, administered orally to participants (21-items)	Range: 0–30 ↑scores = ↑HBP-HL	N/A	Internal Consistency: α = .73 $\alpha = .81_{Black}$ $\alpha = .61_{Latina}$ $\alpha = .68_{Arab}$	CFA[£] c^2, p<.01 TLI[7] = .91 RMSEA[1] = .04 Concurrent: Education, *p* = .02	[US] (*n* = 543) 39% Black, 29% Latina 32% Arab	Functional
Williams and Templin, 2013	Cervical Cancer Literacy Assessment Tool (C-CLAT)	Cervical Cancer focused HL measure (16-items)	Range: 0–16 ↑scores = ↑HBP-HL	~5	Internal Consistency: $\alpha = .72_{TOT}$ $\alpha = .73_{Black}$ $\alpha = .76_{Latina}$ $\alpha = .60_{Arab}$	CFA[£] $X^2 p < .001$ CFI[5]= .76 RMSEA[1]= .07 Concurrent: Education, *p* = .02	[US] (*n* = 543) 39% Black, 29% Latina 32% Arab	Functional
Diviani et al., 2012	Cancer Literacy Scale	Cancer Literacy scale assesses cancer (37-items)	Scores reported as% correct Possible range:0–10 0% ↑scores = ↑Cancer HL	N/A	Internal Consistency: α = .77 Test re-test: *r* = .73	Concurrent: Attitude towards cancer screening, *r* = .26	[Switzerland] (*n* = 639) 72% Swiss 28% Other	Print

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Pendlimari et al, 2011	Assessment of Colon Cancer Literacy (ACCL)	ACCL assessed cancer-related terms (10-items)	Possible range: 0–10 0–6: limited Cancer HL 7–10: adequate Cancer HL	N/A	N/A	ROC[¥] analysis : vs. NVS Cutoff=2 91% sensitivity 34% specificity 46% PPV 87% NPV	[US] (*n* =61) unspecified	Print
Mazor et al., 2012	Cancer Message Literacy Test-Listening (CMLT-L)	Measure that assesses comprehension of spoken health messages related to cancer prevention, administered via touchscreen laptop (48-items)	Unclear	~60	Internal Consistency: α = .82	N/A	[US] (*n* = 79) 50% White 20% Black 16% Asian 14% Other	Comprehension
Mazor et al., 2012	Cancer Message Literacy Test-Reading (CMLT-R)	Measure that assesses comprehension of written health messages related to cancer prevention, printed on standard paper in 14 and 16 pt font (23-items)	Unclear	~10	Internal Consistency: α = .55	N/A	[US] (*n* = 79) 50% White 20% Black 16% Asian 14% Other	Print
Osborn et al., 2010	Brief Estimate of Health Knowledge and Action in HIV (BEHKA-HIV)	Measure of patient's HIV treatment knowledge and action	Possible range: 0–8 ↑scores = ↑HL	N/A	Internal Consistency: $\alpha = .73_{Knowledge}$ $\alpha = .79_{Action}$	Predictive: Med Adherence, p<.05 REALM[β] p<.05	[US] (*n* = 204) unspecified	Knowledge Action
Kim et al., 2011	High Blood Pressure Health Literacy Scale (HBP-HLS)	High blood pressure focused HL assesses print literacy(30 items) and functional HL (13 items)	Print Literacy Possible range: 0–30 Functional HL Range: 0–13 ↑scores = ↑HBP-HL	~10–15	Internal Consistency: Whole scale, KR-20 = .98 Print Literacy, KR-20 = .98 Functional HL, α = .93	Content: Print Literacy, CVI[a] > .80 Convergent: Modified TOFHLA[¢] *r* = .80 Modified NVS[÷] *r* = .76 FHL[¢] subscale, *r* = .82 Concurrent: Education, *r* = .67 Yrs in US, *r* = .32 HBP knowledge, *r* = .33	[US] (*n* = 440) Korean American	Print Functional Numeracy
Levin-Zamair et al., 2011	Media Health Literacy (MHL)	Measures media HL after viewing six television segments, then assesses content identification, perceived influence on	Possible range: 0–24 ↑ scores = ↑ MHL	N/A	Internal Consistency: α = .74	N/A	[Israel] (*n* = 1316) Israeli adolescents	Print

Patient Educ Couns. Author manuscript; available in PMC 2017 January 29.

