BMJ Open

The quality of health literacy instruments used in children and adolescents: which one is the best?

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020080
Article Type:	Research
Date Submitted by the Author:	13-Oct-2017
Complete List of Authors:	Guo, Shuaijun; University of Melbourne School of Population and Global Health, Armstrong, Rebecca; University of Melbourne, Waters, E; University of Melbourne, School of Population and Global Health Sathish, Thirunavukkarasu; University of Melbourne School of Population and Global Health, Centre for Health Equity Alif, Sheikh; University of Melbourne, School of Population and Global Health Browne, Geoff; University of Melbourne, School of Population and Global Health Yu, Xiaoming; Peking University, Institute of Child and Adolescent Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Global health
Keywords:	Measurement properties, Health literacy, Children, Adolescents, Systematic review

The quality of health literacy instruments used in children and adolescents: which one is the best?

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1], Sheikh M Alif[1], Geoffrey R Browne[1], Xiaoming Yu[2*]

[1] School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[2] Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China

**** Corresponding Author***

Shuaijun Guo Postal address: 526-9, Level 5, 207 Bouverie Street, Carlton, Victoria, Australia 3010 Email: gshj1986@gmail.com Tel: +61 452 110 331

Xiaoming Yu Postal address: Room 209, Institute of Child and Adolescent Health, School of Public Health, 38 Xueyuan Road, Haidian, Beijing, China 100083 Email: yxm@bjmu.edu.cn Tel: +86 10 8280 2631

Word count: 4534

ABSTRACT

Objective: Improving health literacy at an early age is crucial to childhood and adolescent health and development. Although health literacy in children and adolescents has gained increasing attention in the past decade, it remains an under-researched area, particularly health literacy measurement. Given that it is still unclear which health literacy instrument is the best in terms of its validity, reliability and feasibility for children and adolescents, this study aimed to examine the quality of health literacy instruments used in children and adolescents and to identify the best instrument for field use.

Design: Systematic review.

Setting: A wide range of settings including schools, hospitals and communities.

Participants: Children or adolescents aged 6 to 24.

Primary and secondary outcome measures: Measurement properties (i.e. reliability, validity and responsiveness) and other important characteristics (e.g. health topics, components and scoring systems) of health literacy instruments.

Results: There were 15 health literacy instruments identified from the screening process. When measuring health literacy in children and adolescents, researchers mainly focus on the functional domain and participant characteristics of cognitive development, dependency and demographic patterns. The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. More than half (70.8%) of measurement properties were unknown, due to either poor methodological quality of included studies or a lack of reporting or assessment.

Conclusions: More rigorous and high-quality studies are needed to fill the knowledge gap in measurement properties of child and adolescent health literacy instruments. Although it is challenging to draw a robust conclusion about which instrument is the most reliable and the most valid, this review provides important evidence that

supports the use of the 8-item Health Literacy Assessment Tool (HLAT-8) to measure childhood and adolescent health literacy in future school-based research.

Keywords: Measurement properties; health literacy; children; adolescents; systematic review

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The COSMIN checklist was used as a methodological framework to rate the methodological quality of included studies.
- This review has updated previous reviews of childhood and adolescent health literacy measurement and identified eight additional health literacy instruments.
- Including only studies that aimed to develop or validate a health literacy instrument may eliminate studies that used a health literacy instrument for other purposes.
- Individual subjectivity exists in the screening and data synthesis stages.

INTRODUCTION

Health literacy is a personal resource that enables an individual to make decisions for healthcare, disease prevention and health promotion in everyday life.[1] As defined by the World Health Organisation,[2] health literacy refers to '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*.' The literature has shown that health literacy is an independent and more direct predictor of health outcomes than socio-demographics.[3 4] People with low health literacy are likely to have worse health-compromising behaviours, higher healthcare costs and poorer health status.[5] Given the close relationship between health literacy and health outcomes, many countries have adopted the promotion of health literacy as a key strategy to reduce health inequities.[6]

From a health promotion perspective, improving health literacy at an early age is crucial to childhood and adolescent health and development.[7] As demonstrated by Diamond *et al.*[8] and Robinson *et al.*,[9] health literacy interventions for children and adolescents can bring about improvements in healthy behaviours and decreased used of emergency department services. Although health literacy in young people has gained increasing attention, with a rapidly growing number of publications in the past decade,[10-12] childhood and adolescent health literacy is still under-researched. According to Forrest *et al.*'s 4D model,[13 14] health literacy in children is mediated by four additional factors compared to adults: (1) *developmental* change: children have less well-developed cognitive ability than adults; (2) *dependency*: children depend more on their parents and peers than adults do; (3) *differential* epidemiology: children experience a unique pattern of health, illness and disability; and (4) *demographic* patterns: many children and adolescents living in poverty or in single-parent families are neglected and so require additional care. These four differences pose significant challenges for researchers when measuring health literacy in children and adolescents.

Health literacy is a broad and multi-dimensional concept with varying definitions.[15] This paper uses the definition by Nutbeam who states that health literacy consists of three domains: functional, interactive and critical[16]. The *functional* domain refers to basic skills in reading and writing health information, which are important for

functioning effectively in everyday life. The *interactive* domain represents advanced skills that allow individuals to extract health information and derive meaning from different forms of communication. And the *critical* domain represents more advanced skills that can be used to critically evaluate health information and take control over health determinants.[16] Although health literacy is sufficiently explained in terms of its definitions[16-18] and theoretical models,[4 7] its measurement remains a contested issue. There are two possible reasons for this. One reason is the large variety of health literacy definitions and conceptual models,[12 15] and the other reason is that researchers may have different study aims, populations and contexts when measuring health literacy.[19 20]

Currently, there are two systematic reviews describing and analysing the methodology and measurement of childhood and adolescent health literacy.[10 11] In 2013, Ormshaw *et al.*[10] conducted a systematic review of child and adolescent health literacy measures. This review used four questions to explore health literacy measurement in children and adolescents: "*What measurement tools were used? What health topics were involved? What components were identified?* and *Did studies achieve their stated aims?*" The authors identified 16 empirical studies, with only six of them evaluating health literacy measurement as their primary aim. The remaining studies used health literacy measures as either a comparison tool when developing other new instruments or as a dependent variable to examine the effect of an intervention program. Subsequently, in 2014, Perry[11] conducted an integrative review of health literacy instruments used in adolescents. In accordance with the eligibility criteria, five instruments were identified.

Although these two reviews provide general knowledge about the methodology and measurement of health literacy in young people, both have limitations. Ormshaw *et al.*[10] did not evaluate measurement properties of each health literacy instrument. Although Perry[11] summarised measurement properties of each instrument, the information provided was limited and mostly descriptive, and lacked a critical appraisal. Notably, Ormshaw *et al.*[10] and Perry[11] did not consider the methodological quality of the included studies. A lack of quality assessment of studies raises concerns about the utility of such reviews for evaluating and selecting health literacy instruments for children and adolescents. Therefore, it is still unclear which

instrument is the best in terms of its validity, reliability and feasibility for field use. In addition, it is also unclear how Nutbeam's three-domain health literacy model and Forrest *et al.*'s 4D model are considered in existing health literacy instruments for children and adolescents.

To fill these knowledge gaps, this study aimed to conduct a systematic review to examine the quality of health literacy instruments used in the young population and to identify the best instrument for field use. We expected the findings would assist researchers in identifying and selecting the most appropriate instrument for different purposes when measuring childhood and adolescent health literacy.

METHODS

Following the methods for conducting systematic reviews outlined in the Cochrane Handbook,[21] we developed a review protocol (See **Appendix 1**) prior to commencing the study. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement[22] (See **Research Checklist**) was used to ensure the reporting quality of this review.

Literature search

The term '*health literacy*' was first used in 1974,[23] and so seven electronic databases were used to search for articles published between 1 January 1974 and 30 May 2014: Medline, PubMed, Embase, PsycINFO, CINAHL, ERIC and the Cochrane Library. The search strategy was designed on the basis of previous reviews[5 10 24 25] and in consultation with two librarian experts. Three types of search terms were used: (1) construct-related terms: '*health literacy*' OR '*health and education and literacy*'; (2) outcome-related terms: '*health literacy assess**' OR '*health literacy measure**' OR '*health literacy evaluat**' OR '*health literacy instrument**' OR '*health literacy tool**'; and (3) age-related terms: '*child**' OR '*adolescent**' OR '*student**' OR '*youth*' OR '*young people*' OR '*teen**' OR '*young adult.*'

No language restriction was applied. The detailed search strategy for each database is available in **Appendix 2**. As per the PRISMA flow diagram,[22] the references from

included studies and from six previously published systematic reviews on health literacy[5 10 24-27] were also included.

Eligibility criteria

Studies had to fulfil the following criteria to be included: (1) the stated aim of the study was to develop or validate a health literacy instrument; (2) participants were children or adolescents aged 6 to 24; (3) the term '*health literacy*' was explicitly defined, although studies assessing health numeracy (the ability to understand and use numbers in healthcare settings) were also considered; and (4) at least one measurement property (reliability, validity and responsiveness) was reported in the outcomes.

Studies were excluded if: (1) the full paper was not available (e.g. conference abstracts); (2) they were not peer-reviewed (e.g. dissertations, government reports); or (3) they were qualitative studies.

Selection process

All references were imported into EndNote X7 software (Thomson Reuters, New York, NY). Duplicate records were initially removed before screening. One author (GS) screened all studies based on title and abstract. Full-text papers of the remaining titles and abstracts were then obtained and screened by two independent authors (GS and SA). At each major step of this systematic review, discrepancies between authors were resolved through discussion.

Data extraction

Data were extracted from full-text papers by two independent authors (GS and TS). The extracted data included: characteristics of included studies (e.g. first author, published year and country), general characteristics of included instruments used in the included studies (e.g. health topics, components and scoring systems), methodological quality of included studies (e.g. internal consistency, reliability and

measurement error) and ratings of measurement properties of included instruments (e.g. internal consistency, reliability and measurement error).

Methodological quality assessment of included studies

The methodological quality of included studies was assessed using the COSMIN checklist.[28] The COSMIN checklist is a critical appraisal tool containing standards for evaluating the methodological quality of studies on measurement properties of health measurement instruments.[29] Specifically, nine measurement properties (internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity and responsiveness) were assessed.[29] Since there is no agreed-upon '*gold standard*' for health literacy measurement,[30 31] criterion validity was not assessed in this review. Each measurement property section contains 5 to 18 evaluating items. For example, '*internal consistency*' is evaluated against 11 items. Each item is scored using a four-point scoring system ('*excellent*', '*good*', '*fair*' or '*poor*'). The overall methodological quality of a study is obtained for each measurement property separately, by taking the lowest rating of any item in that section (i.e. '*worst score counts*'). Two authors (GS and TS) independently assessed the methodological quality of included studies.

Evaluation of measurement properties for included instruments

The quality of each measurement property of an instrument was evaluated using quality criteria proposed by Terwee *et al.*[32], who are members of the group that developed the COSMIN checklist (See **Appendix 3**). Each measurement property was given a rating result ('+' positive, '-' negative, '*?*' indeterminate and '*na*' no information available).

Best evidence synthesis: levels of evidence

As recommended by the COSMIN checklist developer group,[29] '*a best evidence synthesis*' was used to synthesise all the evidence on measurement properties of different instruments. The procedure used was similar to the Grading of

Recommendations, Assessment, Development and Evaluation (GRADE) framework[33], a transparent approach to rating quality of evidence that is often used in reviews of clinical trials.[34] Given that this review did not target clinical trials, the GRADE framework adapted by the COSMIN group was used.[35] Under this procedure, the possible overall rating for a measurement property is '*positive*', '*negative*', '*conflicting*' or '*unknown*', accompanied by levels of evidence ('*strong*', '*moderate*' or '*limited*') (See **Appendix 4**). Specifically, three steps were taken to obtain the overall rating for a measurement property. First, the methodological quality of a study on each measurement property was assessed using the COSMIN checklist.[28] Measurement properties from '*poor*' methodological quality studies did not contribute to '*the best evidence synthesis*'. Second, the quality of each measurement property of an instrument was evaluated using Terwee's quality criteria.[32] Third, the rating results of measurement properties in different studies on the same instrument were examined whether consistent or not. This best evidence synthesis was performed by one author (GS) and then checked by a second author (TS).

Results

The search identified 1804 studies. After duplicates and initial title/abstract screening, 303 full-text articles were identified and obtained. As per the eligibility criteria, 15 studies were included,[36-50] yielding 15 unique health literacy instruments used in children and adolescents (See **Figure 1**).

Characteristics of included studies

Among the 15 studies identified, 11 were published in the last five years (2010 to 2014) (See **Table 1**). Most included studies were conducted in Western countries (n=13), with seven studies carried out in the USA. The target population aged 7 to 25 could be roughly classified into three subgroups: children aged 7 to 12 (n=3), adolescents aged 13 to 17 (n=10) and young adults aged 18 to 25 (n=2). Schools (n=9) were the most common recruitment settings, compared to clinical settings (n=4) and communities (n=2).

Table 1. Characteristics of included studies

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
1	Davis *et al.*[38] (2006)	USA	Adolescents aged 10-19 years (mean age=14.8±1.9)	REALM-Teen	1533 (47.4)	na	Middle schools, high schools, paediatric primary care clinic and summer programs
2	Norman and Skinner[40] (2006)	Canada	Adolescents aged 13-21 years (mean age=14.95±1.24)	eHEALS	664 (55.7)	Sampling from one arm of a randomised controlled trial	Secondary schools
3	Chisolm and Buchanan[45] (2007)	USA	Young people aged 13-17 years (mean age=14.7)	TOFHLA	50 (48.0)	na	Children's hospital
4	Steckelberg *et al.*[44] (2009)	Germany	Students in Grade 10-11 and university	CHC Test	Sample 1: 322 (36.6) Sample 2: 107 (32.7)	na	Secondary schools, university
5	Schmidt *et al.*[43] (2010)	Germany	Children aged 9-13 years (mean age=10.4)	HKACSS	852 (52.9)	na	Primary school
6	Wu *et al.*[37] (2010)	Canada	Students in Grade 8-12	HLAB	275 (48.0)	Convenience sampling	Secondary schools
7	Levin-Zamir *et al.*[46] (2011)	Israel	Adolescents in Grade 7, 9, 11 (approximately age 13, 15 and 17)	MHL	1316 (52.0)	Probability sampling and random cluster sampling	Public schools
8	Chang *et al.*[48] (2012)	Taiwan	Students in high school (mean age=16.01±1.02)	c-sTOFHLAd	300 (52.6)	Multiple-stage stratified random sampling	High schools
9	Hoffman *et al.*[47] (2013)	USA	Youth aged 14-19 years (mean age=17)	REALM-Teen; NVS; s-TOFHLA	229 (61.6)	na	Private high school
10	Massey *et al.*[41] (2013)	USA	Adolescents aged 13-17 years (mean age=14.8)	MMAHL	1208 (37.6)	Sampling from a large health insurance network	Publicly health insurance network
11	Mulvaney *et al.*[50] (2013)	USA	Adolescents aged 12-17 years (Sample 1: mean age=13.92; Sample 2: mean age=15.10)	DNT-39 and DNT-14	Sample 1: 61 (52.5) Sample 2: 72 (55.6)	na	Diabetes clinics
12	Abel *et al.*[42]	Switzerland	Young adults aged 18-25 years	HLAT-8	7428 (95.5)	Sampling from compulsory	Compulsory military

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
	(2014)		(male mean age: 19.6; female mean age=18.8)			military service for males and two-stage random sampling for females	service, communities
13	Driessnack *et al.* (2014)	USA	Children aged 7-12 years	NVS	47 (53.0)	Convenience sampling	The science centre
14	Harper[39] (2014)	New Zealand	Students aged 18-24 years	HLAT-51	144 (41.0)	Purposeful sampling	College
15	Warsh *et al.*[36] (2014)	USA	Children aged 7-17 years (median age=11)	NVS	97 (46.0)	Convenience sampling	Paediatric clinics

Note: na, no information available. CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TOFHLA, the Test of Functional Health Literacy in Adults.

General characteristics of included instruments

Compared to previous systematic reviews,[10 11] this review identified eight additional health literacy instruments (NVS, s-TOFHLA, MMAHL, DNT-39, DNT-14, eHEALS, HLAT-51 and HLAT-8). The 15 health literacy instruments were classified into three groups based on whether the instrument was developed bespoke for the study or not (See **Table 2**).[10] The three groups were: (1) newly-developed instruments for childhood and adolescent health literacy (n=9);[37-44 46 47] (2) adapted instruments that were based on previous instruments for adult health literacy (n=3);[48 50] and (3) original instruments that were developed for adult health literacy (n=3).[36 45 47 49]

Health literacy domains and components

Next, Nutbeam's three-domain health literacy model[16] was used to classify the 15 instruments according to which of the commonly-used components of health literacy were included. Results showed that seven instruments measured only functional health literacy[36 38 45 47-50] and one instrument measured only critical health literacy.[44] There was one instrument measuring functional and interactive health literacy[43] and one measuring functional and critical health literacy.[37] Five instruments measured health literacy by all three domains (functional, interactive and critical).[39-42 46]

Consideration of participants' characteristics

As per Forrest *et al.*'s 4D model,[13 14] the 15 included instruments were examined for whether participant characteristics were considered when developing a new instrument or validating an existing instrument. Results showed most of the health literacy instruments considered developmental change, dependency and demographic patterns. In contrast, only two instruments considered differential epidemiology.[50]

Health topics, contents and readability levels

Health literacy instruments for children and adolescents covered a range of health topics such as nutrition and sexual health. Most instruments (n=12) measured health literacy in healthcare settings or health promotion contexts, while only three

instruments measured health literacy in the specific context of eHealth or media health.[39 40 46] In relation to the readability of tested materials, only five health literacy instruments reported their readability levels, ranging from 4th to 19.5th grade.

Burden and forms of administration

The time to administer was reported in seven instruments, and ranged from 3 to 90 minutes. There were three forms of administration: interviewer-administered instruments (n=7), self-administered instruments (n=7) and video-assisted, interviewer-administered instruments (n=1). As for the method of assessment, ten instruments were performance-based, three instruments were self-report, and two included both performance-based and self-report items.

Table 2. General and important characteristics of included instruments used in children and adolescents

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
1	NVS[36 47 49]	***Functional HL*** 1. Reading comprehension (2) 2. Numeracy (4)	Demographic patterns	Nutrition-related information about the label of an ice cream container (na)	Open-ended	Score: 0-6; Ordinal category: 0-1: high likelihood of limited literacy; 2-3: possibility of limited literacy; 4-6: adequate literacy	No longer than 3 minutes	Interviewer-administered & Performance-based
2	TOFHLA[45]	***Functional HL*** 1. Reading comprehension (50) 2. Numeracy (17)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4.3 grade), a standard informed consent form (10.4 grade), patients' rights and responsibilities section of a Medicaid application form (19.5 grade), actual hospital forms & labelled prescription vials (9.4 grade)	4 response options	Score: 0-100; Ordinal category: 0-59: inadequate health literacy; 60-74: marginal health literacy; 75-100: adequate health literacy	12.9 minutes (8.9-17.3 minutes)	Interviewer-administered & Performance-based
3	s-TOFHLA[47]	***Functional HL*** 1. Reading comprehension (36)	Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	na	Interviewer-administered & Performance-based
4	c-sTOFHLAd[48]	***Functional HL*** 1. Reading comprehension (36)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	20-minute class period	Self-administered & Performance-based
5	REALM-Teen[38 47]	***Functional HL*** 1. Reading recognition (66)	Developmental change Demographic patterns	66 health-related words such as weight, prescription and tetanus (6th grade)	Open-ended	Score: 0-66; Ordinal category: 0-37: ≤ 3rd; 38-47: 4th-5th; 48-58:	2-3 minutes	Interviewer-administered & Performance-

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
						6th-7th; 59-62: 8th-9th; 63-66: ≥ 10th		based
6	HLAB[37]	***Functional and critical HL*** 1. Understanding health information (30) 2. Evaluating health information (17)	Developmental change Demographic patterns	A range of topics such as nutrition and sexual health (pilot-tested)	Open-ended	Score: 0-107; Continuous category	Two regular classroom sessions	Self-Administered & Performance-based
7	MMAHL[41]	***Functional, interactive and critical HL*** 1. Patient-provider encounter (4) 2. Interaction with the health care system (5) 3. Rights and responsibilities (7) 4. Confidence in using health information from personal source (3) 5. Confidence in using health information from media source (3) 6. Health information seeking competency using the Internet (2)	Developmental change Demographic patterns Dependency	Experiences of how to access, navigate and manage one's health care and preventive health needs (6th grade)	5-point Likert scale	Score: na; Continuous category	na	Self-administered & Self-reported
8	MHL[46]	***Functional, interactive and critical HL*** 1. Content identification (6) 2. Perceived influence on behaviour (6) 3. Critical analysis and intended (6) 4. Action/reaction (6)	Dependency Demographic patterns	Nutrition/dieting, physical activity, body image, sexual activity, cigarette smoking, alcohol consumption, violent behaviour, safety habits and/or friendship and family connectedness (pilot-tested)	Open-ended & multiple choice	Score: 0-24; Continuous category	na	Video-assisted interviewer-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
9	DNT-39[50]	***Functional health literacy*** 1. Health numeracy (39)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score: 0-100; Continuous category	na	Interviewer-administered & Performance-based
10	DNT-14[50]	***Functional health literacy*** 1. Health numeracy (14)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score: 0-100; Continuous category	na	Interviewer-administered & Performance-based
11	eHEALS[40]	***Functional, interactive and critical HL*** 1. Accessing health information (4) 2. Evaluating health information (2) 3. Applying health information (2)	Developmental change Dependency Demographic patterns	General health topics about online health information (pilot-tested)	5-point Likert scale	Score: na; Continuous category	na	Self-Administered & Self-reported
12	CHC Test[44]	***Critical HL*** 1. Understanding medical concepts (15) 2. Searching literature skills (22) 3. Basic statistics (18) 4. Design of experiments and sampling (17)	Developmental change Demographic patterns	Echinacea and common cold, magnetic resonance imaging in knee injuries, treatment of acne, breast cancer screening (pilot-tested)	Open-ended & multiple choice	na	Less than 90 minutes	Interviewer-administered & Performance-based
13	HKACSS[43]	***Functional and interactive HL*** 1. Health knowledge (3) 2. Health attitudes (4) 3. Health communication (3) 4. Self-efficacy (3)	Developmental change Dependency Demographic patterns	Physical activities, nutrition, smoking, vaccination, tooth health and general health (na)	2 response options; 5-point Likert scale; 4-point Likert scale	Score: na; Continuous category	na	Self-Administered & Performance-based & Self-reported

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
14	HLAT-51[39]	***Functional, interactive and critical HL*** 1. Comprehension skill (20) 2. Health numeracy (11) 3. Media literacy (8) 4. Digital literacy (12)	Developmental change Dependency Demographic patterns	Health topics such as gout and uric acid, high cholesterol and triglyceride levels, health-information-seeking skills (na)	Yes/no; multiple choice	na	30-45 minutes	Self-administered & Performance-based & Self-reported
15	HLAT-8[42]	***Functional, interactive and critical HL*** 1. Understanding health information (2) 2. Finding health information (2) 3. Interactive health literacy (2) 4. Critical health literacy (2)	Dependency Demographic patterns	General health topics in people's daily life (na)	5-point Likert scale; 4-point Likert scale	Score: 0-37; Continuous category	na	Self-administered & Self-reported

Note: na, no information available. CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HL, Health Literacy; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; IOM, the Institute of Medicine; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TOFHLA, the Test of Functional Health Literacy in Adults; WHO, the World Health Organization.

Evaluation of methodological quality of included studies

According to the COSMIN checklist,[28] the methodological quality of each instrument as assessed by each study is presented in **Table 3**. All studies (n=15) examined content validity, 12 studies assessed internal consistency and hypotheses testing, six studies examined structural validity, five studies assessed reliability, and only one study assessed cross-cultural validity.

Evaluation of instruments' measurement properties

After the methodological quality assessment of included studies, measurement properties of each health literacy instrument were examined according to Terwee's quality criteria (See **Appendix 5**).[32] The rating results of measurement properties of each instrument are summarised in **Table 4**.

The synthesised evidence for the overall rating of measurement properties

Finally, a synthesis was conducted for the overall rating of measurement properties for each instrument according to '*the best evidence synthesis*' guidelines recommended by the COSMIN checklist developer group.[29] This synthesis result was derived from information presented in **Table 3** and **Table 4**. The overall rating of each measurement property for each health literacy instrument is presented in **Table 5**. In summary, most information (70.8%, 85/120) on measurement properties was unknown due to either poor methodological quality of studies or a lack of information on reporting or assessment.

Table 3. Methodological quality of each study for each measurement property according to the COSMIN checklist

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsiveness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [47]	Poor	na	na	Poor	na	Fair	na	na
NVS (Driessnack *et al.*, 2014)	Poor	na	na	Poor	na	Poor	na	na
NVS (Warsh *et al.*, 2014) [36]	na	na	na	Poor	na	Fair	na	na
TOFHLA (Chisolm and Buchanan, 2007) [45]	na	na	na	Poor	na	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [47]	Poor	na	na	Poor	na	Fair	na	na
c-sTOFHLAd (Chang *et al.*, 2012) [48]	Fair	Fair	na	Good	Fair	Fair	Fair	na
REALM-Teen (Davis *et al.*, 2006) [38]	Poor	Fair	na	Good	na	Fair	na	na
REALM-Teen (Hoffman *et al.*, 2013) [47]	Poor	na	na	Poor	na	Poor	na	na
HLAB (Wu *et al.*, 2010) [37]	Fair	Poor	na	Good	na	Fair	na	na
MMAHL (Massey *et al.*, 2013) [41]	Good	na	na	Good	Good	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [46]	Poor	na	na	Good	na	Good	na	na
DNT-39 (Mulvaney *et al.*, 2013) [50]	Fair	na	na	Poor	na	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013) [50]	Fair	na	na	Poor	na	Fair	na	na
eHEALS (Norman and Skinner, 2006) [40]	Fair	Fair	na	Good	Fair	Fair	na	na
CHC Test (Steckelberg *et al.*, 2009) [44]	na	Poor	na	Good	Poor	na	na	na
HKACSS (Schmidt *et al.*, 2010) [43]	Excellent	na	na	Good	na	Good	na	na
HLAT-51 (Harper, 2014) [39]	Poor	na	na	Good	Poor	na	na	na
HLAT-8 (Abel *et al.*, 2014)	Excellent	na	na	Poor	Excellent	Good	na	na

Note: na, no information available. CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TOFHLA, the Test of Functional Health Literacy in Adults.

Table 4. Evaluation of measurement properties for included instruments according to Terwee's quality criteria

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [47]	-	na	na	?	na	-	na	na
NVS (Driessnack *et al.*, 2014) [49]	+	na	na	?	na	-	na	na
NVS (Warsh *et al.*, 2014) [36]	na	na	na	?	na	+	na	na
TOFHLA (Chisolm and Buchanan, 2007) [45]	na	na	na	?	na	+ (TOFHLA-R) -(TOFHLA-N)	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [47]	+	na	na	?	na	-	na	na
c-sTOFHLAd (Chang *et al.*, 2012)	+	+	na	+	?	+	?	na
REALM-Teen (Davis *et al.*, 2006) [38]	+	+	na	+	na	+	na	na
REALM-Teen (Hoffman *et al.*, 2013) [47]	+	na	na	?	na	-	na	na
HLAB (Wu *et al.*, 2010) [37]	+	+	na	+	na	-	na	na
MMAHL (Massey *et al.*, 2013) [41]	+	na	na	+	-	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [46]	+	na	na	+	na	+	na	na
DNT-39 (Mulvaney *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
DNT-14 (Mulvaney *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
eHEALS (Norman and Skinner, 2006) [40]	+	-	na	+	+	-	na	na
CHC Test (Steckelberg *et al.*, 2009) [44]	na	+	na	+	+	na	na	na
HKACSS (Schmidt *et al.*, 2010) [43]	+ (Health communication) - (Health attitude)	na	na	+	na	+	na	na
HLAT-51 (Harper, 2014)	?	na	na	+	?	na	na	na
HLAT-8 (Abel *et al.*, 2014) [42]	-	na	na	?	+	+	na	na

Note: na, no information available; + positive rating; ? indeterminate rating; - negative rating. CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TOFHLA, the Test of Functional Health Literacy in Adults; TOFHLA-N, the Numeracy part of the Test of Functional Health Literacy in Adults; TOFHLA-R, the Reading part of the Test of Functional Health Literacy in Adults.

Table 5. The overall quality of measurement properties for each health literacy instrument used in children and adolescents

Health literacy instrument	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS [36 47 49]	?	na	na	?	na	±	na	na
TOFHLA [45]	na	na	na	?	na	+ (TOFHLA-R) - (TOFHLA-N)	na	na
s-TOFHLA [47]	?	na	na	?	na	-	na	na
c-sTOFHLAd [48]	+	+	na	++	?	+	?	na
REALM-Teen [38 47]	?	+	na	++	na	+	na	na
HLAB [37]	+	?	na	++	na	-	na	na
MMAHL [41]	++	na	na	++	--	na	na	na
MHL [46]	?	na	na	++	na	++	na	na
DNT-39 [50]	+	na	na	?	na	-	na	na
DNT-14 [50]	+	na	na	?	na	-	na	na
eHEALS [40]	+	-	na	++	+	-	na	na
CHC Test [44]	na	?	na	++	?	na	na	na
HKACSS [43]	+++ (Health communication) --- (Health attitude)	na	na	++	na	++	na	na
HLAT-51 [39]	?	na	na	++	?	na	na	na
HLAT-8 [42]	---	na	na	?	+++	++	na	na

Note: na, no information available; +++ or --- strong evidence and positive/negative result; ++ or -- moderate evidence and positive/negative result; + or - limited evidence and positive/negative result; ± conflicting evidence; ? unknown, due to poor methodological quality or indeterminate rating of a measurement property. CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TOFHLA, the Test of Functional Health Literacy in Adults; TOFHLA-N, the Numeracy part of the Test of Functional Health Literacy in Adults; TOFHLA-R, the Reading part of the Test of Functional Health Literacy in Adults.

Discussion

Summary of main results

This study identified and examined 15 health literacy instruments used in children and adolescents and exemplified the large variety of methods to measure childhood and adolescent health literacy. It shows that to date, health literacy instruments generally focus on the functional domain, and less on the interactive and critical domains. When measuring health literacy in children and adolescents, researchers mainly focus on participant characteristics of developmental change, dependency and demographic patterns, rather than differential epidemiology. The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. Most information (70.8%) on measurement properties was unknown due to either the poor methodological quality of studies or a lack of reporting or assessment. It is therefore difficult to draw a robust conclusion about which instrument is the best.

Health literacy measurement in children and adolescents

This review found that health literacy measurement in children and adolescents still focused on the functional domain (n=7) rather than three domains (n=5). Unlike health literacy research for patients in clinics, health literacy research for children and adolescents (a comparatively healthy population) should be considered from a health promotion perspective,[51] rather than a health care or disease management perspective. Integrating interactive and critical domains into health literacy measurement is aligned with the rationale of emphasising empowerment in health promotion for children and adolescents.[52] The focus of health literacy for this population group should therefore include all three domains and so there is a need for future research to integrate the three domains within health literacy instruments.

Similar to previous findings by Ormshaw *et al.*,[10] this review also revealed that childhood and adolescent health literacy measurement varied by its dimensions, health topics, forms of administration, and by the level to which participant characteristics were considered. There

are likely four main reasons for these disparities. First, definitions of health literacy were inconsistent. Some researchers measured general health literacy,[37 42] while others measured eHealth literacy or media health literacy.[40 46] Second, researchers had different research purposes for their studies. Some researchers used what were originally adult instruments to measure adolescent health literacy,[36 45 49] whereas others developed new or adapted instruments.[37-39 50] Third, the research settings affected the measurement process. As clinical settings were busy, short surveys were more appropriate than long surveys.[36 38 41] On the other hand, health literacy in school settings was often measured by long and comprehensive surveys.[37 39 44] Fourth, researchers considered different participant characteristics when measuring health literacy in children and adolescents. For example, some researchers took considerations of students' cognitive development,[37 38 41 43 48] some focused on adolescents' resources and environments (e.g. friends and family contexts, eHealth contexts, media contexts),[40 42 46] and others looked at the effect of different cultural backgrounds and socio-economic status.[37 38 40 41 43 44 46-49] Based on Forrest *et al.*'s 4D model,[13 14] this review showed that most health literacy instruments considered participants' development, dependency and demographic patterns, with only two instruments considering differential epidemiology.[50] Although the '4D' model cannot be used to reduce the disparities in health literacy measurement, it does provide an opportunity to identify gaps in current research and assist researchers to consider participants' characteristics comprehensively in future research.

The methodological quality of included studies

This review included a methodological quality assessment of included studies, which was absent from previous reviews on this subject.[10 11] Methodological quality assessment is important because strong conclusions about the measurement properties of instruments can only be drawn from high-quality studies. In this review, the COSMIN checklist was shown to be a useful framework for critically appraising the methodological quality of studies via each measurement property. Findings suggested that there was wide variation in the methodological quality of studies for all instruments. Poor methodological quality of studies was often seen in the original or adapted health literacy instruments (the NVS, the TOFHLA, the s-TOFHLA, the DNT-39 and the DNT-14) for two main reasons. The first reason was the vague description of the target population involved. This suggested that researchers were less

likely to consider an instrument's content validity when using the original, adult instrument for children and adolescents. Given that children and adolescents have less well-developed cognitive abilities, it is essential to assess whether all items within an instrument are understood in future. The second reason was a lack of uni-dimensionality analysis for internal consistency. As explained in the COSMIN manual,[53] a set of items can be inter-related and multi-dimensional, uni-dimensionality is a prerequisite for a clear interpretation of the internal consistency statistics (Cronbach's alpha). Future research on the use of health literacy instruments therefore needs to assess and report both internal consistency statistics and uni-dimensionality analysis (e.g. factor analysis).

Critical appraisal of measurement properties for included instruments

This review demonstrated that of all instruments reviewed only the c-sTOFHLAd showed satisfactory internal consistency and test-retest reliability. The c-sTOFHLAd was a translated tool of the s-TOFHLA from English to Chinese.[48] Compared to the overall reliability rating of the s-TOFHLA,[47] the c-sTOFHLAd showed better results. The reason for this was probably the different methodological quality of included studies between the s-TOFHLA and the c-sTOFHLAd. The c-sTOFHLAd study had fair methodological quality in terms of internal consistency and test-retest reliability, whereas the original s-TOFHLA study had poor methodological quality for internal consistency and unknown information for test-retest reliability. Given the large disparity of rating results between the original and translated instrument, further evidence is needed to confirm whether the s-TOFHLA has the same or a different reliability within different cultures, thus assisting researchers to understand the generalisability of the s-TOFHLA's reliability results.

Six instruments were found to show satisfactory content validity and construct validity (i.e. structural validity, hypotheses testing and cross-cultural validity). Construct validity is a fundamental aspect of psychometrics and was examined for two reasons. First, it enables an instrument to be assessed for the extent to which operational variables adequately represent underlying theoretical constructs.[54] Second, the overall rating results of content validity for all instruments were similar. The only difference was that the target population was involved or not. Given that all instruments' items reflected the measured construct, in this review,

construct validity was determined to be key to examining the overall validity of included instruments. In this context, only the HLAT-8 showed positive evidence of structural validity and hypotheses testing. However, in the original paper,[42] the HLAT-8 was only tested for its known-group validity, not for convergent validity. Examination of convergent validity is important because it assists researchers in understanding the extent to which two examined measures' constructs are theoretically and practically related.[55] Therefore, future research on the convergent validity of the HLAT-8 would be beneficial for complementing that which exists for its construct validity.

As was the case in a previous study by Jordan *et al.*,[25] this review demonstrated that none of the 15 studies contained evidence of responsiveness. Responsiveness is the ability of an instrument to detect change over time in the construct being measured, and it is particularly important for longitudinal studies.[28] However, most studies included in this review were cross-sectional studies, and only one study (on the MMAHL[41]) discussed the potential to measure health literacy over time. Studies that measure health literacy over time in populations are needed, not only because this is a prerequisite for longitudinal studies, but also so that the responsiveness of instruments can be monitored and improved.

Feasibility issues for included instruments

This review showed that the feasibility aspects of instruments varied markedly. In relation to forms of administration, this review identified seven self-administered instruments and eight interviewer-administered instruments. This suggests that both methods of administration are well used. Self-administered instruments are cost-effective and efficient, but may bring about respondent bias, whereas interviewer-administered instruments, while able to ensure high response rates, are always resource intensive and expensive to administer.[56] Although the literature showed that there was no significant difference in scores outcome between these two administration modes,[57 58] the relevant studies mostly concerned health-related quality of life instruments. It is still unknown whether the same is true for health literacy instruments. Among children and adolescents, health literacy research is more likely to be conducted through large-scale surveys in school settings. Therefore, the more cost-effective, self-administered mode seems to have great potential for future research. To further support the

wide use of self-administered instruments, there is a need for future research to confirm the same effect of administration between self-administered and interviewer-administered instruments.

With regard to the type of assessment method, this review revealed that most health literacy instruments for children and adolescents are performance-based. There might be two reasons for this. First, it is due to participant characteristics. Compared with adults, children and adolescents have limited cognitive ability and are dependent on their parents for health decisions.[14] It is challenging for them to accurately self-assess their ability to find, understand, communicate and apply health information. Second, performance-based instruments are objective, whereas self-report instruments are subjective and may bring about over-estimated results.[59] However, the frequent use of performance-based instruments does not mean that they are more appropriate than self-report instruments when measuring childhood and adolescent health literacy. Compared with performance-based instruments, self-report instruments are always time-efficient and help to preserve respondents' dignity.[20] The challenge in using self-report instruments is to consider the readability of tested materials. If children and adolescents can understand what a health literacy instrument measures, then they can accurately self-assess their own health literacy skills.[52] The difference between self-report and performance-based instruments of health literacy has been discussed in the literature,[60] but the evidence is still limited due to a lack of specifically-designed studies for exploring the difference. Further studies are needed to fill this knowledge gap.

Recommendations for future research

This review identified ten instruments (the REALM-Teen, the NVS, the s-TOFHLA, the c-sTOFHLAd, the eHEALS, the CHC Test, the HKACSS, the HLAB, the MHL, the HLAT-51) that were used to measure health literacy in school settings. Although it is difficult to categorically state which instrument is the best, this review provides useful information that will assist researchers to identify the most suitable instrument to use when measuring health literacy in children and adolescents in school contexts.

Among the ten instruments, four tested functional health literacy (the REALM-Teen, the NVS, the s-TOFHLA and the c-sTOFHLAd); one examined critical health literacy (the CHC Test); one measured functional and interactive health literacy (the HKACSS); one examined functional and critical health literacy (the HLAB); and three tested health literacy comprehensively focusing on functional, interactive and critical domains (the eHEALS, the MHL and the HLAT-51); however, none of these comprehensive instruments were considered appropriate for use in schools. This was due to the fact that they focused on non-general health literacy or were burdensome to administer. To ensure a three-domain nature focus, only the MMAHL and the HLAT-8 were available for consideration in this review.

After comparing measurement contexts and measurement purpose, the HLAT-8 was identified as the most suitable instrument for measuring adolescent health literacy in school settings for four reasons: (1) it measures health literacy in the context of family and friends,[42] a highly important attribute because children and adolescents often need support for health decisions from parents and peers;[7 14] (2) it is a short but comprehensive tool that captures Nutbeam's three-domain nature of health literacy;[16] (3) it showed satisfactory structural validity (RMSEA=0.03; CFI=0.99; TLI=0.97; SRMR=0.03);[42] and (4) it has good feasibility (e.g. it is self-administered and time-efficient) for large-scale samples in school-based studies.

Limitations

This review was not without limitation. First, we restricted the search to studies aiming to develop or validate a health literacy instrument. Thus we may have missed relevant instruments in studies that were not aiming to develop instruments[61 62] or the recently-developed instruments.[63-65] Second, although the COSMIN checklist provided us with strong evidence of the methodological quality of a study via an assessment of each measurement property, it cannot evaluate a study's overall methodological quality. Third, individual subjectivity plays a part in the screening, data extraction and synthesis stage of the review. To reduce this subjectivity, two authors independently managed the major stages.

CONCLUSION

This review updated previous reviews of childhood and adolescent health literacy measurement and incorporated a quality assessment framework. It showed that most information on measurement properties was unknown due to either the poor methodological quality of studies or a lack of assessment and reporting. Rigorous and high-quality studies are needed to fill the knowledge gap in relation to health literacy measurement in children and adolescents. Although it is challenging to draw a robust conclusion about which instrument is the best, this review provides important evidence that supports the use of the HLAT-8 to measure childhood and adolescent health literacy in future school-based research.

ACKNOWLEDGMENTS

This paper is part of Shuaijun Guo's PhD research project, which is supported by the Melbourne International Engagement Award. The authors also appreciate the helpful comments received from the anonymous reviewers at *BMJ Open*.

CONTRIBUTORS

SG conceived the review approach. RA and EW provided general guidance for the drafting of the protocol. SG and SA screened the literature. SG and TS extracted data. SG drafted the manuscript. SG, GB, RA and XY reviewed and revised the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

None.

DATA SHARING STATEMENT

There are no additional data available.

REFERENCES

1. Nutbeam D. The evolving concept of health literacy. *Soc Sci Med* 2008;67(12):2072-78.
2. Nutbeam D. Health promotion glossary. *Health Promot Int* 1998;13(4):349-64.
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav* 2007;31(Supplement 1):S19-S26.
4. Squiers L, Peinado S, Berkman N, et al. The health literacy skills framework. *J Health Commun* 2012;17(sup3):30-54.
5. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155(2):97-107.
6. Dodson S, Good S, Osborne R. Health literacy toolkit for low-and middle-income countries: a series of information sheets to empower communities and strengthen health systems. New Delhi: World Health Organization, Regional Office for South-East Asia, 2015.
7. Manganello J. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res* 2008;23(5):840-47.
8. Diamond C, Saintonge S, August P, et al. The development of building wellness™, a youth health literacy program. *J Health Commun* 2011;16(sup3):103-18.
9. Robinson LD, Jr., Calmes DP, Bazargan M. The impact of literacy enhancement on asthma-related outcomes among underserved children. *J Natl Med Assoc* 2008;100(8):892-96.
10. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. *Health Educ* 2013;113(5):433-55.
11. Perry EL. Health literacy in adolescents: an integrative review. *J Spec Pediatr Nurs* 2014;19(3):210-18.
12. Bröder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017;17:361.
13. Forrest CB, Simpson L, Clancy C. Child health services research: challenges and opportunities. *JAMA* 1997;277(22):1787-93.

14. Rothman RL, Yin HS, Mulvaney S, et al. Health literacy and quality: focus on chronic illness care and patient safety. *Pediatrics* 2009;124(Supplement 3):S315-S26.
15. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12(1):80.
16. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15(3):259-67.
17. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? *J Health Commun* 2010;15(S2):9-19.
18. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies? *Int J Public Health* 2009;54(5):303-05.
19. Abel T. Measuring health literacy: moving towards a health-promotion perspective. *Int J Public Health* 2008;53(4):169-70.
20. Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. *J Health Commun* 2014;19(12):1481-96.
21. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. In: Higgins JP, Green S, eds. London, UK: The Cochrane Collaboration, 2011.
22. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151(4):264-69.
23. Simonds SK. Health education as social policy. *Health Educ Monogr* 1974;2(1_suppl):1-10.
24. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. *Nurs Health Sci* 2009;11(1):77-89.
25. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol* 2011;64(4):366-79.
26. Sanders LM, Federico S, Klass P, et al. Literacy and child health: a systematic review. *Arch Pediatr Adolesc Med* 2009;163(2):131-40.
27. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics* 2009;124(Supplement 3):S265-S74.

28. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):651-57.

29. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13.

30. McCormack L, Haun J, Sørensen K, et al. Recommendations for Advancing Health Literacy Measurement. *J Health Commun* 2013;18(sup1):9-14.

31. Pleasant A, McKinney J, Rikard RV. Health Literacy Measurement: A Proposed Research Agenda. *J Health Commun* 2011;16(sup3):11-21.

32. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.

33. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490-94.

34. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-26.

35. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):659-70.

36. Warsh J, Chari R, Badaczewski A, et al. Can the newest vital sign be used to assess health literacy in children and adolescents? *Clin Pediatr* 2014;53(2):141-44.

37. Wu A, Begoray D, Macdonald M, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. *Health Promot Int* 2010;25(4):444-52.

38. Davis TC, Wolf MS, Arnold CL, 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(6):e1707-e14.

39. Harper R. Development of a health literacy assessment for young adult college students: A pilot study. *J Am Coll Health* 2014;62(2):125-34.

40. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006;8(4):e27.
41. Massey P, Prelip M, Calimlim B, et al. Findings Toward a Multidimensional Measure of Adolescent Health Literacy. *Am J Health Behav* 2013;37(3):342-50.
42. Abel T, Hofmann K, Ackermann S, et al. Health literacy among young adults: a short survey tool for public health and health promotion research. *Health Promot Int* 2014;30(3):725-35.
43. Schmidt CO, Fahland RA, Franze M, et al. Health-related behaviour, knowledge, attitudes, communication and social status in school children in Eastern Germany. *Health Educ Res* 2010;25(4):542-51.
44. Steckelberg A, Hülfenhaus C, Kasper J, et al. How to measure critical health competences: development and validation of the Critical Health Competence Test (CHC Test). *Adv Health Sci Educ* 2009;14:11-22.
45. Chisolm DJ, Buchanan L. Measuring adolescent functional health literacy: a pilot validation of the test of functional health literacy in adults. *J Adolescent Health* 2007;41(3):312-14.
46. Levin-Zamir D, Lemish D, Gofin R. Media Health Literacy (MHL): development and measurement of the concept among adolescents. *Health Educ Res* 2011;26(2):323-35. doi: 10.1093/her/cyr007
47. Hoffman S, Trout A, Nelson T, et al. A psychometric assessment of health literacy measures among youth in a residential treatment setting. *J Stud Soc Sci* 2013;5(2):288-300.
48. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese version of short-form Test of Functional Health Literacy in Adolescents. *J Clin Nurs* 2012;21(17-18):2429-37. doi: 10.1111/j.1365-2702.2012.04147.x
49. Driessnack M, Chung S, Perkhounkova E, et al. Using the "newest vital sign" to assess health literacy in children. *J Pediatr Health Care* 2014;28(2):165-71.
50. Mulvaney SA, Lilley JS, Cavanaugh KL, et al. Validation of the diabetes numeracy test with adolescents with type 1 diabetes. *J Health Commun* 2013;18(7):795-804.
51. Kilgour L, Matthews N, Christian P, et al. Health literacy in schools: prioritising health and well-being issues through the curriculum. *Sport Educ Soc* 2015;20(4):485-500.

52. Velardo S, Drummond M. Emphasizing the child in child health literacy research. *J Child Health Care* 2017;21(1):5-13.

53. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist manual. Amsterdam: VU University Medical Centre, 2012.

54. Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008;98(1):9-10.

55. Akpa OM, Bamgboye EA, Baiyewu O. The Adolescents' Psychosocial Functioning Inventory (APFI): scale development and initial validation using Exploratory and Confirmatory Factor Analysis. *Afr J Psychol Study Soc Issues* 2015;18(1):1-21.

56. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health* 2005;27(3):281-91.

57. Lozano F, Lobos JM, March JR, et al. Self-administered versus interview-based questionnaires among patients with intermittent claudication: Do they give different results? A cross-sectional study. *Sao Paulo Med J* 2016;134(1):63-69.

58. Dujaili JA, Sulaiman SAS, Awaisu A, et al. Comparability of Interviewer-Administration Versus Self-Administration of the Functional Assessment of Chronic Illness Therapy-Tuberculosis (FACIT-TB) Health-Related Quality of Life Questionnaire in Pulmonary Tuberculosis Patients. *Pulm Ther* 2016;2(1):127-37.

59. Altin SV, Finke I, Kautz-Freimuth S, et al. The evolution of health literacy assessment tools: a systematic review. *BMC Public Health* 2014;14:1207.

60. Kiechle ES, Bailey SC, Hedlund LA, et al. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. *J Gen Intern Med* 2015;30(10):1538-46.

61. Paek H-J, Reber BH, Lariscy RW. Roles of interpersonal and media socialization agents in adolescent self-reported health literacy: a health socialization perspective. *Health Educ Res* 2011;26(1):131-49.

62. Hubbard B, Rainey J. Health Literacy Instruction and Evaluation among Secondary School Students. *Am J Health Educ* 2007;38(6):332-37.

63. Paakkari O, Torppa M, Kannas L, et al. Subjective health literacy: Development of a brief instrument for school-aged children. *Scand J Public Health* 2016;44(8):751-57.

64. Manganello JA, DeVellis RF, Davis TC, et al. Development of the Health Literacy Assessment Scale for Adolescents (HAS-A). *J Commun Healthc* 2015;8(3):172-84.

65. Ghanbari S, Ramezankhani A, Montazeri A, et al. Health Literacy Measure for Adolescents (HELMA): Development and Psychometric Properties. *PLoS One* 2016;11(2):e0149202.