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
		behavior, critical analysis and intended action/ reaction (24 -items)						
Reynolds et al., 2012	Food Label Literacy for Applied Nutrition Knowledge (FLLANK)	Measures nutrition knowledge/literacy based on food label content in school children (10-items)	Possible range: 0–10 0% ↑scores = ↑Nut. Literacy	~ 15	Internal Consistency: α = .77 Test re-test: ICC = .68	Predictive: ONQI$^{\lambda}$ t_{18} = 2.80, p = .01	[US] (n = 451) 100% Grade school kids Unspecified race/ethnic	Functional
Brega et al., 2012	Special Diabetes Program for Indians Health Heart Health Literacy (SDPI-HH-HL)	7-Item HL measure adopted from existing scales targeted at American Indians and Alaskan Natives with diabetes	Possible range: 0–10 0%	N/A	Internal Consistency: α = .67$_{\text{Print}}$	Predictive: Educational attainment, p <. 05 Knowledge, p <. 05	[US] (n = 3,033) American Indians and Alaska Natives	Print Numeracy
Gibbs et al., 2013	Nutrition Literacy Assessment Instrument (NLAI)	On-line survey that assesses 5 domains of nutritional literacy (35-item)	Possible range: 0–35 Based on the # of correct item, participants were categorizes as: “inadequate NL”, “marginal NL”, or “adequate NL” Cut-off points unclear	N/A	N/A	Content: Section is important: 89.7% (avg) Section is purposeful: 81.5% (avg) Divergent: REALM$^{\beta}$ r = .38 (N/S)	[US] (n = 134) 100% Dietitians Unspecified race/ethnic	Relationship b/t nutrition and health Knowledge of macronutrients Skill r/t food measurement Numeracy skill Food grouping skill
Guttersrud et al., 2013	Critical Nutrition Literacy (CNL)	A measure of critical nutrition literacy that measures 2 main domains (19-items)	Each item measured on a 5 pt Likert Scale Unclear how to interpret scores		Internal Consistency: α = .80$_{\text{Engage}}$ α = .70$_{\text{Claims}}$ IRT b = − 1.66–1.28	N/A	[Norway] (n = 473) 100% Norwegian nursing students	“Engagement” in dietary habits Taking critical stance toward nutritional “claims”
Wong et al., 2013	Hong Kong Oral Health Literacy Assessment Task for Pediatric Dentistry (HKOHLAT-P)	Measure that focuses on pediatric dentistry which is contextually relevant for those living in Hong Kong (Cantonese speakers) (52-items)	Possible range: 0–52 ↑ scores = Oral HL	N/A	Test-retest ICC = .63 Internal Consistency: α = .71	Concurrent: Self-Reading, r = .43 Reading to child, r = .12 (N/S) Convergent: TOFHLiD$^{¢}$ r = .39 HKREALD-30$^{\times}$ r = .36 Predictive: ECOHIS$^{\lambda}$ r = -.24	[Hong Kong] (n = 200) 100% Cantonese speaking children and parents	Knowledge Numeracy Comprehension

Author, Year	Instrument name	Description (items)	Scoring	Time, mins (adimn)	Reliability	Validity	[Country] Sample	Domain(s) of HL measured
Ownby et al., 2013	HIV related Health Literacy Scale (HIV-HL)	Brief computer administered HL measure that focuses on medication adherence and clinical encounters related to HIV (20-items)	Possible range: 0–20 Low literacy: < 15	~ 10–25	Internal Consistency: α = .69	Concurrent: Self-efficacy, *r* = .31 Depression, *r* = .19 Convergent: TOFHLA[¢] *r* = .58 ROC[¥] analysis: vs. TOFHLA[¢] Cutoff= 15 AUROC[¥1] = .77	[US] (*n* = 124) 63% Black 36% While 1% Other	Reading Listening
Leung et al., 2013	Chinese Health Literacy Scale for Chronic Care (CHLCC)	Chines HL measure of chronic care across 4 subscales (24-items)	Possible range: 0–48 ↑ scores = ↑ HL	N/A	IRT: *b* = −1.73–0.65 *a* = 1.09–5.68 Internal Consistency: α = .91 Test-retest ICC = .77	ROC[¥] analysis: vs. re-hospitalization and edu Cutoff = no re-hospitalization within 12 months and ≥ 7th grade edu AUROC[¥1] = .77% 84% sensitivity 66% specificity Discriminant: Age, *r* = −.31 PPCLA[≠] *r* = .80	[Hong Kong] (*n* = 262) 100% Chinese	Remembering Understanding Applying Analyzing
Massy et al., 2013	Multi-Dimensional Measure of Adolescent Health Literacy	Multi-dimensional measure of skills needed by adolescents to manage their health environment	All items measured on a 5-pt Likert Scale Unclear about score interpretation	N/A	Internal Consistency: $\alpha = .83_{TOT}$ $\alpha = .82_{1}$ $\alpha = .80_{2}$ $\alpha = .83_{3}$ $\alpha = .64_{4}$ $\alpha = .83_{5}$ $\alpha = .71_{6}$	N/A	[US] (*n* = 1208) 22% White 13% Black 34% Hispanic 8% Asian 22% Other	[1]Pt-provider encounter [2]Interaction with healthcare system [3]Rights/Resp. [4]Conf. in health info from personal source [5]Conf. in health info from media [6]Competency in seeking information online

[*] Please see Appendix D for acronym key.