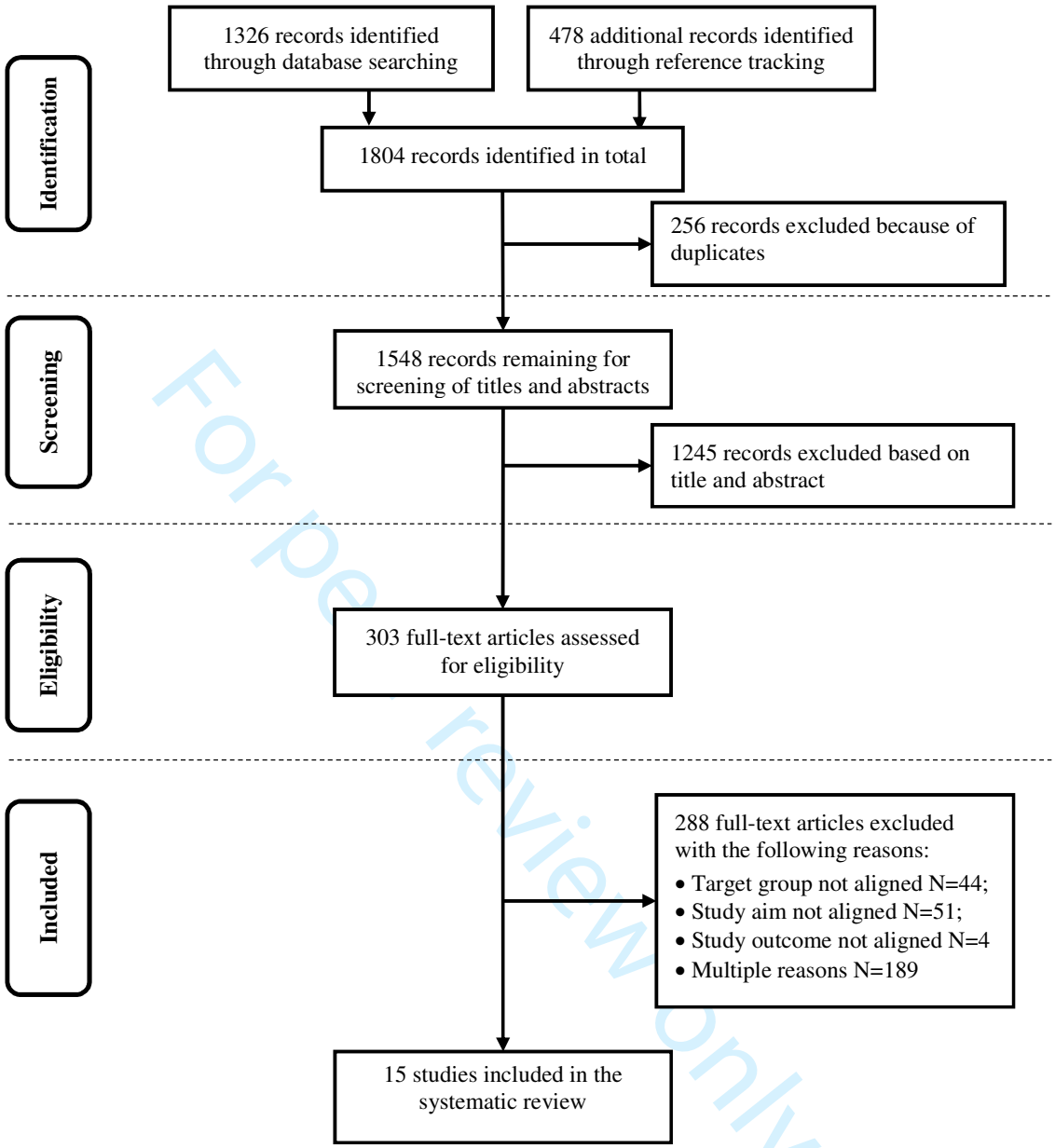

Figure 1. Flowchart of search and selection process according to PRISMA flow diagram

Appendix 1. A systematic review protocol

Measuring the Quality of Child and Adolescent Health Literacy Instruments: A Systematic Review

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1], Sheikh M Alif[1], Geoffrey R Browne[1], Xiaoming Yu[2*]

[1] *School of Population and Global Health, The University of Melbourne, Melbourne, Australia*

[2] *Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China*

* Corresponding author email: gshj1986@gmail.com yxm@bjmu.edu.cn

Background

Health literacy research has been a growing interest by researchers across the globe. The term *'health literacy'* was first used in 1974 in the proceedings of a health education conference discussing health education as a social policy issue affecting the healthcare system, mass communication and the education system (1, 2). However, few references were found regarding health literacy in the literature until 1992 (3). Since 1992, health literacy has been broadly studied both in clinical and public health contexts. In clinical settings, health literacy is typically 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*' by the Institute of Medicine (IOM) in America (4). In such circumstances, health literacy is a derivative concept from literacy and numeracy skills, which is often used as a risk factor that needs to be identified and appropriately managed for patients and health professionals (5). Accordingly, health literacy measurement tools and 'screening aids' for clinicians are developed to assess patient literacy levels, and help health professionals to tailor health information for better communication with their patients (6). From the public health perspective, health literacy is defined and accepted by World Health Organization (WHO) as '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*' (7). This understanding of health literacy identifies it as a broad concept, which is seen as a personal asset to enable individuals to take more control over their health and determinants of health (5). With a different understanding of the concept, health literacy measures vary in a different way. Although health literacy measurement varies and is still

being debated (1, 8-10), there is consistent evidence showing health literacy is of potential importance and considered as a public health goal internationally. A recent WHO report pointed out that poor health literacy skills were associated with riskier behaviours, poorer health status, less self-management and longer hospitalization and more health costs (11).

Based on a preliminary search of health literacy, there were more interests in studies focusing on adult health literacy than adolescent health literacy. However, previous research studies suggested that poor health literacy was a prevalent problem in adolescents. In Australia, the 2006 National Health Literacy Survey reported that 67.6% of adolescents aged 15 to 19 years old did not attain the minimum skills required to deal with health information and service in everyday life (12). Compared with adult health literacy, there are several reasons for the potential importance of adolescent health literacy: 1) adolescents are future mainstream and independent healthcare consumers, a health literate person can contribute to less health care costs, better health status compared to that is not health literate (13); 2) adolescents are at a critical stage of development characterised by physical, emotional and cognitive changes, attempting to prepare for independence but lacking the adequate ability of reasoning and decision-making. Therefore, improving their health literacy skills could support sound health decisions in future (14, 15); 3) low health literacy has been demonstrated to associate with high levels of health-risk behaviors (16, 17) and low levels of health-promoting behaviors for adolescents (18); 4) enhancing health literacy through school-based interventions has great potential for improving students' access to and interpretation of health information (19). Adolescents spend most of their daily time in school, which means they can receive health education and learn how to improve healthy lifestyles and related skills through this setting (20, 21).

Health literacy is more challenging to understand for adolescents than that for adults. Researchers may have different understandings and underlying constructs when using the same definition. That is why there are such a large number of measurement tools of health literacy currently (22, 23), along with some newly-developed health literacy instruments (24). According to Mancuso (1), it is recommended to use specific assessment tools for a specific age group in a specific context. Studies measuring childhood and adolescent health literacy have been a research focus, particularly in the past five years (23). Ormshaw *et al.* (23) conducted a systematic review on measuring childhood and adolescent health literacy in 2011. They found 16 studies that were involved with health literacy measures in children and adolescents. The authors also identified 13 health topics and nine underlying components from existing health literacy instruments. However, the authors did not critically appraise health literacy indices explicitly regarding their validity and reliability. More importantly, the

authors did not assess the methodological quality of each included study. This may undermine the persuasiveness of its conclusion. To fill this knowledge gap, we aim to conduct a systematic review that examines studies' methodological quality and examine reliability and validity of each health literacy instrument, thus providing researchers with unbiased information about which instruments have good psychometric properties. The '*COnsensus-based Standards for the selection of health status Measurement INstruments*' (COSMIN) group has recently developed as a critical appraisal tool (a checklist) to evaluate the methodological quality of studies on measurement properties of health measurement instruments (25). These measurement properties are divided into three domains: reliability, validity, and responsiveness (26). According to the COSMIN checklist, it is possible and scientific to critically appraise and compare psychometric properties of health literacy instruments for children and adolescents.

In this protocol, our target population is adolescent. According to the definition of the WHO, adolescents are those people aged 10 to 19 years and young people aged 10-24 years (27, 28). Given that the term '*adolescent*', '*child*', '*youth*' and '*young people*' is closely related, and Erikson (29) reckoned that children between the ages of 6 and 12 years could learn, compete and co-operate with others, we define our target group as those aged 6-24 years old.

Objectives of the review

This review aims to identify which health literacy instruments have good psychometric properties for children and adolescents. Specifically, there are three objectives:

1) To examine the methodological quality of included studies that aim to measure health literacy in children and adolescents;

2) To examine the measurement properties (i.e. reliability; validity; responsiveness) of health literacy instruments in children and adolescents;

3) To compare the overall rating of measurement properties between each health literacy instrument used in children and adolescents.

Search strategy

Database and search terms

As the term '*health literacy*' was first coined in 1974, articles published from 1st,January 1974 to 30th May 2014 in all languages will be searched. Search strategies will be first designed and then be consulted with two librarian experts. Articles indexed in the following seven databases: Medline, Pubmed, Embase, PsycINFO, CINAHL, ERIC and Cochrane

Library will be searched. The search key terms are '*health literacy*' and '*assessment*' according to previously published studies (1, 23, 30, 31). Age group for '*child, adolescent and young adult*' will be defined in the database settings. The synonyms are listed in **Appendix Table 1**. These synonyms are connected by '*or*' and search strategies are completed by '*and*'.

Appendix Table 1 Searching terms in databases

Key term (1)	Key term (2)
health literacy	health literacy measur*
health AND literacy AND education	health literacy assess*
	health literacy evaluat*
	health literacy instrument*
	health literacy tool*

Other sources of literature

Searching other sources to identify relevant research including:

- Reference lists of identified studies;
- Reference lists of previous systematic reviews on health literacy (1, 23, 30-33).

Eligibility criteria for inclusion and exclusion

According to the guidelines recommended by Cochrane Handbook for systematic reviews (34), inclusion criteria will be addressed regarding population, intervention, comparison, outcome and study design (PICOS):

Inclusion criteria-Participants

The target group should be children and/or adolescents, any age from 6 to 24 years of age.

Inclusion criteria-Interventions and Comparators

As interventional studies are not our interest in this review, it is not applicable to set out guidelines for interventions and comparators

Inclusion criteria-Outcomes

The included studies must be involved with health literacy assessment for children and adolescents, that is, the study should specify the term '*health literacy*', and studies are included if they report on at least one or more attributes of the three measurement properties: 1) reliability; 2) validity; and 3) responsiveness.

Inclusion criteria-Study design

The article should be research-based and peer-reviewed paper including study aim, methods, and results. Also, the study aim should focus on health literacy instrument development or validation.

Exclusion criteria

Studies will be excluded if they are: 1) not focusing on the target group; 2) not focusing on the health literacy instrument development or tool validation; 3) not research-based and peer-reviewed papers including editorials, comments and letters; 4) not reporting findings or results regarding any one of the measurement properties.

Study selection

Search records will be kept including the names of databases searched, keywords, search timeframe, and the search results. All the electronic search results will be initially inputted into the bibliography software of EndNote X7 (Thomson Reuters, New York, NY), and other sources of literature results will be summarised in the print paper. This screening process will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (35). One reviewer will screen studies by titles and abstracts. Secondly, full copies of articles identified will be obtained for thorough screening according to the inclusion criteria by two reviewers independently. Any disagreements in reviewer selections will be resolved at a meeting.

Quality assessment

The methodological quality of each included study will be assessed by two reviewers independently using the COSMIN checklist (25). The checklist consists of nine boxes with 5-18 items concerning methodological standards for how each measurement property should be assessed. Four response options for each item of the COSMIN checklist are defined, representing '*excellent*', '*good*', '*fair*' and '*poor*' quality. An overall score for the methodological quality of a study will be determined for each measurement property separately, by taking the lowest rating of any items in a box ('*worst score counts*') (36). Discrepancies arise between the reviewers will be resolved through discussion, if necessary with a third independent person.

Data extraction

Data extraction will be performed along with the assessment of methodological quality using

the COSMIN checklist (25). In addition, information on the interpretability (e.g. norm scores, floor-ceiling effects, minimal important change of the instruments), generalisability (e.g. characteristics of the study population and sampling procedure), respondent and administrative burden, and forms of administration will be also collected because they are important characteristics of a measurement instrument (26, 37). The data will be entered in an electronic form. Where possible, authors of the original studies will be contacted to obtain essential missing or additional data. Two reviewers will independently extract the data. Consensus should be reached afterward, if necessary with a third independent person.

Data synthesis

The results of the quality of health literacy instruments will be assessed using Terwee's quality criteria (38), to see whether the results of the measurement attributes are '*positive*', '*negative*', or '*indeterminate*'. To summarise the overall ratings of the measurement properties of one health literacy instruments by different authors, the synthesis will be performed by combining the results of the quality of health literacy instruments, the results of methodological quality of health literacy measurement studies and the consistency of their results. The possible overall rating for a measurement property is '*positive*', '*indeterminate*', or '*negative*', accompanied by levels of evidence, similarly as was proposed by the Cochrane Back Review Group (39, 40). One reviewer will perform the data synthesis and a second reviewer will check the synthesised results. Discrepancies of the results will be resolved by discussion.

References

1. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. Nurs Health Sci 2009;11(1):77-89.
2. Simonds SK. Health education as social policy. Health Educ Monogr. 1974;2(1_suppl):1-10.
3. Speros C. Health literacy: concept analysis. J Adv Nurs. 2005;50(6):633-40.
4. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health literacy: a prescription to end confusion. Washington DC, USA: The National Academies Press; 2004.
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med. 2008;67(12):2072-8.
6. Parker RM, Williams MV, Weiss BD, Baker DW, Davis TC, Doak CC, et al. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA. 1999;281(6):552-7.
7. Nutbeam D. Health promotion glossary. Health Promot Int. 1998;13(4):349-64.
8. Mancuso JM. Health literacy: a concept/dimensional analysis. Nurs Health Sci. 2008;10(3):248-55.
9. Higgins JW, Begoray D, MacDonald M. A Social Ecological Conceptual Framework for Understanding Adolescent Health Literacy in the Health Education Classroom. Am J Commun Psychol. 2009;44(3-4):350-62.
10. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int. 2005;20(2):195-203.
11. Kickbusch I, Pelikan JM, Apfel F, Tsouros A. Health literacy: the solid facts. Copenhagen, Denmark: WHO Regional Office for Europe, 2013.
12. Australian Bureau of Statistics. 4233.0-Health Literacy, Australia, 2006. Canberra: Australian Bureau of Statistics; 2008 Jun 25 [cited 2014 Mar 23]; Available from:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4233.02006?OpenDocument.
13. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health. 2012;82(1):28-36.
14. Bacon JL. Adolescent sexuality and pregnancy. Current Opinion in Obstetrics and Gynecology. 2000;12(5):345-7.
15. Paus T. Mapping brain maturation and cognitive development during adolescence. Trends Cogn Sci. 2005;9(2):60-8.
16. Miles, SB, Stipek, D. Contemporaneous and Longitudinal Associations Between Social Behavior and Literacy Achievement in a Sample of Low-Income Elementary School Children. Child Dev. 2006;77(1):103-17.
17. Conwell, LS, O'Callaghan, MJ, Andersen, MJ, Bor, W, Najman, JM, Williams, G. Early adolescent smoking and a web of personal and social disadvantage. J Paediatr Child Health. 2003;39(8):580-5.
18. Chang LC. Health literacy, self-reported status and health promoting behaviours for adolescents in Taiwan. J Clin Nurs. 2011;20(1-2):190-6.
19. Gazmararian, JA, Curran, JW, Parker, RM, Bernhardt, JM, DeBuono, BA. Public health literacy in America: an ethical imperative. Am J Prev Med. 2005;28(3):317-22.
20. Brey RA, Clark SE, Wantz MS. Enhancing health literacy through accessing health information, products, and services: An exercise for children and adolescents. J Sch Health. 2007;77(9):640-4.
21. Kickbusch I. Health literacy: an essential skill for the twenty-first century. Health Educ. 2008;108(2):101-4.
22. Wu A, Begoray D, Macdonald M, Wharf Higgins J, Frankish J, Kwan B, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. Health Promot Int. 2010;25(4):444-52.
23. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. Health Educ. 2013;113(5):433-55.
24. National Institutes of Health (NIH). Office of Behavior and Social Sciences Reasearch. Research Underway in Health Literacy Supported by NIH. 2014. Bethesda, Maryland: NIH; 2014 [cited 2014 Mar 3]; Available: https://obssr-archive.od.nih.gov/scientific_areas/social_culture_factors_in_health/health_literacy/research-underway-health.aspx.
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):539-49
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
27. World Health Organization. Health topics: Adolescent health. Geneva, Switzerland: WHO; 2014 [cited 2014 Mar 7]; Available from: http://www.who.int/topics/adolescent_health/en/.
28. World Health Organization Media Centre. Young people: health risks and solutions, Fact sheet N°345. Geneva, Switzerland: WHO Media Centre; 2014 [cited 2014 Mar 7]; Available: http://www.who.int/mediacentre/factsheets/fs345/en/.
29. Erikson EH. Childhood and society. 2nd ed. New York, USA: W. W. Norton & Company; 1963.
30. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol. 2011;64(4):366-79.
31. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155(2):97-107.
32. Sanders LM, Federico S, Klass P, Abrams MA, Dreyer B. Literacy and child health: a systematic review. Arch Pediatr Adolesc Med 2009;163(2):131-40.
33. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics. 2009;124(Supplement 3):S265-S74.
34. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. London, UK: The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org.
35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.
36. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2012;21(4):651-7.
37. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2002;11(3):193-205.
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.

39. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine. 2009;34(18):1929-41.
40. Van Tulder, M, Furlan, A, Bombardier, C, Bouter, L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine. 2003;28(12):1290-9.

Appendix 2. Search strategy for seven databases

1 MEDLINE (Web of Science) search strategy

MEDLINE database was searched using the Web of Science interface on 16/05/2014 for the period 1974 to 2014.

Basic search:

Set	Results	
# 1	500	**MeSH HEADING:** (health literacy) *OR* ((**TITLE:** (health literacy) *OR* **MeSH HEADING:exp:** (Health Literacy)) *AND* (**TITLE:** (education) *OR* **MeSH HEADING:exp:** (Educational Status) *OR* **MeSH HEADINGS:exp:** (/education) *OR* **MeSH HEADING:exp:** (Teaching) *OR* **MeSH HEADING:exp:** (Educational Status) *OR* **MeSH HEADING:exp:** (Education))) **Refined by: MeSH HEADINGS:** (ADOLESCENT OR YOUNG ADULT OR CHILD) Indexes=MEDLINE Timespan=1974-2014
# 2	3,880	**TOPIC:** ((((health) literacy assess* OR health literacy measur*) OR health literacy evaluat*) OR health literacy instrument*) OR health literacy tool*) Indexes=MEDLINE Timespan=1974-2014
# 3	352	#2 AND #1 Indexes=MEDLINE Timespan=1974-2014

2 PubMed search strategy

PubMed database was searched (Advanced search) on 16/05/2014 for the period 1974 to 16/05/2014.

Set	Results	
# 1	4910	Search **(health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract])** Sort by: **PublicationDate**
# 2	3248385	Search **(child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract])** Sort by: **PublicationDate** **Because if we select age group including child, adolescent, and young adult, the newest papers such as published in 2014 will not be included, the reason maybe the database doesn't update properly. So we use these terms to identify.**
# 3	1887	Search **(health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*)** Sort by: **PublicationDate**
# 4	581	Search **((((health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract]))) AND ((health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*))) AND ((child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract]))** Filters: **Publication date from 1974/01/01 to 2014/05/16** Sort by: **PublicationDate**

3 EMBASE (Ovid) search strategy

EMBASE database was searched using Ovid interface on 16/05/2014 for the period 1974 to current.

Using **.mp** as searching terms (Advanced Search):

Set	Results	
#1	6060	("health literacy" or (health and literacy and education)).mp.
#2	6043	limit 1 to yr="1974 -Current"
#3	671	limit 2 to (school child <7 to 12 years> or adolescent <13 to 17 years>)
#4	170	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	170	limit 4 to yr="1974 -Current"
#6	18	3 and 5

4 PsycINFO (EBSCO) search strategy

PsycINFO database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	786	health literacy OR (health AND literacy AND education)	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#2	133	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#3	133	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

5 CINAHL (EBSCO) search strategy

CINAHL database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	437	health literacy OR (health AND education AND literacy)	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#2	63	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#3	63	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

6 ERIC (EBSCO) search strategy

ERIC database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	59	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#2	2,250	health literacy OR (health AND education AND literacy)	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#3	59	S1 AND S2	Search modes - Boolean/Phrase

7 The Cochrane Library search strategy

The Cochrane Library database was searched on 30/05/2014 for the period January 1974 to May 2014.

Set	Results	Sub-database
#1	4	Cochrane Reviews: There are 4 results from 8483 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Cochrane Reviews'
#2	114	Trials: There are 114 results from 789657 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Trials'
#3	2	Methods Studies: There are 2 results from 15764 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Methods Studies'
#4	120	

Appendix 3. Quality criteria for measurement properties of health literacy instruments

Property	Rating	Quality criteria
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	-	MIC ≤ SDC OR MIC equals or inside LOA
Reliability	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80
	?	Neither ICC/weighted Kappa nor Pearson's r determined
	-	ICC/weighted Kappa < 0.70 OR Pearson's r < 0.80
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement
	-	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity		
Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	-	Factors explain < 50% of the variance
Hypotheses testing	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Responsiveness		
Responsiveness	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs

Note: AUC, Area Under the Curve; ICC, Intra-class Correlation Coefficient; LOA, Limits of Agreement; MIC, Minimal Important Change; SDC, Smallest Detectable Change. + positive rating; ? indeterminate rating; - negative rating.

Appendix 4. Levels of evidence for the overall rating of measurement properties

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodological quality

Note: + positive result; - negative result; ±conflicting result; ? unknown result.

Appendix 5. Reliability and validity results for included instruments

Appendix Table 2. The methodological quality of each study based on reliability for each health literacy instrument

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
NVS (Warsh *et al.*, 2014)	na	na	na			na
NVS (Driessnack *et al.*, 2014)	α=0.71 (n=47)	Poor	na			na
NVS (Hoffman *et al.*, 2013)	α=0.67 (n=229)	Poor	na			na
c-sTOFHLAd (Chang *et al.*, 2012)	α=0.85 (n=300) Item-total correlation=0.44-0.86	Fair	Correlation of test and retest was 0.95 (P<0.001)	Test-retest	1 week	Fair
TOFHLA (Chisolm and Buchanan, 2007)	na	na	na			na
s-TOFHLA (Hoffman *et al.*, 2013)	α=0.89 (n=229)	Poor	na			na
REALM-Teen (Davis *et al.*, 2006)	α=0.94 (n=388)	Poor	γ=0.98	Test-retest	1 week	Fair
REALM-Teen (Hoffman *et al.*, 2013)	α=0.92 (n=229)	Poor	na			
HLAB (Wu *et al.*, 2010)	α=0.92 (n=275) Understanding α=0.88 (n=275) Evaluating α=0.82 (n=275)	Fair	Concordance rate=95%	Inter-rater	na	Poor
MMAHL(Massey *et al.*, 2013)	α=0.83 (n=1208) Item-total correlation=0.39-0.74	Good	na			na
MHL (Levin-Zamir *et al.*, 2011)	α=0.74 (n=1316) Coefficient of reproducibility=0.84 Coefficients of scalability=0.54-0.80	Poor	na			na
DNT-39 (Mulvaney *et al.*, 2013)	α=0.93 (n=61)	Fair	na			na
DNT-14 (Mulvaney *et al.*, 2013)	α=0.82 (n=133) α=0.80 (n=61)	Fair	na			na

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
	α=0.83 (n=72)					
eHEALS (Norman and Skinner, 2006)	α=0.88 (n=664) Item-scale correlation coefficient=0.51-0.76	Fair	The correlations between administrations ranged 0.68-0.40.	Test-retest	Immediately after the intervention; 3-month; 6-month	Fair
CHC Test (Steckelberg *et al.*, 2009)	na	na	Cohen's Kappa was excellent for 277 ratings (κ=0.9-1.0), moderate or good for 31 ratings (κ=0.7-0.89) and poor for 5 ratings (κ=<0.7)	Inter-rater	na	Poor
HKACSS (Schmidt *et al.*, 2010)	Health knowledge χ^2=6.45, *P*=0.17 (n=852) Health communication α=0.73 (n=852) Health attitudes α=0.57 (n=852)	Excellent	na			na
HLAT-51 (Harper, 2014)	Goodness of fit statistic was calculated by each domain (CFI=0.33-0.88; TLI=0.66-0.84; RMSEA=0.09-0.17). The internal consistency statistic was not calculated.	Poor	na			na
HLAT-8 (Abel *et al.*, 2014)	α=0.64 (n=7097 for male) α=0.65 (n=331 for female)	Excellent	na			na

Note: na, no information available. CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; RMSEA, Root Mean Square Error of Approximation; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults.

Appendix Table 3. The methodological quality of each study based on validity for each health literacy instrument

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
NVS (Warsh *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in this study.	Poor	na	na	Hypotheses regarding correlation between scores of a comparator instrument of Gray Silent Reading Test (GSRT) and NVS were formulated before data collection. The NVS and GSRT scores were highly correlated (ρ=0.71, p<0.0001). The NVS score increased with child age (ρ=0.53, p<0.0001).	Fair	na	na
NVS (Driessnack *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in this study.	Poor	na	na	A moderate positive correlation was found between children's NVS scores and their age, and between children's NVS scores and their reports of books numbers (γ_s=0.43, p=0.003; γ_s=0.36, p=0.012, respectively), but not found with their parents' report of the number of children's books at home (γ_s=0.06, p=0.671).	Poor	na	na
NVS (Hoffman *et al.*, 2013)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.49 (p<0.01).	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	this study.							
c-sTOFHLAd (Chang *et al.*, 2012)	The c-sTOFHLAd was translated from the short-version of TOFHLA according to translation procedures and was tested among 30 adolescents to ensure appropriateness.	Good	Confirmatory factor analysis was conducted to determine structural validity. One-factor model indicated an acceptable fit to the data according structural equation modelling analysis.	Fair	Convergent validity was measured between c-sTOFHLAd and the rapid estimate of adult literacy in medicine (REALM), with a correlation coefficient of 0.74 (p<0.001).	Fair	Semantic equivalence was measured by the content validity index (CVI). All items were rated by the experts as having a CVI>0.85. Thirty adolescents were chosen to determine and ensure the cultural congruence of the instrument.	Fair
TOFHLA (Chisolm and Buchanan, 2007)	The TOFHLA was developed from a literacy expert after reviewing commonly used hospital texts and a pilot test. No target population is involved in this study.	Poor	na	na	The reading comprehension component (TOFHLA-R) was significantly collated with the Wide-Ranging Achievement Test (WRAT3) and the rapid estimate of adult literacy in medicine (REALM) (ρ=0.59, p<0.001; ρ=0.60, p<0.001 respectively), however, no correlation were found with the numeracy component (TOFHLA-N) (ρ=0.11, p=0.45; ρ=0.18, p=0.22 respectively).	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013)	The s-TOFHLA was developed based on previous data analysis, perceived importance and frequency of the task in the healthcare settings.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.28 (p<0.01).	Fair	na	na
REALM-	The REALM-Teen was	Good	na	na	Convergent validity was	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
Teen (Davis *et al.*, 2006)	developed based on a preliminary test and a structured interview among adolescents. And a panel of experts reviewed the word list.				measured between REALM-Teen and the WRAT-3 (r=0.83) and SORT-R (r=0.93).			
REALM-Teen (Hoffman *et al.*, 2013)	The REALM-Teen was developed based on a preliminary test and structured interview among adolescents. And a panel of experts reviewed the word list.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.40 (p<0.01).	Poor	na	na
HLAB (Wu *et al.*, 2010)	Previous experience and literature review were used to develop items; 10 students were pilot-tested for appropriateness of wording, content and format of the final instrument.	Good	na	na	Correlations were assumed between socio-demographic variables and the overall scores. Socio-demographics of gender, age when came to Canada to live, speaking a language other than English were correlated with the scores of HLAB (β=-0.18, p=0.004; β=-0.22, p=0.014; β=-0.20, p=0.008 respectively). No convergent validity is assessed.	Fair	na	na
MMAHL (Massey *et al.*, 2013)	Domains were established from literature review and focus group. Items were developed either using adaptation of existing relevant items or created by the research team.	Good	Explorative principal components factor analysis was conducted and 49.8% of the variance was accounted by 6 factors.	Good	na	na	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
MHL (Levin-Zamir *et al.*, 2011)	The face validity was discussed in the focus group during pilot test. The content validity was analysed using theory and operational definitions of health literacy and media literacy, and adolescents were invited to write detailed, anonymous responses.	Good	na	na	As hypothesised, MHL was associated with socio-economic determinants, particularly with gender (β=1.25, p<0.001) and mother's education (β=0.16, p=0.04). In addition, MHL was also associated with health behaviours (β=0.03, p=0.05) and health empowerment (β=0.36, p<0.001).	Good	na	na
DNT-39 (Mulvaney *et al.*, 2013)	The DNT-39 was developed from the original 43-item version DNT-43 by eliminating questions specific to type 2 diabetes. An expert team developed the DNT-43 and refined it.	Poor	na	na	The DNT-39 was associated with WRAT-3 and parent education (ρ=0.40, p=0.001; ρ=0.29, p=0.028 respectively)	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013)	The DNT-14 was developed from the original 15-item version DNT-15 by eliminating 1 question specific to type 2 diabetes. An expert team developed the DNT-15 by data analysis from DNT-43.	Poor	na	na	The DNT-14 was associated with the Wide-Ranging Achievement Test (WRAT3), parent education, diabetes problem solving and HbA1c (ρ=0.36, p=0.005; ρ=0.31, p=0.019; ρ=0.27, p=0.023; ρ=-0.34, p=0.004 respectively)	Fair	na	na
eHEALS (Norman and Skinner, 2006)	The eHEALS was developed by the expert team and pilot-tested and refined by feedback from	Good	Explorative principal components factor analysis was conducted and 56% of the	Fair	Correlations were assumed between eHEALS and other measured variables (gender, age, use of information technology	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	participants.		variance was accounted by a single factor. The factor loadings ranged from 0.60-0.84 among the 8 items.		overall, self-evaluations of health). However, only gender difference was found at baseline level of eHealth literacy (t=2.236, *p*=0.026). No convergent validity is assessed.			
CHC Test (Steckelberg *et al.*, 2009)	The CHC Test was developed by the research team and pre-tested by collecting qualitative data and quantitative field test.	Good	IRT test for determining dimensionality was performed.	Poor	na	na	na	na
HKACSS (Schmidt *et al.*, 2010)	The HKACSS items were taken from a previous health survey and selected basing on consideration of item content.	Good	na	na	As hypothesised, health communication, attitudes and self-efficacy were significantly related to each other (ρ=0.15-0.38, *P*<0.05). And children from higher educational background showed a better knowledge and communicated more about health topics (β=0.16, *p*<0.05).	Good	na	na
HLAT-51 (Harper, 2014)	The expert team evaluated the initial items using a 5-point Likert scale according to their research experience. And 144 college students were invited to complete a pilot test.	Good	Comprehension (CFI=0.80; TLI=0.78; RMSEA=0.09); health numeracy (CFI=0.57; TLI=0.48; RMSEA=0.09); media literacy (CFI=0.88; TLI=0.84; RMSEA=0.07); digital literacy (CFI=0.33; TLI=0.06;	Poor	na		na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
			RMSEA=0.16); health information seeking (CFI=0.80; TLI=0.66; RMSEA=0.17)					
HLAT-8 (Abel *et al.*, 2014)	The research team developed the HALT-8 drawing on literature review and their own experience. No target population is involved in this study.	Poor	Explorative principal components factor analysis was conducted and 72.96% of the variance was accounted by four factors among male. In addition, the factor structure was validated using confirmatory factor analysis (CFI=0.99; TLI=0.97; RMSEA=0.03; SRMR=0.03).	Excellent	Hypotheses were formulated a priori regarding correlations between health literacy and gender, socio-cultural characteristics and health values. Results showed that female, higher educational status, and a stronger health valuation were associated with higher HL scores (p<0.05, respectively).	Good	na	na

Note: na, no information available. CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; RMSEA, Root Mean Square Error of Approximation; SRMR, Standardized Root Mean Square Residual; SORT-R, Slosson Oral Reading Test-Revised; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults; WRAT-3, Wide-Range Achievement Test-Revised.

Research Checklist. PRISMA checklist for reporting systematic review

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Appendix 1
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	9-10

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9-10
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10; Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10; 13-14; Table 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	19; Table 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	19; Table 4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	19; Table 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	19; Table 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	23-28
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	28
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	28-29
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

The Quality of Health Literacy Instruments used in Children and Adolescents: A Systematic Review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020080.R1
Article Type:	Research
Date Submitted by the Author:	29-Mar-2018
Complete List of Authors:	Guo, Shuaijun; University of Melbourne School of Population and Global Health, Armstrong, Rebecca; University of Melbourne, Waters, E; University of Melbourne, School of Population and Global Health Sathish, Thirunavukkarasu; University of Melbourne School of Population and Global Health, Centre for Health Equity; Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Alif, Sheikh; University of Melbourne, School of Population and Global Health Browne, Geoff; University of Melbourne, School of Population and Global Health Yu, Xiaoming; Peking University, Institute of Child and Adolescent Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Global health
Keywords:	Measurement properties, Health literacy, Children, Adolescents, Systematic review

The Quality of Health Literacy Instruments used in Children and Adolescents: A Systematic Review

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1,2], Sheikh M Alif[1], Geoffrey R Browne[1], Xiaoming Yu[3*]

[1] School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[2] Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

[3] Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China

**** Corresponding Author***

Shuaijun Guo Postal address: 526-9, Level 5, 207 Bouverie Street, Carlton, Victoria, Australia 3010 Email: gshj1986@gmail.com Tel: +61 452 110 331

Xiaoming Yu Postal address: Room 209, Institute of Child and Adolescent Health, School of Public Health, 38 Xueyuan Road, Haidian, Beijing, China 100083 Email: yxm@bjmu.edu.cn Tel: +86 10 8280 2631

Word count: 5502

ABSTRACT

Objective: Improving health literacy at an early age is crucial to personal health and development. Although health literacy in children and adolescents has gained momentum in the past decade, it remains an under-researched area, particularly health literacy measurement. This study aimed to examine the quality of health literacy instruments used in children and adolescents and to identify the best instrument for field use.

Design: Systematic review.

Setting: A wide range of settings including schools, clinics and communities.

Participants: Children and/or adolescents aged 6-24 years.

Primary and secondary outcome measures: Measurement properties (reliability, validity and responsiveness) and other important characteristics (e.g. health topics, components or scoring systems) of health literacy instruments.

Results: There were 29 health literacy instruments identified from the screening process. When measuring health literacy in children and adolescents, researchers mainly focus on the functional domain (basic skills in reading and writing) and consider participant characteristics of developmental change (of cognitive ability), dependency (on parents) and demographic patterns (e.g. racial/ethnic backgrounds), less on differential epidemiology (of health and illness). The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. More than half (62.9%) of measurement properties were unknown, due to either poor methodological quality of included studies or a lack of reporting or assessment. The 8-item Health Literacy Assessment Tool (HLAT-8) showed best evidence on construct validity and the Health Literacy Measure for Adolescents showed best evidence on reliability.

Conclusions: More rigorous and high-quality studies are needed to fill the knowledge gap in measurement properties of health literacy instruments. Although it is

challenging to draw a robust conclusion about which instrument is the most reliable and the most valid, this review provides important evidence that supports the use of the HLAT-8 to measure childhood and adolescent health literacy in future school-based research.

Keywords: measurement properties; health literacy; children; adolescents; systematic review

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The COSMIN checklist was used as a methodological framework to rate the methodological quality of included studies.
- This review has updated previous three reviews of childhood and adolescent health literacy measurement tools and identified 19 additional new health literacy instruments.
- Including only studies that aimed to develop or validate a health literacy instrument may eliminate studies that used a health literacy instrument for other purposes.
- Individual subjectivity exists in the screening and data synthesis stages.

INTRODUCTION

Health literacy is a personal resource that enables an individual to make decisions for healthcare, disease prevention and health promotion in everyday life.[1] As defined by the World Health Organisation,[2] health literacy refers to '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*.' The literature has shown that health literacy is an independent and more direct predictor of health outcomes than socio-demographics.[3 4] People with low health literacy are likely to have worse health-compromising behaviours, higher healthcare costs and poorer health status.[5] Given the close relationship between health literacy and health outcomes, many countries have adopted health literacy promotion as a key strategy to reduce health inequities.[6]

From a health promotion perspective, improving health literacy at an early age is crucial to childhood and adolescent health and development.[7] As demonstrated by Diamond *et al.*[8] and Robinson *et al.*,[9] health literacy interventions for children and adolescents can bring about improvements in healthy behaviours and decreased used of emergency department services. Although health literacy in young people has gained increasing attention, with a rapidly growing number of publications in the past decade,[10-13] childhood and adolescent health literacy is still under-researched. According to Forrest *et al.*'s 4D model,[14 15] health literacy in children and adolescents is mediated by four additional factors compared to adults: (1) *developmental* change: children and adolescents have less well-developed cognitive ability than adults; (2) *dependency*: children and adolescents depend more on their parents and peers than adults do; (3) *differential* epidemiology: children and adolescents experience a unique pattern of health, illness and disability; and (4) *demographic* patterns: many children and adolescents living in poverty or in single-parent families are neglected and so require additional care. These four differences pose significant challenges for researchers when measuring health literacy in children and adolescents.

Health literacy is a broad and multi-dimensional concept with varying definitions.[16] This paper uses the definition by Nutbeam who states that health literacy consists of three domains: functional, interactive and critical.[17] The *functional* domain refers to

basic skills in reading and writing health information, which are important for functioning effectively in everyday life. The *interactive* domain represents advanced skills that allow individuals to extract health information and derive meaning from different forms of communication. And the *critical* domain represents more advanced skills that can be used to critically evaluate health information and take control over health determinants.[17] Although health literacy is sufficiently explained in terms of its definitions[17-19] and theoretical models,[4 7] its measurement remains a contested issue. There are two possible reasons for this. One reason is the large variety of health literacy definitions and conceptual models,[12 16] and the other reason is that researchers may have different study aims, populations and contexts when measuring health literacy.[20 21]

Currently, there are three systematic reviews describing and analysing the methodology and measurement of childhood and adolescent health literacy.[10 11 13] In 2013, Ormshaw *et al.*[10] conducted a systematic review of child and adolescent health literacy measures. This review used four questions to explore health literacy measurement in children and adolescents: "*What measurement tools were used? What health topics were involved? What components were identified?* and *Did studies achieve their stated aims?*" The authors identified 16 empirical studies, with only six of them evaluating health literacy measurement as their primary aim. The remaining studies used health literacy measures as either a comparison tool when developing other new instruments or as a dependent variable to examine the effect of an intervention program. Subsequently, in 2014, Perry[11] conducted an integrative review of health literacy instruments used in adolescents. In accordance with the eligibility criteria, five instruments were identified. More recently, Okan *et al.*[13] conducted another systematic review on generic health literacy instruments used for children and adolescents with the aim of identifying and assessing relevant instruments for first-time use. They found fifteen generic health literacy instruments used for this target group.

Although these three reviews provide general knowledge about the methodology and measurement of health literacy in young people, they all have limitations. Ormshaw *et al.*[10] did not evaluate measurement properties of each health literacy instrument. Although Perry[11] and Okan *et al.*[13] summarised measurement properties of each

instrument, the information provided was limited, mostly descriptive, and lacked a critical appraisal. Notably, none of the three reviews considered the methodological quality of included studies[10 11 13]. A lack of quality assessment of studies raises concerns about the utility of such reviews for evaluating and selecting health literacy instruments for children and adolescents. Therefore, it is still unclear which instrument is the best in terms of its validity, reliability and feasibility for field use. In addition, it is also unclear how Nutbeam's three-domain health literacy model and Forrest *et al.*'s 4D model are considered in existing health literacy instruments for children and adolescents.

To fill these knowledge gaps, this systematic review aimed to examine the quality of health literacy instruments used in the young population and to identify the best instrument for field use. We expect the findings will assist researchers in identifying and selecting the most appropriate instrument for different purposes when measuring childhood and adolescent health literacy.

METHODS

Following the methods for conducting systematic reviews outlined in the Cochrane Handbook,[22] we developed a review protocol (See **Appendix 1**, PROSPERO registered ID: CRD42018013759) prior to commencing the study. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement[23] (See **Research Checklist**) was used to ensure the reporting quality of this review.

Literature search

The review took place over two time periods: The initial systematic review covered the period between 1 January 1974 and 16 May 2014 (period 1). The start date of 1974 was chosen because this was the date from which the term '*health literacy*' was first used.[24] A second search was used to update the review in February 2018. It covered the period 17 May 2014 to 31 Jan 2018 (period 2). The databases searched were: Medline, PubMed, Embase, PsycINFO, CINAHL, ERIC and the Cochrane Library. The search strategy was designed on the basis of previous reviews[5 10 25 26] and in consultation with two librarian experts. Three types of search terms were used: (1)

construct-related terms: '*health literacy*' OR '*health and education and literacy*'; (2) outcome-related terms: '*health literacy assess**' OR '*health literacy measure**' OR '*health literacy evaluat**' OR '*health literacy instrument**' OR '*health literacy tool**'; and (3) age-related terms: '*child**' OR '*adolescent**' OR '*student**' OR '*youth*' OR '*young people*' OR '*teen**' OR '*young adult.*'

No language restriction was applied. The detailed search strategy for each database is available in **Appendix 2**. As per the PRISMA flow diagram,[23] the references from included studies and from six previously published systematic reviews on health literacy[5 10 25-28] were also included.

Eligibility criteria

Studies had to fulfil the following criteria to be included: (1) the stated aim of the study was to develop or validate a health literacy instrument; (2) participants were children or adolescents aged 6 to 24. This broad age range was used because the age range for '*children*' (under the age of 18) and '*adolescents*' (aged 10 to 24) overlap[29] and also because children aged over 6 are able to learn and develop their own health literacy[30]; (3) the term '*health literacy*' was explicitly defined, although studies assessing health numeracy (the ability to understand and use numbers in healthcare settings) were also considered; and (4) at least one measurement property (reliability, validity and responsiveness) was reported in the outcomes.

Studies were excluded if: (a) the full paper was not available (i.e. only a conference abstract or protocol was available); (b) they were not peer-reviewed (e.g. dissertations, government reports); or (c) they were qualitative studies.

Selection process

All references were imported into EndNote X7 software (Thomson Reuters, New York, NY) and duplicate records were initially removed before screening. Next, one author (GS) screened all studies based on title and abstract. Full-text papers of the remaining titles and abstracts were then obtained separately for each review round (period 1 and period 2). All papers were screened by two independent authors (GS

and SA). At each major step of this systematic review, discrepancies between authors were resolved through discussion.

Data extraction

The data that were extracted from papers were: characteristics of included studies (e.g. first author, published year and country), general characteristics of instruments (e.g. health topics, components and scoring systems), methodological quality of the study (e.g. internal consistency, reliability and measurement error) and ratings of measurement properties of included instruments (e.g. internal consistency, reliability and measurement error). Data extraction from full-text papers published during period 1 was performed by two independent authors (GS and TS), whereas data extraction from full-text papers published during period 2 was conducted by one author (GS) and then checked by a second author (TS).

Methodological quality assessment of included studies

The methodological quality of included studies was assessed using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist.[31] The COSMIN checklist is a critical appraisal tool containing standards for evaluating the methodological quality of studies on measurement properties of health measurement instruments.[32] Specifically, nine measurement properties (internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity and responsiveness) were assessed.[32] Since there is no agreed-upon '*gold standard*' for health literacy measurement,[33 34] criterion validity was not assessed in this review. Each measurement property section contains 5 to 18 evaluating items. For example, '*internal consistency*' is evaluated against 11 items. Each item is scored using a four-point scoring system ('*excellent*', '*good*', '*fair*' or '*poor*'). The overall methodological quality of a study is obtained for each measurement property separately, by taking the lowest rating of any item in that section (i.e. '*worst score counts*'). Two authors (GS and TS) independently assessed the methodological quality of included studies published during period 1, whereas the quality of included

studies published during period 2 was assessed by one author (GS) and then checked by another (TS).

Evaluation of measurement properties for included instruments

The quality of each measurement property of an instrument was evaluated using quality criteria proposed by Terwee *et al.*[35], who are members of the group that developed the COSMIN checklist (See **Appendix 3**). Each measurement property was given a rating result ('+' positive, '-' negative, '*?*' indeterminate and '*na*' no information available).

Best evidence synthesis: levels of evidence

As recommended by the COSMIN checklist developer group,[32] '*a best evidence synthesis*' was used to synthesise all the evidence on measurement properties of different instruments. The procedure used was similar to the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework[36], a transparent approach to rating quality of evidence that is often used in reviews of clinical trials.[37] Given that this review did not target clinical trials, the GRADE framework adapted by the COSMIN group was used.[38] Under this procedure, the possible overall rating for a measurement property is '*positive*', '*negative*', '*conflicting*' or '*unknown*', accompanied by levels of evidence ('*strong*', '*moderate*' or '*limited*') (See **Appendix 4**). Three steps were taken to obtain the overall rating for a measurement property. First, the methodological quality of a study on each measurement property was assessed using the COSMIN checklist. Measurement properties from '*poor*' methodological quality studies did not contribute to '*the best evidence synthesis*'. Second, the quality of each measurement property of an instrument was evaluated using Terwee's quality criteria.[35] Third, the rating results of measurement properties in different studies on the same instrument were examined whether consistent or not. This best evidence synthesis was performed by one author (GS) and then checked by a second author (TS).

Patient and Public Involvement

Children and adolescents were not involved in setting the research question, the outcome measures, or the design or implementation of this study.

Results

The initial search identified 2790 studies. After duplicates and initial title/abstract screening, 361full-text articles were identified and obtained. As per the eligibility criteria, 29 studies were included,[39-53] yielding 29 unique health literacy instruments used in children and adolescents (See **Figure 1**).

Characteristics of included studies

Of the 29 studies identified, 25 were published between 2010 and 2017 (See **Table 1**). Most included studies were conducted in Western countries (n=20), with eleven studies carried out in the USA. The target population (aged 7 to 25) could be roughly classified into three subgroups: children aged 7 to 12 (n=5), adolescents aged 13 to 17 (n=20) and young adults aged 18 to 25 (n=4). Schools (n=17) were the most common recruitment settings, compared to clinical settings (n=8) and communities (n=4).

Table 1. Characteristics of included studies

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
1	Davis *et al.*[41] (2006)	USA	Adolescents aged 10-19 years (mean age=14.8±1.9)	REALM-Teen	1533 (47.4)	na	Middle schools, high schools, paediatric primary care clinic and summer programs
2	Norman and Skinner[43] (2006)	Canada	Adolescents aged 13-21 years (mean age=14.95±1.24)	eHEALS	664 (55.7)	Sampling from one arm of a randomized controlled trial	Secondary schools
3	Chisolm and Buchanan[48] (2007)	USA	Young people aged 13-17 years (mean age=14.7)	TOFHLA	50 (48.0)	na	Children's hospital
4	Steckelberg *et al.*[47] (2009)	Germany	Students in Grade 10-11 and university	CHC Test	Sample 1: 322 (36.6) Sample 2: 107 (32.7)	na	Secondary schools, university
5	Schmidt *et al.*[46] (2010)	Germany	Children aged 9-13 years (mean age=10.4)	HKACSS	852 (52.9)	na	Primary school
6	Wu *et al.*[40] (2010)	Canada	Students in Grade 8-12	HLAB	275 (48.0)	Convenience sampling	Secondary schools
7	Levin-Zamir *et al.*[49] (2011)	Israel	Adolescents in Grade 7, 9, 11 (approximately age 13, 15 and 17)	MHL	1316 (52.0)	Probability sampling and random cluster sampling	Public schools
8	Chang *et al.*[51] (2012)	Taiwan	Students in high school (mean age=16.01±1.02)	c-sTOFHLAd	300 (52.6)	Multiple-stage stratified random sampling	High schools
9	Hoffman *et al.*[50] (2013)	USA	Youth aged 14-19 years (mean age=17)	REALM-Teen; NVS; s-TOFHLA	229 (61.6)	na	Private high school
10	Massey *et al.*[44] (2013)	USA	Adolescents aged 13-17 years (mean age=14.8)	MMAHL	1208 (37.6)	Sampling from a large health insurance network	Publicly health insurance network
11	Mulvaney *et al.*[53] (2013)	USA	Adolescents aged 12-17 years (Sample 1: mean age=13.92; Sample 2: mean age=15.10)	DNT-39 and DNT-14	Sample 1: 61 (52.5) Sample 2: 72 (55.6)	na	Diabetes clinics

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
12	Abel *et al.*[45] (2014)	Switzerland	Young adults aged 18-25 years (male mean age: 19.6; female mean age=18.8)	HLAT-8	7428 (95.5)	Sampling from compulsory military service for males and two-stage random sampling for females	Compulsory military service, communities
13	Driessnack *et al.*[54] (2014)	USA	Children aged 7-12 years	NVS	47 (53.0)	Convenience sampling	The science centre
14	Harper[42] (2014)	New Zealand	Students aged 18-24 years	HLAT-51	144 (41.0)	Purposeful sampling	College
15	Warsh *et al.*[39] (2014)	USA	Children aged 7-17 years (median age=11)	NVS	97 (46.0)	Convenience sampling	Paediatric clinics
16	Liu et al.[55] (2014)	Taiwan	Children in grade six	CHLT	162609 (51.1)	National sampling	Primary schools
17	Ueno *et al.*[56] (2014)	Japan	Students in high school Grade 1 (age range: 15-16 years)	VOHL	162 (46.3)	Convenience sampling	A senior high school
18	Manganello *et al.*[57] (2015)	USA	Youth aged 12-19 years (mean age=15.6)	HAS-A	272 (37.0)	Convenience sampling	A paediatric clinic and the community
19	Naigaga *et al.*[58] (2015)	Uganda	Pregnant adolescents aged 15-19 years	MaHeLi	384 (0)	Random sampling	Health centres
20	de Jesus Loureiro *et al.*[59] (2015)	Portugal	Adolescents and young people aged 14-24 years (mean age=16.75±1.62)	QuALiSMental	4938 (43.3)	Multi-stage cluster random sampling	Schools
21	McDonald *et al.*[60] (2016)	Australia	Adolescents and young adults diagnosed with cancer (age range: 12-24 years)	FCCHL-AYAC	105 (33.3)	Sampling from a support organisation	An organisation for young people living with cancer
22	Smith *et al.*[61] (2016)	USA	Deaf/hard-of hearing and hearing adolescents in high school (mean age=17.0±0.84 and 15.8±1.1)	ICHL	Sample 1: 154 (53.2) Sample 2: 89 (33.0)	Convenience sampling	Medical centre summer programs
23	Ghanbari *et al.*[62] (2016)	Iran	Adolescents aged 15-18 years (mean age=16.2±1.03)	HELMA	582 (48.8)	Multi-stage sampling	High schools
24	Paakkari *et al.*[63] (2016)	Finland	Pupils (7th graders aged 13 years: n=1918; 9th graders aged 15 years: n=1935)	HLSAC	3853 (na)	Cluster sampling	Secondary schools
25	Manganello *et*	USA	Adolescents aged 14-19 years	REALM-TeenS	174 (na)	na	Adolescent medicine

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
	al.[64] (2017)		(mean age=16.6)				clinics
26	Tsubakita *et al.*[65] (2017)	Japan	Young adults aged 18-26 years (mean age=19.65±1.34)	funHLS-YA	1751 (76.8)	Convenience sampling	A private university
27	Intarakamhang *et al.*[66] (2017)	Thailand	Overweight children aged 9-14 years	HLS-TCO	2000 (na)	Quota-stratified random sampling	Schools
28	Bradley-Klug *et al.*[67] (2017)	USA	Youth and young adults with chronic health conditions aged 13-21 years (mean age=17.6)	HLRS-Y	204 (24.3)	National sampling	Community-based agencies and social media outlets
29	Quemelo *et al.*[68] (2017)	Brazil	University students (mean age=22.7±5.3)	p_HLAT-8	472 (33.9)	na	A university

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

General characteristics of included instruments

Compared to previous systematic reviews,[10 11 13] this review identified 19 additional new health literacy instruments (eHEALS, s-TOFHLA, DNT-39, DNT-14, HLAT-51, HLAT-8, CHLT, VOHL, HAS-A, QuALiSMental, FCCHL-AYAC, ICHL, HELMA, HLSAC, REALM-TeenS, funHLS-YA, HLS-TCO, HLRS-Y and p_HLAT-8). The 29 health literacy instruments were classified into three groups based on whether the instrument was developed bespoke for the study or not (See **Table 2**).[10] The three groups were: (1) newly-developed instruments for childhood, adolescent and youth health literacy (n=20);[40-47 49 50 55-58 61-63 65-67] (2) adapted instruments that were based on previous instruments for adult/adolescent health literacy (n=6);[51 53 59 60 64 68] and (3) original instruments that were developed for adult health literacy (n=3).[39 48 50 52]

Health literacy domains and components

Next, Nutbeam's three-domain health literacy model[17] was used to classify the 29 instruments according to which of the commonly-used components of health literacy were included. Results showed that ten instruments measured only functional health literacy[39 41 48 50-53 56 64 65] and one instrument measured only critical health literacy.[47] There was one instrument measuring functional and interactive health literacy[46], one measuring functional and critical health literacy[40], and one measuring interactive and critical health literacy.[61] Fifteen instruments measured health literacy by all three domains (functional, interactive and critical).[42-45 49 55 57-60 62 63 66-68]

Consideration of participants' characteristics

As per Forrest *et al.*'s 4D model,[14 15] the 29 included instruments were examined for whether participant characteristics were considered when developing a new instrument or validating an existing instrument. Results showed most of the health literacy instruments considered developmental change, dependency and demographic patterns. In contrast, only seven instruments considered differential epidemiology.[53 58 60 61 66 67]

Health topics, contents and readability levels

Health literacy instruments for children and adolescents covered a range of health topics such as nutrition and sexual health. Most instruments (n=26) measured health literacy in healthcare settings or health promotion contexts (e.g. general health topics, oral health, or mental health), while only three instruments measured health literacy in the specific context of eHealth or media health.[42 43 49] In relation to the readability of tested materials, only eight health literacy instruments reported their readability levels, ranging from 2th to 19.5th grade.

Burden and forms of administration

The time to administer was reported in seven instruments, ranging from 3 to 90 minutes. There were three forms of administration: self-administered instruments (n=19), interviewer-administered instruments (n=9), and video-assisted, interviewer-administered instruments (n=1). Regarding the method of assessment, fifteen instruments were performance-based, eleven instruments were self-report, and three included both performance-based and self-report items.

Table 2. General and important characteristics of included instruments used in children and adolescents

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
1	NVS[50 54 39]	***Functional HL*** 1. Reading comprehension (2) 2. Numeracy (4)	Demographic patterns	Nutrition-related information about the label of an ice cream container (na)	Open-ended	Score range: 0-6; Ordinal category: 0-1: high likelihood of limited literacy; 2-3: possibility of limited literacy; 4-6: adequate literacy	No longer than 3 minutes	Interviewer-administered & Performance-based
2	TOFHLA[48]	***Functional HL*** 1. Reading comprehension (50) 2. Numeracy (17)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4.3 grade), a standard informed consent form (10.4 grade), patients' rights and responsibilities section of a Medicaid application form (19.5 grade), actual hospital forms & labelled prescription vials (9.4 grade)	4 response options	Score range: 0-100; Ordinal category: 0-59: inadequate health literacy; 60-74: marginal health literacy; 75-100: adequate health literacy	12.9 minutes (8.9-17.3 minutes)	Interviewer-administered & Performance-based
3	s-TOFHLA[50]	***Functional HL*** 1. Reading comprehension (36)	Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score range: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	na	Interviewer-administered & Performance-based
4	c-sTOFHLAd[51]	***Functional HL*** 1. Reading comprehension (36)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score range: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	20-minute class period	Self-administered & Performance-based
5	REALM-Teen[41 50]	***Functional HL*** 1. Reading recognition (66)	Developmental change Demographic patterns	66 health-related words such as weight, prescription and tetanus (6th grade)	Open-ended	Score range: 0-66; Ordinal category: 0-37: ≤ 3rd; 38-47: 4th-	2-3 minutes	Interviewer-administered & Performance-

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
						5th; 48-58: 6th-7th; 59-62: 8th-9th; 63-66: ≥ 10th		based
6	HLAB[40]	***Functional and critical HL*** 1. Understanding health information (30) 2. Evaluating health information (17)	Developmental change Demographic patterns	A range of topics such as nutrition and sexual health (pilot-tested)	Open-ended	Score range: 0-107; Continuous score	Two regular classroom sessions	Self-Administered & Performance-based
7	MMAHL[44]	***Functional, interactive and critical HL*** 1. Patient-provider encounter (4) 2. Interaction with the health care system (5) 3. Rights and responsibilities (7) 4. Confidence in using health information from personal source (3) 5. Confidence in using health information from media source (3) 6. Health information seeking competency using the Internet (2)	Developmental change Demographic patterns Dependency	Experiences of how to access, navigate and manage one's health care and preventive health needs (6th grade)	5-point Likert scale	Score range: na; Continuous score	na	Self-administered & Self-reported
8	MHL[49]	***Functional, interactive and critical HL*** 1. Content identification (6) 2. Perceived influence on behaviour (6) 3. Critical analysis and intended (6)	Dependency Demographic patterns	Nutrition/dieting, physical activity, body image, sexual activity, cigarette smoking, alcohol consumption, violent behaviour, safety habits and/or friendship and family connectedness (pilot-tested)	Open-ended & multiple choice	Score range: 0-24; Continuous score	na	Video-assisted interviewer-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
		4. Action/reaction (6)						
9	DNT-39[53]	***Functional health literacy*** 1. Health numeracy (39)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score range: 0-100; Continuous score	na	Interviewer-administered & Performance-based
10	DNT-14[53]	***Functional health literacy*** 1. Health numeracy (14)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score range: 0-100; Continuous score	na	Interviewer-administered & Performance-based
11	eHEALS[43]	***Functional, interactive and critical HL*** 1. Accessing health information (4) 2. Evaluating health information (2) 3. Applying health information (2)	Developmental change Dependency Demographic patterns	General health topics about online health information (pilot-tested)	5-point Likert scale	Score range: na; Continuous score	na	Self-Administered & Self-reported
12	CHC Test[47]	***Critical HL*** 1. Understanding medical concepts (15) 2. Searching literature skills (22) 3. Basic statistics (18) 4. Design of experiments and sampling (17)	Developmental change Demographic patterns	Echinacea and common cold, magnetic resonance imaging in knee injuries, treatment of acne, breast cancer screening (pilot-tested)	Open-ended & multiple choice	na	Less than 90 minutes	Interviewer-administered & Performance-based
13	HKACSS[46]	***Functional and interactive HL*** 1. Health knowledge (3) 2. Health attitudes (4) 3. Health communication (3) 4. Self-efficacy (3)	Developmental change Dependency Demographic patterns	Physical activities, nutrition, smoking, vaccination, tooth health and general health (na)	2 response options; 5-point Likert scale; 4-point Likert	Score range: na; Continuous score	na	Self-Administered & Performance-based & Self-reported

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
					scale			
14	HLAT-51[42]	***Functional, interactive and critical HL*** 1. Comprehension skill (20) 2. Health numeracy (11) 3. Media literacy (8) 4. Digital literacy (12)	Developmental change Dependency Demographic patterns	Health topics such as gout and uric acid, high cholesterol and triglyceride levels, health-information-seeking skills (na)	Yes/no; multiple choice	na	30-45 minutes	Self-administered & Performance-based & Self-reported
15	HLAT-8[45]	***Functional, interactive and critical HL*** 1. Understanding health information (2) 2. Finding health information (2) 3. Interactive health literacy (2) 4. Critical health literacy (2)	Dependency Demographic patterns	General health topics in people's daily life (na)	5-point Likert scale; 4-point Likert scale	Score range: 0-37; Continuous score	na	Self-administered & Self-reported
16	CHLT[55]	***Functional, interactive and critical HL*** 1. Health knowledge (11) 2. Health attitude (16) 3. Health skills (5)	Developmental change Dependency Demographic patterns	Personal hygiene, growth and aging, sexual education and mental health, healthy eating, safety and first aid, medicine safety, substance abuse prevention, health promotion and disease prevention, consumer health, health and environment (pilot-tested)	Multiple choice	Score range: 0-32; Continuous score	na	Self-administered & Performance-based
17	VOHL[56]	***Functional HL*** 1. Health knowledge (2)	Developmental change	Oral health for tooth & gingiva (na)	Visual drawing	Score range: 0-6; Continuous score	na	Self-administered & Performance-based
18	HAS-A[57]	***Functional, interactive and***	Developmental change	General health topics in daily	5-point	Score range: 0-24	na	Self-

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
		critical HL 1. Understanding health information (6) 2. Communication health information (5) 3. Confusion about health information (4)	Demographic patterns Dependency	life (pilot-tested)	Likert scale	(understanding), 0-20 (communication), 0-16 (confusion); Continuous score		administered & Self-reported
19	MaHeLi[58]	***Functional, interactive and critical HL*** 1. Health seeking-behaviour (1) 2. Competence and coping skills (6) 3. Appraisal of health information (5)	Developmental change Demographic patterns Dependency Differential epidemiology	General health and maternal health topics (na)	6-point Likert scale	na	na	Interviewer-administered & Self-reported
20	QuALiSMental[59]	***Functional, interactive and critical HL*** 1. Recognition disorders (14) 2. Knowledge about the professionals and treatments available (16) 3. Knowledge of the effectiveness of self-help strategies (12) 4. Knowledge and skills needed to provide support and first aid to others (10) 5. Knowledge of how to prevent mental disorders (8)	Developmental change Demographic patterns Dependency	Mental health vignettes (na)	Yes/no; Multiple choice	na	40-50 minutes	Self-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
21	FCCHL-AYA[60]	***Functional, interactive and critical HL*** 1. Functional HL (6) 2. Communicative HL(3) 3. Critical HL (4)	Developmental change Dependency Differential epidemiology	Health topics regarding cancer in daily life (2nd grade)	4-point Likert scale	Score range: 13-52; Continuous score	na	Self-administered & Self-reported
22	ICHL[61]	***Interactive and critical HL*** 1. Interactive HL (2) 2. Critical HL (7)	Developmental change Demographic patterns Dependency Differential epidemiology	General health topics in daily life (pilot-tested)	5-point Likert scale	na	40-55 minutes (together with other measures)	Self-administered & Self-reported
23	HELMA[62]	***Functional, interactive and critical HL*** 1. Access (5) 2. Reading (5) 3. Understanding (10) 4. Appraise (5) 5. Use (4) 6. Communication (8) 7. Self-efficacy (4) 8. Numeracy (3)	Developmental change Dependency	General health topics in daily life (pilot-tested)	5-point Likert scale	Score range: 0-100; Ordinal category: 0-50: inadequate; 50.1-66: problematic; 66.1-84: sufficient; 84.1-100: excellent	15 minutes	Self-administered & Self-reported
24	HLSAC[63]	***Functional, interactive and critical HL*** 1. Theoretical knowledge (2) 2. Practical knowledge (2) 3. Individual critical thinking (2) 4. Self-awareness (2) 5. Citizenship (2)	Developmental change Dependency	General health topics in daily life (7th grade)	4-point Likert scale	na	45 minutes (together with the HBSC survey)	Self-administered & Self-reported
25	REALM-TeenS[64]	***Functional HL*** 1. Reading recognition (10)	Developmental change Demographic patterns	10 health-related words such as diabetes (6th grade)	Open-ended	Score range: 0-10; Ordinal category: 0-2: ≤ 3rd; 3-4: 4th-5th; 5-6: 6th-7th; 7-8: 8th-	13.6 seconds (range: 7.8-23.0)	Interviewer-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
26	funHLS-YA[65]	***Functional HL*** 1. Word recognition and comprehension (19)	Developmental change	Diseases and symptoms, nutrition and diet, biology of the human body (na)	Multiple choice	9^{th}; 9-10: $\geq 10^{th}$ Score range: na; Continuous score	5 minutes	Self-administered & Performance-based
27	HLS-TCO[66]	***Functional, interactive and critical HL*** 1. Health knowledge and understanding (10) 2. Accessing information and services (5) 3. Communicating skills (6) 4. Managing health conditions (5) 5. Media literacy (5) 6. Making decisions (4)	Developmental change Demographic patterns Dependency Differential epidemiology	Health information for obesity preventive behaviours (pilot-tested)	Multiple choice; Likert scale	Score range: 0-135; Ordinal category: low: <21 for FHL, <33 for IHL, <27 for CHL; fair: 21-27.99 for FHL, 33-43.99 for IHL, 27-35.99 for CHL; high: 28-35 for FHL, 44-54.9 for IHL, 36-45 for CHL	na	Self-administered & Performance-based & Self-reported
28	HLRS-Y[67]	***Functional, interactive and critical HL*** 1. Knowledge (10) 2. Self-advocacy/support (14) 3. Resiliency (13)	Developmental change Demographic patterns Differential epidemiology	Health information about chronic health conditions (pilot-tested)	4-point Likert scale	Score range: na; Continuous score	15-20 minutes	Self-administered & Self-reported
29	p_HLAT-8[68]	***Functional, interactive and critical HL*** 1. Understanding health information (2) 2. Searching health information (2) 3. Communicating health information (2) Appraising health information (2)	Dependency	General health topics in daily life (pilot-tested)	5-point Likert scale; 4-point Likert scale	Score range: 0-37; Continuous score	na	Self-administered & Self-reported

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HBSC, Health Behaviour in School-aged Children; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HL, Health Literacy; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

Evaluation of methodological quality of included studies

According to the COSMIN checklist, the methodological quality of each instrument as assessed by each study is presented in **Table 3**. Almost all studies (n=28) examined content validity, 24 studies assessed internal consistency and hypotheses testing, 17 studies examined structural validity, eight studies assessed test-retest/inter-rater reliability, two studies assessed cross-cultural validity and only one study assessed responsiveness.

Evaluation of instruments' measurement properties

After the methodological quality assessment of included studies, measurement properties of each health literacy instrument were examined according to Terwee's quality criteria (See **Appendix 5**).[35] The rating results of measurement properties of each instrument are summarised in **Table 4**.

The synthesised evidence for the overall rating of measurement properties

Finally, a synthesis was conducted for the overall rating of measurement properties for each instrument according to '*the best evidence synthesis*' guidelines recommended by the COSMIN checklist developer group.[32] This synthesis result was derived from information presented in **Table 3** and **Table 4**. The overall rating of each measurement property for each health literacy instrument is presented in **Table 5**. In summary, most information (62.9%, 146/232) on measurement properties was unknown due to either poor methodological quality of studies or a lack of information on reporting or assessment.

Table 3. Methodological quality of each study for each measurement property according to the COSMIN checklist

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Fair	na	na
NVS (Driessnack *et al.*, 2014) [54]	Poor	na	na	Poor	na	Poor	na	na
NVS (Warsh *et al.*, 2014) [39]	na	na	na	Poor	na	Fair	na	na
TOFHLA (Chisolm and Buchanan, 2007) [48]	na	na	na	Poor	na	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Fair	na	na
c-sTOFHLAd (Chang *et al.*, 2012) [51]	Fair	Fair	na	Good	Fair	Fair	Fair	na
REALM-Teen (Davis *et al.*, 2006) [41]	Poor	Fair	na	Good	na	Fair	na	na
REALM-Teen (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Poor	na	na
HLAB (Wu *et al.*, 2010) [40]	Fair	Poor	na	Good	na	Fair	na	na
MMAHL (Massey *et al.*, 2013) [44]	Good	na	na	Good	Good	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [49]	Poor	na	na	Good	na	Good	na	na
DNT-39 (Mulvaney *et al.*, 2013) [53]	Fair	na	na	Poor	na	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013) [53]	Fair	na	na	Poor	na	Fair	na	na
eHEALS (Norman *et al.*, 2006) [43]	Fair	Fair	na	Good	Fair	Fair	na	na
CHC Test (Steckelberg *et al.*, 2009) [47]	na	Poor	na	Good	Poor	na	na	na
HKACSS (Schmidt *et al.*, 2010) [46]	Excellent	na	na	Good	na	Good	na	na
HLAT-51 (Harper, 2014) [42]	Poor	na	na	Good	Poor	na	na	na
HLAT-8 (Abel *et al.*, 2014) [45]	Excellent	na	na	Poor	Excellent	Good	na	na
CHLT (Liu *et al.*, 2014) [55]	Fair	na	na	Good	Fair	Fair	na	na
VOHL (Ueno *et al.*, 2014) [56]	na	Fair	na	na	na	Fair	na	Fair
HAS-A (Manganello *et al.* 2015) [57]	Fair	na	na	Good	Fair	Fair	na	na
MaHeLi (Naigaga *et al.* 2015) [58]	Fair	na	na	Poor	Fair	na	na	na
QuALiSMental (de Jesus Loureiro *et al.*, 2015) [59]	Fair	na	na	Excellent	Fair	Fair	na	na
FCCHL-AYAC (McDonald *et al.*, 2016) [60]	Fair	na	na	Good	Fair	Fair	na	na

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive -ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
ICHL (Smith *et al.*, 2016) [61]	na	na	na	Good	na	Fair	na	na
HELMA (Ghanbari *et al.*, 2016) [62]	Good	Good	na	Good	Good	na	na	na
HLSAC (Paakkari *et al.*, 2016) [63]	Fair	Fair	na	Good	Fair	Fair	na	na
REALM-TeenS (Manganello *et al.*, 2017) [64]	Good	na	na	Good	na	Good	na	na
funHLS-YA (Tsubakita *et al.*, 2017) [65]	Fair	na	na	Poor	Fair	Fair	na	na
HLS-TCO (Intarakamhang *et al.*, 2017) [66]	Fair	na	na	Good	Fair	Fair	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017) [67]	Fair	na	na	Excellent	Fair	Fair	na	na
p_HLAT-8 (Quemelo *et al.*, 2017) [68]	Fair	na	na	Good	Fair	Fair	Fair	na

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

Table 4. Evaluation of measurement properties for included instruments according to Terwee's quality criteria

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [50]	-	na	na	?	na	-	na	na
NVS (Driessnack *et al.*, 2014) [54]	+	na	na	?	na	-	na	na
NVS (Warsh *et al.*, 2014) [39]	na	na	na	?	na	+	na	na
TOFHLA (Chisolm and Buchanan, 2007) [48]	na	na	na	?	na	-	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
c-sTOFHLAd (Chang *et al.*, 2012) [51]	+	+	na	+	?	+	?	na
REALM-Teen (Davis *et al.*, 2006) [41]	+	+	na	+	na	+	na	na
REALM-Teen (Hoffman *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
HLAB (Wu *et al.*, 2010) [40]	+	+	na	+	na	-	na	na
MMAHL (Massey *et al.*, 2013) [44]	+	na	na	+	-	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [49]	+	na	na	+	na	+	na	na
DNT-39 (Mulvaney *et al.*, 2013) [53]	+	na	na	?	na	-	na	na
DNT-14 (Mulvaney *et al.*, 2013) [53]	+	na	na	?	na	-	na	na
eHEALS (Norman and Skinner, 2006) [43]	+	-	na	+	+	-	na	na
CHC Test (Steckelberg *et al.*, 2009) [47]	na	+	na	+	+	na	na	na
HKACSS (Schmidt *et al.*, 2010) [46]	+ (HC) - (HA)	na	na	+	na	+	na	na
HLAT-51 (Harper, 2014) [42]	?	na	na	+	?	na	na	na
HLAT-8 (Abel *et al.*, 2014) [45]	-	na	na	?	+	+	na	na
CHLT (Liu *et al.*, 2014) [55]	+	na	na	+	+	+	na	na
VOHL (Ueno *et al.*, 2014) [56]	na	- (TS) + (GS)	na	na	na	-	na	+
HAS-A (Manganello *et al.* 2015) [57]	+	na	na	+	+	-	na	na
MaHeLi (Naigaga *et al.* 2015) [58]	+	na	na	?	+	na	na	na
QuALiSMental (de Jesus Loureiro *et al.*, 2015) [59]	-	na	na	+	+	+	na	na
FCCHL-AYAC (McDonald *et al.*, 2016) [60]	+ (FHL) - (IHL) + (CHL)	na	na	+	+	-	na	na
ICHL (Smith *et al.*, 2016) [61]	na	na	na	+	na	+	na	na
HELMA (Ghanbari *et al.*, 2016) [62]	+	+	na	+	+	na	na	na
HLSAC (Paakkari *et al.*, 2016) [63]	+	+	na	+	-	+	na	na
REALM-TeenS (Manganello *et al.*,	+	na	na	+	na	+	na	na

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
2017) [64]								
funHLS-YA (Tsubakita *et al.*, 2017) [65]	+	na	na	?	+	-	na	na
HLS-TCO (Intarakamhang *et al.*, 2017) [66]	+	na	na	+	+	+	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017) [67]	+	na	na	+	+	+	na	na
p_HLAT-8 (Quemelo *et al.*, 2017) [68]	+	na	na	+	+	-	+	na

Note: na, no information available; + positive rating; ? indeterminate rating; - negative rating. CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; GS, Gingiva Score; HA, Health Attitude; HAS-A, Health Literacy Assessment Scale for Adolescents; HC, Health Communication; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; TS, Tooth Score; VOHL, Visual Oral Health Literacy.

Table 5. The overall quality of measurement properties for each health literacy instrument used in children and adolescents

Health literacy instrument	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsiveness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS [50 54 39]	?	na	na	?	na	±	na	na
TOFHLA [48]	na	na	na	?	na	-	na	na
s-TOFHLA [50]	?	na	na	?	na	-	na	na
c-sTOFHLAd [51]	+	+	na	++	?	+	?	na
REALM-Teen [41 50]	?	+	na	++	na	+	na	na
HLAB [40]	+	?	na	++	na	-	na	na
MMAHL [44]	++	na	na	++	--	na	na	na
MHL [49]	?	na	na	++	na	++	na	na
DNT-39 [53]	+	na	na	?	na	-	na	na
DNT-14 [53]	+	na	na	?	na	-	na	na
eHEALS [43]	+	-	na	++	+	-	na	na
CHC Test [47]	na	?	na	++	?	na	na	na
HKACSS [46]	+++ (HC) --- (HA)	na	na	++	na	++	na	na
HLAT-51 [42]	?	na	na	++	?	na	na	na
HLAT-8 [45]	---	na	na	?	+++	++	na	na
CHLT [55]	+	na	na	++	+	+	na	na
VOHL [56]	na	- (TS) + (GS)	na	na	na	-	na	+
HAS-A [57]	+	na	na	++	+	-	na	na
MaHeLi [58]	+	na	na	?	+	na	na	na
QuALiSMental [59]	-	na	na	+++	+	+	na	na
FCCHL-AYAC [60]	+ (FHL) - (IHL) + (CHL)	na	na	++	+	-	na	na
ICHL [61]	na	na	na	++	na	+	na	na
HELMA [62]	++	++	na	++	++	na	na	na
HLSAC [63]	+	+	na	++	-	+	na	na
REALM-TeenS [64]	++	na	na	++	na	++	na	na
funHLS-YA [65]	+	na	na	?	+	-	na	na
HLS-TCO [66]	+	na	na	++	+	+	na	na
HLRS-Y [67]	+	na	na	+++	+	+	na	na
p_HLAT-8 [68]	+	na	na	++	+	-	+	na

Note: na, no information available; +++ or --- strong evidence and positive/negative result; ++ or -- moderate evidence and positive/negative result; + or - limited evidence and positive/negative

result; ± conflicting evidence; ? unknown, due to poor methodological quality or indeterminate rating of a measurement property. CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; GS, Gingiva Score; HA, Health Attitude; HAS-A, Health Literacy Assessment Scale for Adolescents; HC, Health Communication; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; TS, Tooth Score; VOHL, Visual Oral Health Literacy.

Discussion

Summary of main results

This study identified and examined 29 health literacy instruments used in children and adolescents and exemplified the large variety of methods used. It showed that to date, only half of included health literacy instruments (15/29) measure all three domains (functional, interactive and critical) and that the functional domain is still the focus of attention when measuring health literacy in children and adolescents. Additionally, researchers mainly focus on participant characteristics of developmental change (of cognitive ability), dependency (on parents) and demographic patterns (e.g. racial/ethnic backgrounds), and less so on differential epidemiology (of health and illness). The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. Most information (62.9%) on measurement properties was unknown due to either the poor methodological quality of studies or a lack of reporting or assessment. It is therefore difficult to draw a robust conclusion about which instrument is the best.

Health literacy measurement in children and adolescents

This review found that health literacy measurement in children and adolescents tends to include Nutbeam's three-domain health literacy construct (i.e. functional, interactive and critical), especially in the past five years. However, almost one third of included instruments focused only on the functional domain (n=10). Unlike health literacy research for patients in clinics, health literacy research for children and adolescents (a comparatively healthy population) should be considered from a health promotion perspective,[69] rather than a health care or disease management perspective. Integrating interactive and critical domains into health literacy measurement is aligned with the rationale of emphasising empowerment in health promotion for children and adolescents.[70] The focus of health literacy for this population group should therefore include all three domains and so there is a need for future research to integrate the three domains within health literacy instruments.

Similar to previous findings by Ormshaw *et al.*[10] and Okan *et al.*,[13] this review also revealed that childhood and adolescent health literacy measurement varied by its dimensions, health topics, forms of administration, and by the level to which participant characteristics were considered. There are likely four main reasons for these disparities. First, definitions of health literacy were inconsistent. Some researchers measured general health literacy,[40 45] while others measured eHealth literacy or media health literacy.[43 49] Second, researchers had different research purposes for their studies. Some researchers used what were originally adult instruments to measure adolescent health literacy,[39 48 52] whereas others developed new or adapted instruments.[40-42 53] Third, the research settings affected the measurement process. As clinical settings were busy, short surveys were more appropriate than long surveys.[39 41 44] On the other hand, health literacy in school settings was often measured using long and comprehensive surveys.[40 42 47] Fourth, researchers considered different participant characteristics when measuring health literacy in children and adolescents. For example, some researchers took considerations of students' cognitive development,[40 41 44 46 51] some focused on adolescents' resources and environments (e.g. friends and family contexts, eHealth contexts, media contexts),[43 45 49] and others looked at the effect of different cultural backgrounds and socio-economic status.[40 41 43 44 46 47 49-52] Based on Forrest *et al.*'s 4D model,[14 15] this review showed that most health literacy instruments considered participants' development, dependency and demographic patterns, with only seven instruments considering differential epidemiology.[53 58 60 61 66 67] Although the '4D' model cannot be used to reduce the disparities in health literacy measurement, it does provide an opportunity to identify gaps in current research and assist researchers to consider participants' characteristics comprehensively in future research.

The methodological quality of included studies

This review included a methodological quality assessment of included studies, which was absent from previous reviews on this subject.[10 11] Methodological quality assessment is important because strong conclusions about the measurement properties of instruments can only be drawn from high-quality studies. In this review, the COSMIN checklist was shown to be a useful framework for critically appraising the methodological quality of studies via each measurement property. Findings suggested

that there was wide variation in the methodological quality of studies for all instruments. Poor methodological quality of studies was often seen in the original or adapted health literacy instruments (the NVS, the TOFHLA, the s-TOFHLA, the DNT-39 and the DNT-14) for two main reasons. The first reason was the vague description of the target population involved. This suggested that researchers were less likely to consider an instrument's content validity when using the original, adult instrument for children and/or adolescents. Given that children and adolescents have less well-developed cognitive abilities, in future it is essential to assess whether all items within an instrument are understood. The second reason was a lack of uni-dimensionality analysis for internal consistency. As explained by the COSMIN group,[71] a set of items can be inter-related and multi-dimensional, whereas uni-dimensionality is a prerequisite for a clear interpretation of the internal consistency statistics (Cronbach's alpha). Future research on the use of health literacy instruments therefore needs to assess and report both internal consistency statistics and uni-dimensionality analysis (e.g. factor analysis).

Critical appraisal of measurement properties for included instruments

This review demonstrated that of all instruments reviewed three instruments (the c-sTOFHLAd, the HELMA and the HLSAC) showed satisfactory evidence about internal consistency and test-retest reliability. Based on the synthesised evidence, the HELMA showed moderate evidence and positive results of internal consistency (α=0.93) and test-retest reliability (intraclass correlation coefficient (ICC)=0.93), whereas the HLSAC (α=0.93; standardised stability estimate=0.83) and the c-sTOFHLAd (α=0.85; ICC=0.95) showed limited evidence and positive results. Interestingly, compared to the overall reliability rating of the s-TOFHLA,[50] the c-sTOFHLAd showed better results.[72] The reason for this was probably the different methodological quality of the studies that examined the s-TOFHLA and the c-sTOFHLAd. The c-sTOFHLAd study had fair methodological quality in terms of internal consistency and test-retest reliability, whereas the original s-TOFHLA study had poor methodological quality for internal consistency and unknown information for test-retest reliability. Given the large disparity of rating results between the

original and translated instrument, further evidence is needed to confirm whether the s-TOFHLA has the same or a different reliability within different cultures, thus assisting researchers to understand the generalisability of the s-TOFHLA's reliability results.

Four instruments were found to show satisfactory evidence about both content validity and construct validity (structural validity and hypotheses testing). Construct validity is a fundamental aspect of psychometrics and was examined in this review for two reasons. First, it enables an instrument to be assessed for the extent to which operational variables adequately represent underlying theoretical constructs.[73] Second, the overall rating results of content validity for all included instruments were similar (i.e. unknown or moderate/strong evidence and positive result). The only difference was that the target population was involved or not. Given that all instruments' items reflected the measured construct, in this review, construct validity was determined to be key to examining the overall validity of included instruments. In this context, only the HLAT-8 showed strong evidence and positive result for structural validity (CFI=0.99; TLI=0.97; RMSEA=0.03; SRMR=0.03) and moderate evidence on hypotheses testing (known-group validity results showed differences of health literacy by gender, educational status and health valuation). However, in the original paper,[45] the HLAT-8 was only tested for its known-group validity, not for convergent validity. Examination of convergent validity is important because it assists researchers in understanding the extent to which two examined measures' constructs are theoretically and practically related.[74] Therefore, future research on the convergent validity of the HLAT-8 would be beneficial for complementing that which exists for its construct validity.

Similar to a previous study by Jordan *et al.*,[26] this review demonstrated that only one included study contained evidence of responsiveness. Ueno *et al.*[56] developed a visual oral health literacy instrument and examined responsiveness by comparing changes in health literacy before and after oral health education. Their results showed students' health literacy scores increased significantly after health education. Responsiveness is the ability of an instrument to detect change over time in the construct being measured, and it is particularly important for longitudinal studies.[31] However, most studies included in this review were cross-sectional studies, and only one study (on the

MMAHL[44]) discussed the potential to measure health literacy over time. Studies that measure health literacy over time in populations are needed, not only because this is a prerequisite for longitudinal studies, but also so that the responsiveness of instruments can be monitored and improved.

Feasibility issues for included instruments

This review showed that the feasibility aspects of instruments varied markedly. In relation to forms of administration, this review identified 19 self-administered instruments and 10 interviewer-administered instruments. This suggests that self-administered instruments are more commonly used in practice than interviewer-administered instruments. However, both administration modes have limitations. Self-administered instruments are cost-effective and efficient, but may bring about respondent bias, whereas interviewer-administered instruments, while able to ensure high response rates, are always resource intensive and expensive to administer.[75] Although the literature showed that there was no significant difference in scores outcome between these two administration modes,[76 77] the relevant studies mostly concerned health-related quality of life instruments. It is still unknown whether the same is true for health literacy instruments. Among children and adolescents, health literacy research is more likely to be conducted through large-scale surveys in school settings. Therefore, the more cost-effective, self-administered mode seems to have great potential for future research. To further support the wide use of self-administered instruments, there is a need for future research to confirm the same effect of administration between self-administered and interviewer-administered instruments.

With regard to the type of assessment method, this review revealed that performance-based health literacy instruments (n=15) are more preferable than self-report instruments (n=11). There might be two reasons for this. First, it is due to participant characteristics. Compared with adults, children and adolescents are more dependent on their parents for health-related decisions.[15] Measurement error is more likely to occur when children and adolescents answer self-report items.[78] Therefore, performance-based assessment is often selected to avoid such inaccuracy. Second, performance-based instruments are objective, whereas self-report instruments are

subjective and may bring about over-estimated results.[79] However, the frequent use of performance-based instruments does not mean that they are more appropriate than self-report instruments when measuring childhood and adolescent health literacy. Compared with performance-based instruments, self-report instruments are always time-efficient and help to preserve respondents' dignity.[21] The challenge in using self-report instruments is to consider the readability of tested materials. If children and adolescents can understand what a health literacy instrument measures, then they are more able to accurately self-assess their own health literacy skills.[70] The difference between self-report and performance-based instruments of health literacy has been discussed in the literature,[80] but the evidence about the difference is still limited due to a lack of specifically-designed studies for exploring the difference. Further studies are needed to fill this knowledge gap.

Recommendations for future research

This review identified 18 instruments (the REALM-Teen, the NVS, the s-TOFHLA, the c-sTOFHLAd, the eHEALS, the CHC Test, the HKACSS, the HLAB, the MHL, the HLAT-51, the CHLT, the VOHL, the QuALiSMental, the HELMA, the HLSAC, the funHLS-YA, the HLS-TCO and the p_HLAT-8) that were used to measure health literacy in school settings. Although it is difficult to categorically state which instrument is the best, this review provides useful information that will assist researchers to identify the most suitable instrument to use when measuring health literacy in children and adolescents in school contexts.

Among the 18 instruments, six tested functional health literacy (the REALM-Teen, the NVS, the s-TOFHLA, the c-sTOFHLAd, the VOHL and the funHLS-YA); one examined critical health literacy (the CHC Test); one measured functional and interactive health literacy (the HKACSS); one examined functional and critical health literacy (the HLAB); and nine tested health literacy comprehensively focusing on functional, interactive and critical domains (the eHEALS, the MHL, the HLAT-51, the CHLT, the QuALiSMental, the HELMA, the HLSAC, the HLS-TCO and the p_HLAT-8). However, only one of these three-domain instruments (the HLSAC) was considered appropriate for use in schools because of its quick administration, satisfactory reliability and one-factor validity. Eight three-domain instruments were

excluded due to the fact that they focused on non-general health literacy (the eHEALS, the MHL, the QuALiSMental, the HLS-TCO) or were burdensome to administer (the HLAT-51, the HELMA-44) or were not published in English (the CHLT and the p_HLAT-8).

Compared with the HLSAC, the HLAT-8 examines the construct of health literacy via three domains rather than one-factor structure, thus enabling a more comprehensive examination of the constrcut. Meanwhile, although the p_HLAT-8 (Portuguese version) is not available in English, the original HLAT-8 is. After comparing measurement domains and measurement perperties, the HLAT-8 was deemed to be more suitable for measuring health literacy in school settings for four reasons: (1) it measures health literacy in the context of family and friends,[45] a highly important attribute because children and adolescents often need support for health decisions from parents and peers;[7 15] (2) it is a short but comprehensive tool that captures Nutbeam's three-domain nature of health literacy;[17] (3) it showed satisfactory structural validity (RMSEA=0.03; CFI=0.99; TLI=0.97; SRMR=0.03);[45] and (4) it has good feasibility (e.g. the p_HLAT-8 is self-administered and time-efficient) in school-based studies. However, there are still two main aspects that need to be considered in future. One aspect is its use in the target population. Given the HLAT-8 has not been tested for children and adolescents under 18, its readability and measurement properties need to be evaluated. The other aspect is that its convergent validity (the strength of association between two measures of a similar construct, an essential part of construct validity) has not been examined. Testing convergent validity of the HLAT-8 is important because high convergent validity assists researchers to understand the extent to which two examined measures' constructs are theoretically and practically related.

Limitations

This review was not without limitation. First, we restricted the search to studies aiming to develop or validate a health literacy instrument. Thus we may have missed relevant instruments in studies that were not aiming to develop instruments.[81 82] Second, although the COSMIN checklist provided us with strong evidence of the methodological quality of a study via an assessment of each measurement property, it

cannot evaluate a study's overall methodological quality. Third, criterion validity was not examined due to lack of 'gold standard' for health literacy measurement. However, we examined convergent validity under the domain of 'hypotheses testing'. This can ascertain the validity of newly-developed instruments against existing commonly-used instruments. Finally, individual subjectivity inevitably played a part in the screening, data extraction and synthesis stage of the review. To reduce this subjectivity, two authors independently managed the major stages.

CONCLUSION

This review updated previous reviews of childhood and adolescent health literacy measurement and incorporated a quality assessment framework. It showed that most information on measurement properties was unknown due to either the poor methodological quality of studies or a lack of assessment and reporting. Rigorous and high-quality studies are needed to fill the knowledge gap in relation to health literacy measurement in children and adolescents. Although it is challenging to draw a robust conclusion about which instrument is the best, this review provides important evidence that supports the use of the HLAT-8 to measure childhood and adolescent health literacy in future research.

ACKNOWLEDGMENTS

This paper is part of Shuaijun Guo's PhD research project, which is supported by the Melbourne International Engagement Award. The authors also appreciate the helpful comments received from the reviewers (Martha Driessnack and Debi Bhattacharya) at *BMJ Open*.

CONTRIBUTORS

SG conceived the review approach. RA and EW provided general guidance for the drafting of the protocol. SG and SA screened the literature. SG and TS extracted data. SG drafted the manuscript. SG, GB, RA, EW, XY, SA and TS reviewed and revised the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

None.

DATA SHARING STATEMENT

There are no additional data available.

REFERENCES

1. Nutbeam D. The evolving concept of health literacy. *Soc Sci Med* 2008;67(12):2072-78.
2. Nutbeam D. Health promotion glossary. *Health Promot Int* 1998;13(4):349-64.
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav* 2007;31(Supplement 1):S19-S26.
4. Squiers L, Peinado S, Berkman N, et al. The health literacy skills framework. *J Health Commun* 2012;17(sup3):30-54.
5. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155(2):97-107.
6. Dodson S, Good S, Osborne R. Health literacy toolkit for low-and middle-income countries: a series of information sheets to empower communities and strengthen health systems. New Delhi: World Health Organization, Regional Office for South-East Asia, 2015.
7. Manganello J. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res* 2008;23(5):840-47.
8. Diamond C, Saintonge S, August P, et al. The development of building wellness™, a youth health literacy program. *J Health Commun* 2011;16(sup3):103-18.
9. Robinson LD, Jr., Calmes DP, Bazargan M. The impact of literacy enhancement on asthma-related outcomes among underserved children. *J Natl Med Assoc* 2008;100(8):892-96.
10. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. *Health Educ* 2013;113(5):433-55.
11. Perry EL. Health literacy in adolescents: an integrative review. *J Spec Pediatr Nurs* 2014;19(3):210-18.
12. Bröder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017;17:361.

13. Okan O, Lopes E, Bollweg TM, et al. Generic health literacy measurement instruments for children and adolescents: a systematic review of the literature. *BMC public health* 2018;18(1):166.
14. Forrest CB, Simpson L, Clancy C. Child health services research: challenges and opportunities. *JAMA* 1997;277(22):1787-93.
15. Rothman RL, Yin HS, Mulvaney S, et al. Health literacy and quality: focus on chronic illness care and patient safety. *Pediatrics* 2009;124(Supplement 3):S315-S26.
16. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12(1):80.
17. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15(3):259-67.
18. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? *J Health Commun* 2010;15(S2):9-19.
19. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies? *Int J Public Health* 2009;54(5):303-05.
20. Abel T. Measuring health literacy: moving towards a health-promotion perspective. *Int J Public Health* 2008;53(4):169-70.
21. Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. *J Health Commun* 2014;19(12):1481-96.
22. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. In: Higgins JP, Green S, eds. London, UK: The Cochrane Collaboration, 2011.
23. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151(4):264-69.
24. Simonds SK. Health education as social policy. *Health Educ Monogr* 1974;2(1_suppl):1-10.
25. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. *Nurs Health Sci* 2009;11(1):77-89.
26. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol* 2011;64(4):366-79.
27. Sanders LM, Federico S, Klass P, et al. Literacy and child health: a systematic review. *Arch Pediatr Adolesc Med* 2009;163(2):131-40.
28. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics* 2009;124(Supplement 3):S265-S74.
29. Sawyer S, Sawyer R, Afifi L, et al. Adolescence: a foundation for future health. *Lancet (London, England)* 2012;379(9826):1630-40.
30. Fok TKS, Wong. What does health literacy mean to children? *Contemporary Nurse : a Journal for the Australian Nursing Profession* 2002;13(2-3):249-58.
31. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):651-57.
32. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to

select an outcome measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13.

33. McCormack L, Haun J, Sørensen K, et al. Recommendations for Advancing Health Literacy Measurement. *J Health Commun* 2013;18(sup1):9-14.
34. Pleasant A, McKinney J, Rikard RV. Health Literacy Measurement: A Proposed Research Agenda. *J Health Commun* 2011;16(sup3):11-21.
35. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
36. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490-94.
37. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-26.
38. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):659-70.
39. Warsh J, Chari R, Badaczewski A, et al. Can the newest vital sign be used to assess health literacy in children and adolescents? *Clin Pediatr* 2014;53(2):141-44.
40. Wu A, Begoray D, Macdonald M, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. *Health Promot Int* 2010;25(4):444-52.
41. Davis TC, Wolf MS, Arnold CL, 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(6):e1707-e14.
42. Harper R. Development of a health literacy assessment for young adult college students: A pilot study. *J Am Coll Health* 2014;62(2):125-34.
43. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006;8(4):e27.
44. Massey P, Prelip M, Calimlim B, et al. Findings Toward a Multidimensional Measure of Adolescent Health Literacy. *Am J Health Behav* 2013;37(3):342-50.
45. Abel T, Hofmann K, Ackermann S, et al. Health literacy among young adults: a short survey tool for public health and health promotion research. *Health Promot Int* 2014;30(3):725-35.
46. Schmidt CO, Fahland RA, Franze M, et al. Health-related behaviour, knowledge, attitudes, communication and social status in school children in Eastern Germany. *Health Educ Res* 2010;25(4):542-51.
47. Steckelberg A, Hülfenhaus C, Kasper J, et al. How to measure critical health competences: development and validation of the Critical Health Competence Test (CHC Test). *Adv Health Sci Educ* 2009;14:11-22.
48. Chisolm DJ, Buchanan L. Measuring adolescent functional health literacy: a pilot validation of the test of functional health literacy in adults. *J Adolescent Health* 2007;41(3):312-14.
49. Levin-Zamir D, Lemish D, Gofin R. Media Health Literacy (MHL): development and measurement of the concept among adolescents. *Health Educ Res* 2011;26(2):323-35. doi: 10.1093/her/cyr007

50. Hoffman S, Trout A, Nelson T, et al. A psychometric assessment of health literacy measures among youth in a residential treatment setting. *J Stud Soc Sci* 2013;5(2):288-300.
51. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese version of short-form Test of Functional Health Literacy in Adolescents. *J Clin Nurs* 2012;21(17-18):2429-37. doi: 10.1111/j.1365-2702.2012.04147.x
52. Driessnack M, Chung S, Perkhounkova E, et al. Using the "newest vital sign" to assess health literacy in children. *J Pediatr Health Care* 2014;28(2):165-71.
53. Mulvaney SA, Lilley JS, Cavanaugh KL, et al. Validation of the diabetes numeracy test with adolescents with type 1 diabetes. *J Health Commun* 2013;18(7):795-804.
54. Driessnack M, Chung S, Perkhounkova E, et al. Using the "Newest Vital Sign" to assess health literacy in children. *Journal of Pediatric Health Care* 2014;28(2):165-71.
55. Liu CH, Liao LL, Shih SF, et al. Development and implementation of Taiwan's child health literacy test. *Taiwan Journal of Public Health* 2014;33(3):251-70.
56. Ueno M, Takayama A, Adiatman M, et al. Application of visual oral health literacy instrument in health education for senior high school students. *International Journal of Health Promotion and Education* 2014;52(1):38-46. doi: http://dx.doi.org/10.1080/14635240.2013.845412
57. Manganello JA, DeVellis RF, Davis TC, et al. Development of the Health Literacy Assessment Scale for Adolescents (HAS-A). *Journal of communication in healthcare* 2015;8(3):172-84. doi: 10.1179/1753807615y.0000000016 [published Online First: 2015/01/01]
58. Naigaga MDAS, Pettersen KS. Measuring Maternal Health Literacy in Adolescents Attending Antenatal Care in Uganda: Exploring the Dimensionality of the Health Literacy Concept Studying a Composite Scale. *Journal of nursing measurement* 2015;23(2):E50.
59. de Jesus Loureiro LM. Questionnaire for Assessment of Mental Health Literacy--QuALiSMental: study of psychometric properties. *Revista de Enfermagem Referência* 2015;4(4):79-88. doi: 10.12707/riv14031
60. McDonald FE, Patterson P, Costa DS, et al. Validation of a Health Literacy Measure for Adolescents and Young Adults Diagnosed with Cancer. *Journal of adolescent and young adult oncology* 2016;5(1):69-75. doi: 10.1089/jayao.2014.0043 [published Online First: 2016/01/27]
61. Smith SR, Samar VJ. Dimensions of Deaf/Hard-of-Hearing and Hearing Adolescents' Health Literacy and Health Knowledge. *Journal of health communication* 2016;21:141-54. doi: 10.1080/10810730.2016.1179368
62. Ghanbari S, Ramezankhani A, Montazeri A, et al. Health Literacy Measure for Adolescents (HELMA): Development and Psychometric Properties. *PloS one* 2016;11(2):e0149202. doi: 10.1371/journal.pone.0149202 [published Online First: 2016/02/18]
63. Paakkari O, Torppa M, Kannas L, et al. Subjective health literacy: Development of a brief instrument for school-aged children. *Scandinavian journal of public health* 2016;44(8):751-57. doi: 10.1177/1403494816669639
64. Manganello JA, Colvin KF, Chisolm DJ, et al. Validation of the Rapid Estimate for Adolescent Literacy in Medicine Short Form (REALM-TeenS). *Pediatrics* 2017;139(5) doi: 10.1542/peds.2016-3286 [published Online First: 2017/05/31]

65. Tsubakita T, Kawazoe N, Kasano E. A New Functional Health Literacy Scale for Japanese Young Adults Based on Item Response Theory. *Asia-Pacific journal of public health* 2017;29(2):149-58. doi: 10.1177/1010539517690226 [published Online First: 2017/02/17]
66. Intarakamhang U, Intarakamhang P. Health Literacy Scale and Causal Model of Childhood Overweight. *Journal of research in health sciences* 2017;17(1):e00368.
67. Bradley-Klug K, Shaffer-Hudkins E, Lynn C, et al. Initial development of the Health Literacy and Resiliency Scale: Youth version. *Journal of communication in healthcare* 2017;10(2):100-07. doi: 10.1080/17538068.2017.1308689
68. Quemelo PRV, Milani D, Bento VF, et al. Health literacy: translation and validation of a research instrument on health promotion in Brazil. *Cadernos de saude publica* 2017;33(2):e00179715. doi: 10.1590/0102-311x00179715
69. Kilgour L, Matthews N, Christian P, et al. Health literacy in schools: prioritising health and well-being issues through the curriculum. *Sport Educ Soc* 2015;20(4):485-500.
70. Velardo S, Drummond M. Emphasizing the child in child health literacy research. *J Child Health Care* 2017;21(1):5-13.
71. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist manual. Amsterdam: VU University Medical Centre, 2012.
72. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese version of short-form Test of Functional Health Literacy in Adolescents. *Journal of clinical nursing* 2012;21(17-18):2429-37. doi: 10.1111/j.1365-2702.2012.04147.x
73. Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008;98(1):9-10.
74. Akpa OM, Bamgboye EA, Baiyewu O. The Adolescents' Psychosocial Functioning Inventory (APFI): scale development and initial validation using Exploratory and Confirmatory Factor Analysis. *Afr J Psychol Study Soc Issues* 2015;18(1):1-21.
75. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health* 2005;27(3):281-91.
76. Lozano F, Lobos JM, March JR, et al. Self-administered versus interview-based questionnaires among patients with intermittent claudication: Do they give different results? A cross-sectional study. *Sao Paulo Med J* 2016;134(1):63-69.
77. Dujaili JA, Sulaiman SAS, Awaisu A, et al. Comparability of Interviewer-Administration Versus Self-Administration of the Functional Assessment of Chronic Illness Therapy-Tuberculosis (FACIT-TB) Health-Related Quality of Life Questionnaire in Pulmonary Tuberculosis Patients. *Pulm Ther* 2016;2(1):127-37.
78. Vaz S, Parsons R, Passmore AE, et al. Internal consistency, test–retest reliability and measurement error of the self-report version of the Social Skills Rating System in a sample of Australian adolescents. *PloS one* 2013;8(9):e73924.
79. Altin SV, Finke I, Kautz-Freimuth S, et al. The evolution of health literacy assessment tools: a systematic review. *BMC Public Health* 2014;14:1207.
80. Kiechle ES, Bailey SC, Hedlund LA, et al. Different measures, different outcomes? A systematic review of performance-based versus self-reported

measures of health literacy and numeracy. *J Gen Intern Med* 2015;30(10):1538-46.
81. Paek H-J, Reber BH, Lariscy RW. Roles of interpersonal and media socialization agents in adolescent self-reported health literacy: a health socialization perspective. *Health Educ Res* 2011;26(1):131-49.
82. Hubbard B, Rainey J. Health Literacy Instruction and Evaluation among Secondary School Students. *Am J Health Educ* 2007;38(6):332-37.

FIGURE

Figure 1. Flowchart of search and selection process according to PRISMA flow diagram

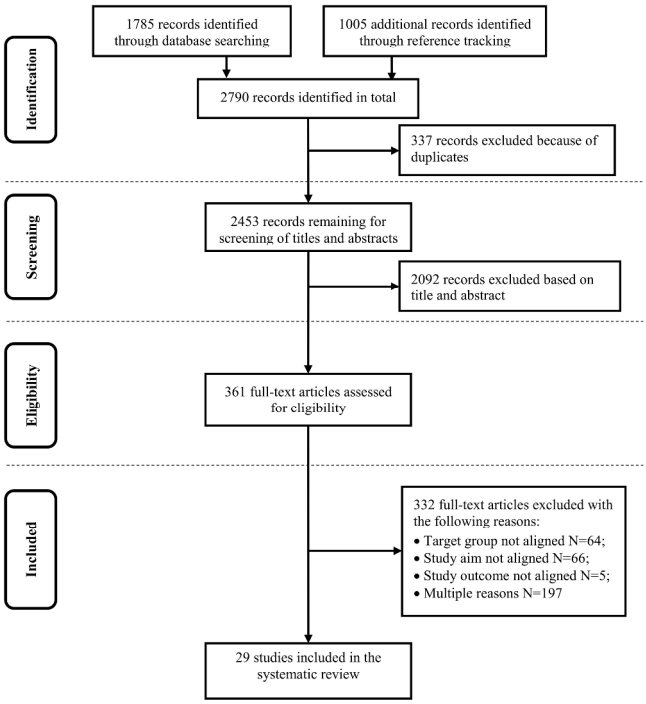

Figure 1. Flowchart of search and selection process according to PRISMA flow diagram

297x420mm (300 x 300 DPI)

Appendix 1. A systematic review protocol

Measuring the Quality of Child and Adolescent Health Literacy Instruments: A Systematic Review

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1], Sheikh M Alif[1], Geoffrey R Browne[1], Xiaoming Yu[2*]

[1] *School of Population and Global Health, The University of Melbourne, Melbourne, Australia*

[2] *Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China*

* Corresponding author email: gshj1986@gmail.com yxm@bjmu.edu.cn

Background

Health literacy research has been a growing interest by researchers across the globe. The term *'health literacy'* was first used in 1974 in the proceedings of a health education conference discussing health education as a social policy issue affecting the healthcare system, mass communication and the education system (1, 2). However, few references were found regarding health literacy in the literature until 1992 (3). Since 1992, health literacy has been broadly studied both in clinical and public health contexts. In clinical settings, health literacy is typically 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*' by the Institute of Medicine (IOM) in America (4). In such circumstances, health literacy is a derivative concept from literacy and numeracy skills, which is often used as a risk factor that needs to be identified and appropriately managed for patients and health professionals (5). Accordingly, health literacy measurement tools and 'screening aids' for clinicians are developed to assess patient literacy levels, and help health professionals to tailor health information for better communication with their patients (6). From the public health perspective, health literacy is defined and accepted by World Health Organization (WHO) as '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*' (7). This understanding of health literacy identifies it as a broad concept, which is seen as a personal asset to enable individuals to take more control over their health and determinants of health (5). With a different understanding of the concept, health literacy measures vary in a different way. Although health literacy measurement varies and is still being debated (1, 8-10), there is

consistent evidence showing health literacy is of potential importance and considered as a public health goal internationally. A recent WHO report pointed out that poor health literacy skills were associated with riskier behaviours, poorer health status, less self-management and longer hospitalization and more health costs (11).

Based on a preliminary search of health literacy, there were more interests in studies focusing on adult health literacy than adolescent health literacy. However, previous research studies suggested that poor health literacy was a prevalent problem in adolescents. In Australia, the 2006 National Health Literacy Survey reported that 67.6% of adolescents aged 15 to 19 years old did not attain the minimum skills required to deal with health information and service in everyday life (12). Compared with adult health literacy, there are several reasons for the potential importance of adolescent health literacy: 1) adolescents are future mainstream and independent healthcare consumers, a health literate person can contribute to less health care costs, better health status compared to that is not health literate (13); 2) adolescents are at a critical stage of development characterised by physical, emotional and cognitive changes, attempting to prepare for independence but lacking the adequate ability of reasoning and decision-making. Therefore, improving their health literacy skills could support sound health decisions in future (14, 15); 3) low health literacy has been demonstrated to associate with high levels of health-risk behaviors (16, 17) and low levels of health-promoting behaviors for adolescents (18); 4) enhancing health literacy through school-based interventions has great potential for improving students' access to and interpretation of health information (19). Adolescents spend most of their daily time in school, which means they can receive health education and learn how to improve healthy lifestyles and related skills through this setting (20, 21).

Health literacy is more challenging to understand for adolescents than that for adults. Researchers may have different understandings and underlying constructs when using the same definition. That is why there are such a large number of measurement tools of health literacy currently (22, 23), along with some newly-developed health literacy instruments (24). According to Mancuso (1), it is recommended to use specific assessment tools for a specific age group in a specific context. Studies measuring childhood and adolescent health literacy have been a research focus, particularly in the past five years (23). Ormshaw *et al.* (23) conducted a systematic review on measuring childhood and adolescent health literacy in 2011. They found 16 studies that were involved with health literacy measures in children and adolescents. The authors also identified 13 health topics and nine underlying components from existing health literacy instruments. However, the authors did not critically appraise health literacy indices explicitly regarding their validity and reliability. More importantly, the authors

did not assess the methodological quality of each included study. This may undermine the persuasiveness of its conclusion. To fill this knowledge gap, we aim to conduct a systematic review that examines studies' methodological quality and examine reliability and validity of each health literacy instrument, thus providing researchers with unbiased information about which instruments have good psychometric properties. The '*COnsensus-based Standards for the selection of health status Measurement INstruments*' (COSMIN) group has recently developed as a critical appraisal tool (a checklist) to evaluate the methodological quality of studies on measurement properties of health measurement instruments (25). These measurement properties are divided into three domains: reliability, validity, and responsiveness (26). According to the COSMIN checklist, it is possible and scientific to critically appraise and compare psychometric properties of health literacy instruments for children and adolescents.

In this protocol, our target population is adolescent. According to the definition of the WHO, adolescents are those people aged 10 to 19 years and young people aged 10-24 years (27, 28). Given that the term '*adolescent*', '*child*', '*youth*' and '*young people*' is closely related, and Erikson (29) reckoned that children between the ages of 6 and 12 years could learn, compete and co-operate with others, we define our target group as those aged 6-24 years old.

Objectives of the review

This review aims to identify which health literacy instruments have good psychometric properties for children and adolescents. Specifically, there are three objectives:

1) To examine the methodological quality of included studies that aim to measure health literacy in children and adolescents;

2) To examine the measurement properties (i.e. reliability; validity; responsiveness) of health literacy instruments in children and adolescents;

3) To compare the overall rating of measurement properties between each health literacy instrument used in children and adolescents.

Search strategy

Database and search terms

As the term '*health literacy*' was first coined in 1974, articles published from 1st,January 1974 to 30th May 2014 in all languages will be searched. Search strategies will be first designed and then be consulted with two librarian experts. Articles indexed in the following seven databases: Medline, Pubmed, Embase, PsycINFO, CINAHL, ERIC and Cochrane Library will be searched. The search key terms are '*health literacy*' and '*assessment*' according to previously published

studies (1, 23, 30, 31). Age group for '*child, adolescent and young adult*' will be defined in the database settings. The synonyms are listed in **Appendix Table 1**. These synonyms are connected by '*or*' and search strategies are completed by '*and*'.

Appendix Table 1 Searching terms in databases

Key term (1)	Key term (2)
health literacy	health literacy measur*
health AND literacy AND education	health literacy assess*
	health literacy evaluat*
	health literacy instrument*
	health literacy tool*

Other sources of literature

Searching other sources to identify relevant research including:

- Reference lists of identified studies;
- Reference lists of previous systematic reviews on health literacy (1, 23, 30-33).

Eligibility criteria for inclusion and exclusion

According to the guidelines recommended by Cochrane Handbook for systematic reviews (34), inclusion criteria will be addressed regarding population, intervention, comparison, outcome and study design (PICOS):

Inclusion criteria-Participants

The target group should be children and/or adolescents, any age from 6 to 24 years of age.

Inclusion criteria-Interventions and Comparators

As interventional studies are not our interest in this review, it is not applicable to set out guidelines for interventions and comparators

Inclusion criteria-Outcomes

The included studies must be involved with health literacy assessment for children and adolescents, that is, the study should specify the term '*health literacy*', and studies are included if they report on at least one or more attributes of the three measurement properties: 1) reliability; 2) validity; and 3) responsiveness.

Inclusion criteria-Study design

The article should be research-based and peer-reviewed paper including study aim, methods,

and results. Also, the study aim should focus on health literacy instrument development or validation.

Exclusion criteria

Studies will be excluded if they are: 1) not focusing on the target group; 2) not focusing on the health literacy instrument development or tool validation; 3) not research-based and peer-reviewed papers including editorials, comments and letters; 4) not reporting findings or results regarding any one of the measurement properties.

Study selection

Search records will be kept including the names of databases searched, keywords, search timeframe, and the search results. All the electronic search results will be initially inputted into the bibliography software of EndNote X7 (Thomson Reuters, New York, NY), and other sources of literature results will be summarised in the print paper. This screening process will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (35). One reviewer will screen studies by titles and abstracts. Secondly, full copies of articles identified will be obtained for thorough screening according to the inclusion criteria by two reviewers independently. Any disagreements in reviewer selections will be resolved at a meeting.

Quality assessment

The methodological quality of each included study will be assessed by two reviewers independently using the COSMIN checklist (25). The checklist consists of nine boxes with 5-18 items concerning methodological standards for how each measurement property should be assessed. Four response options for each item of the COSMIN checklist are defined, representing '*excellent*', '*good*', '*fair*' and '*poor*' quality. An overall score for the methodological quality of a study will be determined for each measurement property separately, by taking the lowest rating of any items in a box ('*worst score counts*') (36). Discrepancies arise between the reviewers will be resolved through discussion, if necessary with a third independent person.

Data extraction

Data extraction will be performed along with the assessment of methodological quality using the COSMIN checklist (25). In addition, information on the interpretability (e.g. norm scores,

floor-ceiling effects, minimal important change of the instruments), generalisability (e.g. characteristics of the study population and sampling procedure), respondent and administrative burden, and forms of administration will be also collected because they are important characteristics of a measurement instrument (26, 37). The data will be entered in an electronic form. Where possible, authors of the original studies will be contacted to obtain essential missing or additional data. Two reviewers will independently extract the data. Consensus should be reached afterward, if necessary with a third independent person.

Data synthesis

The results of the quality of health literacy instruments will be assessed using Terwee's quality criteria (38), to see whether the results of the measurement attributes are '*positive*', '*negative*', or '*indeterminate*'. To summarise the overall ratings of the measurement properties of one health literacy instruments by different authors, the synthesis will be performed by combining the results of the quality of health literacy instruments, the results of methodological quality of health literacy measurement studies and the consistency of their results. The possible overall rating for a measurement property is '*positive*', '*indeterminate*', or '*negative*', accompanied by levels of evidence, similarly as was proposed by the Cochrane Back Review Group (39, 40). One reviewer will perform the data synthesis and a second reviewer will check the synthesised results. Discrepancies of the results will be resolved by discussion.

References

1. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. Nurs Health Sci 2009;11(1):77-89.
2. Simonds SK. Health education as social policy. Health Educ Monogr. 1974;2(1_suppl):1-10.
3. Speros C. Health literacy: concept analysis. J Adv Nurs. 2005;50(6):633-40.
4. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health literacy: a prescription to end confusion. Washington DC, USA: The National Academies Press; 2004.
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med. 2008;67(12):2072-8.
6. Parker RM, Williams MV, Weiss BD, Baker DW, Davis TC, Doak CC, et al. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA. 1999;281(6):552-7.
7. Nutbeam D. Health promotion glossary. Health Promot Int. 1998;13(4):349-64.
8. Mancuso JM. Health literacy: a concept/dimensional analysis. Nurs Health Sci. 2008;10(3):248-55.
9. Higgins JW, Begoray D, MacDonald M. A Social Ecological Conceptual Framework for Understanding Adolescent Health Literacy in the Health Education Classroom. Am J Commun Psychol. 2009;44(3-4):350-62.
10. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int. 2005;20(2):195-203.
11. Kickbusch I, Pelikan JM, Apfel F, Tsouros A. Health literacy: the solid facts. Copenhagen, Denmark: WHO Regional Office for Europe, 2013.
12. Australian Bureau of Statistics. 4233.0-Health Literacy, Australia, 2006. Canberra: Australian Bureau of Statistics; 2008 Jun 25 [cited 2014 Mar 23]; Available from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4233.02006?OpenDocument.
13. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health. 2012;82(1):28-36.
14. Bacon JL. Adolescent sexuality and pregnancy. Current Opinion in Obstetrics and Gynecology.

2000;12(5):345-7.
15. Paus T. Mapping brain maturation and cognitive development during adolescence. Trends Cogn Sci. 2005;9(2):60-8.
16. Miles, SB, Stipek, D. Contemporaneous and Longitudinal Associations Between Social Behavior and Literacy Achievement in a Sample of Low-Income Elementary School Children. Child Dev. 2006;77(1):103-17.
17. Conwell, LS, O'Callaghan, MJ, Andersen, MJ, Bor, W, Najman, JM, Williams, G. Early adolescent smoking and a web of personal and social disadvantage. J Paediatr Child Health. 2003;39(8):580-5.
18. Chang LC. Health literacy, self-reported status and health promoting behaviours for adolescents in Taiwan. J Clin Nurs. 2011;20(1-2):190-6.
19. Gazmararian, JA, Curran, JW, Parker, RM, Bernhardt, JM, DeBuono, BA. Public health literacy in America: an ethical imperative. Am J Prev Med. 2005;28(3):317-22.
20. Brey RA, Clark SE, Wantz MS. Enhancing health literacy through accessing health information, products, and services: An exercise for children and adolescents. J Sch Health. 2007;77(9):640-4.
21. Kickbusch I. Health literacy: an essential skill for the twenty-first century. Health Educ. 2008;108(2):101-4.
22. Wu A, Begoray D, Macdonald M, Wharf Higgins J, Frankish J, Kwan B, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. Health Promot Int. 2010;25(4):444-52.
23. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. Health Educ. 2013;113(5):433-55.
24. National Institutes of Health (NIH). Office of Behavior and Social Sciences Reasearch. Research Underway in Health Literacy Supported by NIH. 2014. Bethesda, Maryland: NIH; 2014 [cited 2014 Mar 3]; Available: https://obssr-archive.od.nih.gov/scientific_areas/social_culture_factors_in_health/health_literacy/research-underway-health.aspx.
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):539-49
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
27. World Health Organization. Health topics: Adolescent health. Geneva, Switzerland: WHO; 2014 [cited 2014 Mar 7]; Available from: http://www.who.int/topics/adolescent_health/en/.
28. World Health Organization Media Centre. Young people: health risks and solutions, Fact sheet N°345. Geneva, Switzerland: WHO Media Centre; 2014 [cited 2014 Mar 7]; Available: http://www.who.int/mediacentre/factsheets/fs345/en/.
29. Erikson EH. Childhood and society. 2nd ed. New York, USA: W. W. Norton & Company; 1963.
30. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol. 2011;64(4):366-79.
31. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155(2):97-107.
32. Sanders LM, Federico S, Klass P, Abrams MA, Dreyer B. Literacy and child health: a systematic review. Arch Pediatr Adolesc Med 2009;163(2):131-40.
33. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics. 2009;124(Supplement 3):S265-S74.
34. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. London, UK: The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org.
35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.
36. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2012;21(4):651-7.
37. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2002;11(3):193-205.
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
39. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine. 2009;34(18):1929-41.
40. Van Tulder, M, Furlan, A, Bombardier, C, Bouter, L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine. 2003;28(12):1290-9.

Appendix 2. Search strategy for seven databases

This section has two parts for SEARCH STRATEGY. The first part focuses on the timeline of 1974 to 2014. The second part focuses on the timeline of May 2014 to Jan 2018.

Part 1:

1 MEDLINE (Web of Science) search strategy

MEDLINE database was searched using the Web of Science interface on 16/05/2014 for the period 1974 to 2014.

Basic search:

Set	Results	
# 1	500	MeSH HEADING: (health literacy) OR ((TITLE: (health literacy) OR MeSH HEADING:exp: (Health Literacy)) AND (TITLE: (education) OR MeSH HEADING:exp: (Educational Status) OR MeSH HEADINGS:exp: (/education) OR MeSH HEADING:exp: (Teaching) OR MeSH HEADING:exp: (Educational Status) OR MeSH HEADING:exp: (Education))) Refined by: MeSH HEADINGS: (ADOLESCENT OR YOUNG ADULT OR CHILD) Indexes=MEDLINE Timespan=1974-2014
# 2	3,880	TOPIC: ((((health) literacy assess* OR health literacy measur*) OR health literacy evaluat*) OR health literacy instrument*) OR health literacy tool*) Indexes=MEDLINE Timespan=1974-2014
# 3	352	#2 AND #1 Indexes=MEDLINE Timespan=1974-2014

2 PubMed search strategy

PubMed database was searched (Advanced search) on 16/05/2014 for the period 1974 to 16/05/2014.

Set	Results	
# 1	4910	Search (health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract]) Sort by: PublicationDate
# 2	3248385	Search (child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract]) Sort by: PublicationDate Because if we select age group including child, adolescent, and young adult, the newest papers such as published in 2014 will not be included, the reason maybe the database doesn't update properly. So we use these terms to identify.
# 3	1887	Search (health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*) Sort by: PublicationDate
# 4	581	Search ((((health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract]))) AND ((health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*))) AND ((child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract])) Filters: Publication date from 1974/01/01 to 2014/05/16 Sort by: PublicationDate

3 EMBASE (Ovid) search strategy

EMBASE database was searched using Ovid interface on 16/05/2014 for the period 1974 to current.

Using **.mp** as searching terms (Advanced Search):

Set	Results	
#1	6060	("health literacy" or (health and literacy and education)).mp.
#2	6043	limit 1 to yr="1974 -Current"
#3	671	limit 2 to (school child <7 to 12 years> or adolescent <13 to 17 years>)
#4	170	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	170	limit 4 to yr="1974 -Current"
#6	18	3 and 5

4 PsycINFO (EBSCO) search strategy

PsycINFO database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	786	health literacy OR (health AND literacy AND education)	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#2	133	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#3	133	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

5 CINAHL (EBSCO) search strategy

CINAHL database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	437	health literacy OR (health AND education AND literacy)	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#2	63	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#3	63	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

6 ERIC (EBSCO) search strategy

ERIC database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	59	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#2	2,250	health literacy OR (health AND education AND literacy)	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#3	59	S1 AND S2	Search modes - Boolean/Phrase

7 The Cochrane Library search strategy

The Cochrane Library database was searched on 30/05/2014 for the period January 1974 to May 2014.

Set	Results	Sub-database
#1	4	Cochrane Reviews: There are 4 results from 8483 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Cochrane Reviews'
#2	114	Trials: There are 114 results from 789657 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Trials'
#3	2	Methods Studies: There are 2 results from 15764 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Methods Studies'
#4	120	

Part 2:

The above seven databases were searched using similar rationale as describe before for the timeframe of May 17 2014 to Jan 31 2018.

MEDLINE was searched using the Web of Science interface on 17/02/2018 for the period 2014 to 2018.

Basic search:

Set	Results	
# 5	35	#4 AND #3 Indexes=MEDLINE Timespan=2014-2018
# 4	14,198	MeSH MAJOR TOPIC:exp: ((((((child*) OR adolescent*) OR student*) OR youth) OR young people) OR teen*) OR young adult) Indexes=MEDLINE Timespan=2014-2018

# 3	1,779	#2 AND #1 Indexes=MEDLINE Timespan=2014-2018
# 2	3,482	((((TOPIC: (health literacy assess*) OR TOPIC: (health literacy measur*)) OR TOPIC: (health literacy instrument*)) OR TOPIC: (health literacy tool*)) ORTOPIC: (health literacy evaluat*)) Indexes=MEDLINE Timespan=2014-2018
# 1	2,654	((MeSH HEADING:exp: (health literacy) OR MeSH MAJOR TOPIC:exp: (health literacy)) OR TITLE: (health literacy)) OR MeSH MAJOR TOPIC: ((health) AND education) AND literacy) Indexes=MEDLINE Timespan=2014-2018

Pubmed was searched (Advanced search) on 17/02/2018 for the period 2014 to 31/01/2018.

Set	Results	
#6	26	Search (((((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract])))) AND ((child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract])) Filters:Publication date from 2014/05/16 to 2018/01/31
#5	48	Search (((((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract])))) AND ((child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract]))
#4	288	Search (((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract]))
#3	288	Search (health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract])
#2	1636528	Search (child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract])
#1	8495	Search (((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract])

EMBASE was searched using Ovid interface on 17/02/2018 for the period 2014 to current.

Using **.mp** as searching terms (Basic Search):

Set	Results	
#1	11966	("health literacy" or (health and literacy and education)).mp.
#2	5862	limit 1 to yr="2014 -Current"
#3	639	limit 2 to (school child <7 to 12 years> or adolescent <13 to 17 years>)
#4	372	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	**26**	3 and 4

PsycINFO was searched using Ovid interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results	
#1	4331	("health literacy" or (health and literacy and education)).mp.
#2	2077	limit 1 to yr="2014 -Current"
#3	754	limit 2 to (100 childhood <birth to age 12 yrs> or 180 school age <age 6 to 12 yrs> or 200 adolescence <age 13 to 17 yrs> or 320 young adulthood <age 18 to 29 yrs>)
#4	216	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	**40**	3 and 4

CINAHL was searched using EBSCO interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results		
S1	health literacy OR ((health AND education AND literacy))	Limiters - Published Date: 20140501-20180131; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase	View Results (467)
S2	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 20140501-20180131; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase	View Results (118)
S3	S1 AND S2	Search modes - Boolean/Phrase	View Results (118)

ERIC was searched using EBSCO interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results		
S1	health literacy OR ((health AND education AND literacy))	Limiters - Date Published: 20140501-20180131 Search modes - Boolean/Phrase	View Results (292)
S2	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Date Published: 20140501-20180131 Search modes - Boolean/Phrase	View Results (13)
S3	(S1 AND S2)	Search modes - Boolean/Phrase	View Results (13)

Cochrane Library was searched on 17/02/2018 for the period May 2014 to Jan 2018.

Set	Results	Sub-database
#1	2	Cochrane Reviews: There are 2 results from 10210 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Year from 2014 to 2018 in Cochrane Reviews'
#2	199	Trials: There are 199 results from 1121096 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Year from 2014 to 2018 in Trials'
#3	201	

Appendix 3. Quality criteria for measurement properties of health literacy instruments

Property	Rating	Quality criteria
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	-	MIC ≤ SDC OR MIC equals or inside LOA
Reliability	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80
	?	Neither ICC/weighted Kappa nor Pearson's r determined
	-	ICC/weighted Kappa < 0.70 OR Pearson's r < 0.80
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement
	-	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity		
Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	-	Factors explain < 50% of the variance
Hypotheses testing	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Responsiveness		
Responsiveness	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs

Note: AUC, Area Under the Curve; ICC, Intra-class Correlation Coefficient; LOA, Limits of Agreement; MIC, Minimal Important Change; SDC, Smallest Detectable Change. + positive rating; ? indeterminate rating; - negative rating.

Appendix 4. Levels of evidence for the overall rating of measurement properties

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodological quality

Note: + positive result; - negative result; ±conflicting result; ? unknown result.

Appendix 5. Reliability and validity results for included instruments

Appendix Table 1. The methodological quality of each study based on reliability for each health literacy instrument

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
NVS (Warsh *et al.*, 2014)	na	na	na	na	na	na
NVS (Driessnack *et al.*, 2014)	α=0.71 (n=47)	Poor	na	na	na	na
NVS (Hoffman *et al.*, 2013)	α=0.67 (n=229)	Poor	na	na	na	na
c-sTOFHLAd (Chang *et al.*, 2012)	α=0.85 (n=300) Item-total correlation=0.44-0.86	Fair	Correlation of test and retest was 0.95 (P<0.001)	Test-retest	1 week	Fair
TOFHLA (Chisolm and Buchanan, 2007)	na	na	na	na	na	na
s-TOFHLA (Hoffman *et al.*, 2013)	α=0.89 (n=229)	Poor	na	na	na	na
REALM-Teen (Davis *et al.*, 2006)	α=0.94 (n=388)	Poor	γ=0.98	Test-retest	1 week	Fair
REALM-Teen (Hoffman *et al.*, 2013)	α=0.92 (n=229)	Poor	na	na	na	
HLAB (Wu *et al.*, 2010)	α=0.92 (n=275) Understanding α=0.88 (n=275) Evaluating α=0.82 (n=275)	Fair	Concordance rate=95%	Inter-rater	na	Poor
MMAHL(Massey *et al.*, 2013)	α=0.83 (n=1208) Item-total correlation=0.39-0.74	Good	na	na	na	na
MHL (Levin-Zamir *et al.*, 2011)	α=0.74 (n=1316) Coefficient of reproducibility=0.84 Coefficients of scalability=0.54-0.80	Poor	na	na	na	na
DNT-39 (Mulvaney *et al.*, 2013)	α=0.93 (n=61)	Fair	na	na	na	na
DNT-14 (Mulvaney *et al.*, 2013)	α=0.82 (n=133)	Fair	na	na	na	na

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
	α=0.80 (n=61) α=0.83 (n=72)					
eHEALS (Norman and Skinner, 2006)	α=0.88 (n=664) Item-scale correlation coefficient=0.51-0.76	Fair	The correlations between administrations ranged 0.68-0.40.	Test-retest	Immediately after the intervention; 3-month; 6-month	Fair
CHC Test (Steckelberg *et al.*, 2009)	na	na	Cohen's Kappa was excellent for 277 ratings (κ=0.9-1.0), moderate or good for 31 ratings (κ=0.7-0.89) and poor for 5 ratings (κ=<0.7)	Inter-rater	na	Poor
HKACSS (Schmidt *et al.*, 2010)	Health knowledge χ^2=6.45, *P*=0.17 (n=852) Health communication α=0.73 (n=852) Health attitudes α=0.57 (n=852)	Excellent	na	na	na	na
HLAT-51 (Harper, 2014)	Goodness of fit statistic was calculated by each domain (CFI=0.33-0.88; TLI=0.66-0.84; RMSEA=0.09-0.17). The internal consistency statistic was not calculated.	Poor	na	na	na	na
HLAT-8 (Abel *et al.*, 2014)	α=0.64 (n=7097 for male) α=0.65 (n=331 for female)	Excellent	na	na	na	na
CHLT (Liu *et al.*, 2014)	α=0.87 (the entire scale); subscales α ranged 0.59 to 0.81	Fair	na	na	na	na
VOHL (Ueno *et al.*, 2014)	na	na	The kappa value of scoring among the dentists ranged from 0.60 tooth score to 0.70 for gingiva score.	Inter-rater	na	Fair
HAS-A (Manganello *et al.*, 2015)	α=0.77 (communication) α=0.73 (confusion) α=0.76 (understanding)	Fair	na	na	na	na
MaHeLi (Naigaga *et al.* 2015)	The person separation index for the original 20-item scale	Fair	na	na	na	na

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
	was 0.91 and α=0.92. After item reduction, the person separation index for 12-item scale was 0.90.					
QuALiSMental (de Jesus Loureiro *et al.*, 2015)	α=0.55-0.72 (component 2 and 3) α=0.44-0.59 (component 4) α=0.60-0.82 (component 5)	Fair	na	na	na	na
FCCHL-AYAC (McDonald *et al.*, 2016)	α=0.73 (FHL) α=0.63 (IHL) α=0.85 (CHL)	Fair	na	na	na	na
ICHL (Smith *et al.*, 2016)	na	na	na	na	na	na
HELMA (Ghanbari *et al.*, 2016)	α=0.93 (the entire scale); subscales α ranged 0.61 to 0.89	Good	The intraclass correlation coefficient was 0.93.	Test-retest	Two weeks	Good
HLSAC (Paakkari *et al.*, 2016)	α=0.93 (the entire scale); subscales α ranged 0.69 to 0.77	Fair	The standardised stability estimate was 0.83.	Test-retest	Two weeks	Fair
REALM-TeenS (Manganello *et al.*, 2017)	α=0.82	Good	na	na	na	na
funHLS-YA (Tsubakita *et al.*, 2017)	α=0.75	Fair	na	na	na	na
HLS-TCO (Intarakamhang *et al.*, 2017)	α=0.70-0.82 for five subscales; KR-20=0.76 for health knowledge scale	Fair	na	na	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017)	α=0.88 (Knowledge) α=0.94 (Self-advocacy/ support) α=0.93 (Resiliency)	Fair	na	na	na	na
p_HLAT-8 (Quemelo *et al.*, 2017)	α=0.74 (the entire scale), subscales α ranged 0.41 to 0.71	Fair	na	na	na	na

Note: na, no information available. CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; FCCHL-

AYAC, the Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, the Health Literacy Assessment Scale for Adolescents; HELMA, the Health Literacy Measure for Adolescents; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, The Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, the Maternal Health Literacy; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, the Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, the Rapid Estimate of Adolescent Literacy in Medicine Short Form; RMSEA, Root Mean Square Error of Approximation; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults; VOHL, the Visual Oral Health Literacy.

Appendix Table 2. The methodological quality of each study based on validity for each health literacy instrument

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
NVS (Warsh *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in item generation.	Poor	na	na	Hypotheses regarding correlation between scores of a comparator instrument of Gray Silent Reading Test (GSRT) and NVS were formulated before data collection. The NVS and GSRT scores were highly correlated (ρ=0.71, p<0.0001). The NVS score increased with child age (ρ=0.53, p<0.0001).	Fair	na	na
NVS (Driessnack *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in item generation.	Poor	na	na	A moderate positive correlation was found between children's NVS scores and their age, and between children's NVS scores and their reports of books numbers (γ_s=0.43, p=0.003; γ_s=0.36, p=0.012, respectively), but not found with their parents' report of the number of children's books at home (γ_s=0.06, p=0.671).	Poor	na	na
NVS (Hoffman *et al.*, 2013)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in item generation.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.49 (p<0.01).	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
c-sTOFHLAd (Chang *et al.*, 2012)	The c-sTOFHLAd was translated from the short-version of TOFHLA according to translation procedures and was tested among 30 adolescents to ensure appropriateness.	Good	Confirmatory factor analysis was conducted to determine structural validity. One-factor model indicated an acceptable fit to the data according structural equation modelling analysis.	Fair	Convergent validity was measured between c-sTOFHLAd and the rapid estimate of adult literacy in medicine (REALM), with a correlation coefficient of 0.74 (p<0.001).	Fair	Semantic equivalence was measured by the content validity index (CVI). All items were rated by the experts as having a CVI>0.85. Thirty adolescents were chosen to determine and ensure the cultural congruence of the instrument.	Fair
TOFHLA (Chisolm and Buchanan, 2007)	The TOFHLA was developed from a literacy expert after reviewing commonly used hospital texts and a pilot test. No target population is involved in item generation.	Poor	na	na	The reading comprehension component was significantly correlated with the WRAT3 and the REALM (ρ=0.59, p<0.001; ρ=0.60, p<0.001 respectively), however, no correlation were found with the numeracy component (ρ=0.11, p=0.45; ρ=0.18, p=0.22 respectively).	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013)	The s-TOFHLA was developed based on previous data analysis, perceived importance and frequency of the task in the healthcare settings.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.28 (p<0.01).	Fair	na	na
REALM-Teen (Davis *et al.*, 2006)	The REALM-Teen was developed based on a preliminary test and a structured interview among adolescents. And a panel of experts reviewed the word	Good	na	na	Convergent validity was measured between REALM-Teen and the WRAT-3 (r=0.83) and SORT-R (r=0.93).	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	list.							
REALM-Teen (Hoffman *et al.*, 2013)	The REALM-Teen was developed based on a preliminary test and structured interview among adolescents. And a panel of experts reviewed the word list.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.40 (p<0.01).	Poor	na	na
HLAB (Wu *et al.*, 2010)	Previous experience and literature review were used to develop items; 10 students were pilot-tested for appropriateness of wording, content and format of the final instrument.	Good	na	na	Correlations were assumed between socio-demographic variables and the overall scores. Socio-demographics of gender, age when came to Canada to live, speaking a language other than English were correlated with the scores of HLAB (β=-0.18, p=0.004; β=-0.22, p=0.014; β=-0.20, p=0.008 respectively). No convergent validity is assessed.	Fair	na	na
MMAHL (Massey *et al.*, 2013)	Domains were established from literature review and focus group. Items were developed either using adaptation of existing relevant items or created by the research team.	Good	Explorative principal components factor analysis was conducted and 49.8% of the variance was accounted by 6 factors.	Good	na	na	na	na
MHL (Levin-Zamir *et al.*, 2011)	The face validity was discussed in the focus group during pilot test. The content validity was analysed using theory and operational definitions of	Good	na	na	As hypothesised, MHL was associated with socio-economic determinants, particularly with gender (β=1.25, p<0.001) and mother's education (β=0.16, p=0.04). In addition, MHL was	Good	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	health literacy and media literacy, and adolescents were invited to write detailed, anonymous responses.				also associated with health behaviours (β=0.03, *p*=0.05) and health empowerment (β=0.36, *p*<0.001).			
DNT-39 (Mulvaney *et al.*, 2013)	The DNT-39 was developed from the original 43-item version DNT-43 by eliminating questions specific to type 2 diabetes. An expert team developed the DNT-43 and refined it.	Poor	na	na	The DNT-39 was associated with WRAT-3 and parent education (ρ=0.40, *p*=0.001; ρ=0.29, *p*=0.028 respectively)	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013)	The DNT-14 was developed from the original 15-item version DNT-15 by eliminating 1 question specific to type 2 diabetes. An expert team developed the DNT-15 by data analysis from DNT-43.	Poor	na	na	The DNT-14 was associated with the Wide-Ranging Achievement Test (WRAT3), parent education, diabetes problem solving and HbA1c (ρ=0.36, *p*=0.005; ρ=0.31, *p*=0.019; ρ=0.27, *p*=0.023; ρ=-0.34, *p*=0.004 respectively)	Fair	na	na
eHEALS (Norman and Skinner, 2006)	The eHEALS was developed by the expert team and pilot-tested and refined by feedback from participants.	Good	Explorative principal components factor analysis was conducted and 56% of the variance was accounted by a single factor. The factor loadings ranged from 0.60-0.84 among the 8 items.	Fair	Correlations were assumed between eHEALS and other measured variables (gender, age, use of information technology overall, self-evaluations of health). However, only gender difference was found at baseline level of eHealth literacy (t=2.236, *p*=0.026). No convergent validity is assessed.	Fair	na	na
CHC Test (Steckelberg	The CHC Test was developed by the research	Good	IRT test for determining	Poor	na	na	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
et al., 2009)	team and pre-tested by collecting qualitative data and quantitative field test.		dimensionality was performed.					
HKACSS (Schmidt *et al.*, 2010)	The HKACSS items were taken from a previous health survey and selected basing on consideration of item content.	Good	na	na	As hypothesised, health communication, attitudes and self-efficacy were significantly related to each other (ρ=0.15-0.38, P<0.05). And children from higher educational background showed a better knowledge and communicated more about health topics (β=0.16, p<0.05).	Good	na	na
HLAT-51 (Harper, 2014)	The expert team evaluated the initial items using a 5-point Likert scale according to their research experience. And 144 college students were invited to complete a pilot test.	Good	Comprehension (CFI=0.80; TLI=0.78; RMSEA=0.09); health numeracy (CFI=0.57; TLI=0.48; RMSEA=0.09); media literacy (CFI=0.88; TLI=0.84; RMSEA=0.07); digital literacy (CFI=0.33; TLI=0.06; RMSEA=0.16); health information seeking (CFI=0.80; TLI=0.66; RMSEA=0.17)	Poor	na	na	na	na
HLAT-8 (Abel *et al.*, 2014)	The research team developed the HALT-8 drawing on literature review and their own	Poor	Explorative principal components factor analysis was conducted and 72.96% of the	Excellent	Hypotheses were formulated a priori regarding correlations between health literacy and gender, socio-cultural	Good	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	experience. No target population is involved in item generation.		variance was accounted by four factors among male. In addition, the factor structure was validated using confirmatory factor analysis (CFI=0.99; TLI=0.97; RMSEA=0.03; SRMR=0.03).		characteristics and health values. Results showed that female, higher educational status, and a stronger health valuation were associated with higher HL scores (p<0.05, respectively).			
CHLT (Liu *et al.*, 2014)	The research team developed the CHLT drawing on literature review, expert consultation and pilot test. 12 six graders were piloted about the instrument's readability.	Good	Confirmatory factor analysis was conducted to test the uni-dimensionality of each subscale. The factor loadings ranged from 0.20-0.58.	Fair	Hypotheses were formulated a priori regarding correlations between health literacy and gender, self-reported health and health behaviours. Results showed that female, better health status, normal BMI and healthy behaviours were positively associated with HL scores (p<0.05, respectively). Health-risky behaviours were negatively associated with health literacy scores (p<0.05).	Fair	na	na
VOHL (Ueno *et al.*, 2014)	na	na	na	na	Correlations were conducted between health literacy and gender. Results showed female students had higher gingiva scores than male students (p<0.05). However, no gender differences were found regarding tooth scores.	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
HAS-A (Manganello *et al.*, 2015)	The research team developed the HAS-A drawing on literature review, expert consultation and pilot test. Scale items were piloted with undergraduates.	Good	Exploratory factor analysis was conducted and 41% of the variance was accounted by three factors.	Fair	Communication scale, confusion scale, and understanding scale were all correlated with the AURA scale (r=0.69, p<0.001; r=-0.50, p<0.001; r=-0.42, p<0.001). The correlation between communication scale, confusion scale and understanding scale and REALM-Teen and NVS were small, ranging from -0.26 to 0.08. Also health literacy scores were compared by demographics. There was no difference in scores by sex or age, but a significant difference by race/ethnicity (p<0.001).	Fair	na	na
MaHeLi (Naigaga *et al.* 2015)	The research team developed the MaHeLi based on the health belief model and integrated model of health literacy. No target population is involved in item generation.	Poor	The health-seeking behaviour (HSB) subscale brought substantial multidimensionality into the MeHeLi scale. After removing most items of the HSB subscale, the MeHeLi scale showed a uni-dimensionality construct with some but not too noticeable multi-dimensionality.	Fair	na	na	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
QuALiSMental (de Jesus Loureiro *et al.*, 2015)	The questionnaire was developed based on mental health literacy framework and adapted among Portuguese adolescents and young people.	Excellent	Exploratory factor analysis was conducted for each component of the questionnaire. A five-factor solution explained 46.84% of the total variance for component 1 and 40.00% for components 2 and 3. A three-factor solution explained 47.24% of the variance for component 4 and a two-factor solution explained 55.63% for component 5.	Fair	The relationship between mental health components and mental health help-seeking intension was examined using a binary logistic regression analysis. Results showed higher levels of mental health literacy tended to associate with mental health-seeking intentions.	Fair	na	na
FCCHL-AYAC (McDonald *et al.*, 2016)	The instrument was adapted from the functional, communicative, and critical health literacy scale to be suitable for adolescents and young adults diagnosed with cancer.	Good	Exploratory factor analysis was conducted for the entire scale. The screen plot suggested the extraction of three factors (53.1% variance explained)	Fair	Health literacy scores were examined by gender, whether the measure was completed online or on paper, whether the participant was on or off treatment. Results showed no significant difference was found.	Fair	na	na
ICHL (Smith *et al.*, 2016)	The instrument was developed from formative interviews with 20 deaf/hard-of hearing high school students. Also the instrument was piloted with 18 individuals including content-expert and content-	Good	na	na	The relationship between ICHL and standard health literacy measures were examined. Result showed most ICHL items were related to health literacy skills instrument-short form, s-TOFHLA, and comprehensive heart disease knowledge	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	naïve deaf and hearing colleagues, teachers interpreters and students.				questionnaire (p<0.05).			
HELMA (Ghanbari *et al.*, 2016)	All items were initially generated by in-depth interviews with 67 adolescents. Then items were assessed by an expert panel review and 16 adolescents.	Good	Exploratory factor analysis was conducted and 53.37% of the variance was accounted by eight factors.	Good	na	na	na	na
HLSAC (Paakkari *et al.*, 2016)	The research team developed the HLSAC drawing on literature review, expert review and pilot test. Scale items were piloted with 401 pupils (7th graders and 9th graders).	Good	The five-factor structure was tested using confirmatory factor analysis (RMSEA=0.08; CFI=0.96; TLI=0.92; SRMR=0.03). However, due to high correlations between factors, one-factor structure was finally determined (RMSEA=0.08; CFI=0.94; TLI=0.92; SRMR=0.04).	Fair	Correlations were assumed between the final 10-item scale and the original 15-item scale. Results showed the 10-item HLSAC predicted approximately 97% of the variance of the 15-item instrument.	Fair	na	na
REALM-TeenS (Manganello *et al.*, 2017)	This instrument was derived from the original 66-item REALM-Teen using the item response theory. Also, ten teenage patients were piloted.	Good	na	na	The REALM-TeenS scores were correlated with the REALM-Teen (r=0.92, p<0.001). Item fit analysis using the differential item functioning showed the REALM-TeenS functioned well for different groups of sex,	Good	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
					race/ethnicity, and language spoken at home.			
funHLS-YA (Tsubakita *et al.*, 2017)	Items were generated from health materials that were frequently used in young adults and reviewed by the research team. No target population was involved in pilot test.	Poor	1-factor model was supported by the exploratory factor analysis.	Fair	The correlation between funHLS-YA and the comparator instrument of functional health literacy was 0.191 (p<0.001).	Fair	na	na
HLS-TCO (Intarakamhang *et al.*, 2017)	Items were developed from theories, documents and related research. Also, focus group and expert review were used to develop the instrument. 100 samples of overweight children were piloted.	Good	Confirmatory factor analysis was conducted for each subscale and results showed the model was acceptable, with factor loading ranging 0.39-0.73.	Fair	The path model of health literacy for obesity prevention behaviours was conducted using structural equation modelling. Results showed the hypothetical causal model was consistent with empirical data (chi-square=60.10, p=0.00, df=12, RMSEA=0.05, CFI=0.99; AGFI=0.99).	Fair	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017)	Items were generated by focus group, expert review and a pilot test with 25 participants.	Excellent	Exploratory factor analysis was conducted, and results showed a three-factor structure of the instrument.	Fair	The relationships between health literacy scores and demographics were examined and results showed insurance type and knowledge, time since diagnosis and knowledge and self-advocacy.	Fair	na	na
p_HLAT-8 (Quemelo *et al.*, 2017)	The p_HLAT-8 was translated from the HLAT-8 according to translation procedures and was tested among 10 university	Good	Confirmatory factor analysis was conducted, and results showed the 4-factor model fit was fair	Fair	Convergent validity was examined for each sub-scale, but the results showed that only the factor 'search for information' was adequate. Discriminant	Fair	Three experts in the field of health forward and backward translated the scale independently. Ten university students	Fair

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	students to ensure appropriateness.		(CFI=0.97, GFI=0.98, TLI=0.95, RMSEA=0.03).		validity was only adequate for two factors ('search for information' and 'understanding information').		were piloted to test and ensure the cultural congruence of the scale. Confirmatory factor analysis showed a 4-factor structure fit the model.	

Note: na, no information available. AGFI, Adjusted Goodness of Fit Index; AURA, Ask, Understand, Remember and Assessment; CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; FCCHL-AYAC, the Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, the Health Literacy Assessment Scale for Adolescents; HELMA, the Health Literacy Measure for Adolescents; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, The Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, the Maternal Health Literacy; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, the Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, the Rapid Estimate of Adolescent Literacy in Medicine Short Form; RMSEA, Root Mean Square Error of Approximation; SRMR, Standardized Root Mean Square Residual; SORT-R, Slosson Oral Reading Test-Revised; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults; VOHL, the Visual Oral Health Literacy; WRAT-3, Wide-Range Achievement Test-Revised.

Appendix Table 3. The methodological quality of each study based on responsiveness for each health literacy instrument

Instrument	Responsiveness	
	Results	COSMIN score
VOHL (Ueno *et al.*, 2014)	Comparison of health literacy scores before and after health education showed both tooth and gingiva scores significantly increased after health education.	Fair

Note: As there was only one study examining the instrument's responsiveness, we only presented the instrument of VOHL. VOHL, the Visual Oral Health Literacy.

Research Checklist. PRISMA checklist for reporting systematic review

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Appendix 1 (CRD42018013759)
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8-9
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	10

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	10
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	11; Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11; 15-16; Table 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	25; Table 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	25; Table 4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	25; Table 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	25; Table 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	32-38
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38-39
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	39
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

The Quality of Health Literacy Instruments used in Children and Adolescents: A Systematic Review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020080.R2
Article Type:	Research
Date Submitted by the Author:	25-Apr-2018
Complete List of Authors:	Guo, Shuaijun; University of Melbourne School of Population and Global Health, Armstrong, Rebecca; University of Melbourne, Waters, E; University of Melbourne, School of Population and Global Health Sathish, Thirunavukkarasu; University of Melbourne School of Population and Global Health, Centre for Health Equity; Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Alif, Sheikh; University of Melbourne, School of Population and Global Health; Monash University, School of Public Health and Preventive Medicine Browne, Geoff; University of Melbourne, School of Population and Global Health Yu, Xiaoming; Peking University, Institute of Child and Adolescent Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Global health
Keywords:	Measurement properties, Health literacy, Children, Adolescents, Systematic review

The Quality of Health Literacy Instruments used in Children and Adolescents: A Systematic Review

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1,2], Sheikh M Alif[1,3], Geoffrey R Browne[1], Xiaoming Yu[4*]

[1] School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[2] Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

[3] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[4] Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China

**** Corresponding Author***

Shuaijun Guo Postal address: 526-9, Level 5, 207 Bouverie Street, Carlton, Victoria, Australia 3010 Email: gshj1986@gmail.com Tel: +61 452 110 331

Xiaoming Yu Postal address: Room 209, Institute of Child and Adolescent Health, School of Public Health, 38 Xueyuan Road, Haidian, Beijing, China 100083 Email: yxm@bjmu.edu.cn Tel: +86 10 8280 2631

Word count: 5536

ABSTRACT

Objective: Improving health literacy at an early age is crucial to personal health and development. Although health literacy in children and adolescents has gained momentum in the past decade, it remains an under-researched area, particularly health literacy measurement. This study aimed to examine the quality of health literacy instruments used in children and adolescents and to identify the best instrument for field use.

Design: Systematic review.

Setting: A wide range of settings including schools, clinics and communities.

Participants: Children and/or adolescents aged 6-24 years.

Primary and secondary outcome measures: Measurement properties (reliability, validity and responsiveness) and other important characteristics (e.g. health topics, components or scoring systems) of health literacy instruments.

Results: There were 29 health literacy instruments identified from the screening process. When measuring health literacy in children and adolescents, researchers mainly focus on the functional domain (basic skills in reading and writing) and consider participant characteristics of developmental change (of cognitive ability), dependency (on parents) and demographic patterns (e.g. racial/ethnic backgrounds), less on differential epidemiology (of health and illness). The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. More than half (62.9%) of measurement properties were unknown, due to either poor methodological quality of included studies or a lack of reporting or assessment. The 8-item Health Literacy Assessment Tool (HLAT-8) showed best evidence on construct validity and the Health Literacy Measure for Adolescents showed best evidence on reliability.

Conclusions: More rigorous and high-quality studies are needed to fill the knowledge gap in measurement properties of health literacy instruments. Although it is

challenging to draw a robust conclusion about which instrument is the most reliable and the most valid, this review provides important evidence that supports the use of the HLAT-8 to measure childhood and adolescent health literacy in future school-based research.

Keywords: measurement properties; health literacy; children; adolescents; systematic review

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The COSMIN checklist was used as a methodological framework to rate the methodological quality of included studies.
- This review has updated previous three reviews of childhood and adolescent health literacy measurement tools and identified 19 additional new health literacy instruments.
- Including only studies that aimed to develop or validate a health literacy instrument may eliminate studies that used a health literacy instrument for other purposes.
- Individual subjectivity exists in the screening and data synthesis stages.

INTRODUCTION

Health literacy is a personal resource that enables an individual to make decisions for healthcare, disease prevention and health promotion in everyday life.[1] As defined by the World Health Organisation,[2] health literacy refers to '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*.' The literature has shown that health literacy is an independent and more direct predictor of health outcomes than socio-demographics.[3 4] People with low health literacy are likely to have worse health-compromising behaviours, higher healthcare costs and poorer health status.[5] Given the close relationship between health literacy and health outcomes, many countries have adopted health literacy promotion as a key strategy to reduce health inequities.[6]

From a health promotion perspective, improving health literacy at an early age is crucial to childhood and adolescent health and development.[7] As demonstrated by Diamond *et al.*[8] and Robinson *et al.*,[9] health literacy interventions for children and adolescents can bring about improvements in healthy behaviours and decreased used of emergency department services. Although health literacy in young people has gained increasing attention, with a rapidly growing number of publications in the past decade,[10-13] childhood and adolescent health literacy is still under-researched. According to Forrest *et al.*'s 4D model,[14 15] health literacy in children and adolescents is mediated by four additional factors compared to adults: (1) *developmental* change: children and adolescents have less well-developed cognitive ability than adults; (2) *dependency*: children and adolescents depend more on their parents and peers than adults do; (3) *differential* epidemiology: children and adolescents experience a unique pattern of health, illness and disability; and (4) *demographic* patterns: many children and adolescents living in poverty or in single-parent families are neglected and so require additional care. These four differences pose significant challenges for researchers when measuring health literacy in children and adolescents.

Health literacy is a broad and multi-dimensional concept with varying definitions.[16] This paper uses the definition by Nutbeam who states that health literacy consists of three domains: functional, interactive and critical.[17] The *functional* domain refers to

basic skills in reading and writing health information, which are important for functioning effectively in everyday life. The *interactive* domain represents advanced skills that allow individuals to extract health information and derive meaning from different forms of communication. And the *critical* domain represents more advanced skills that can be used to critically evaluate health information and take control over health determinants.[17] Although health literacy is sufficiently explained in terms of its definitions[17-19] and theoretical models,[4 7] its measurement remains a contested issue. There are two possible reasons for this. One reason is the large variety of health literacy definitions and conceptual models,[12 16] and the other reason is that researchers may have different study aims, populations and contexts when measuring health literacy.[20 21]

Currently, there are three systematic reviews describing and analysing the methodology and measurement of childhood and adolescent health literacy.[10 11 13] In 2013, Ormshaw *et al.*[10] conducted a systematic review of child and adolescent health literacy measures. This review used four questions to explore health literacy measurement in children and adolescents: "*What measurement tools were used? What health topics were involved? What components were identified?* and *Did studies achieve their stated aims?*" The authors identified 16 empirical studies, with only six of them evaluating health literacy measurement as their primary aim. The remaining studies used health literacy measures as either a comparison tool when developing other new instruments or as a dependent variable to examine the effect of an intervention program. Subsequently, in 2014, Perry[11] conducted an integrative review of health literacy instruments used in adolescents. In accordance with the eligibility criteria, five instruments were identified. More recently, Okan *et al.*[13] conducted another systematic review on generic health literacy instruments used for children and adolescents with the aim of identifying and assessing relevant instruments for first-time use. They found fifteen generic health literacy instruments used for this target group.

Although these three reviews provide general knowledge about the methodology and measurement of health literacy in young people, they all have limitations. Ormshaw *et al.*[10] did not evaluate measurement properties of each health literacy instrument. Although Perry[11] and Okan *et al.*[13] summarised measurement properties of each

instrument, the information provided was limited, mostly descriptive, and lacked a critical appraisal. Notably, none of the three reviews considered the methodological quality of included studies[10 11 13]. A lack of quality assessment of studies raises concerns about the utility of such reviews for evaluating and selecting health literacy instruments for children and adolescents. Therefore, it is still unclear which instrument is the best in terms of its validity, reliability and feasibility for field use. In addition, it is also unclear how Nutbeam's three-domain health literacy model and Forrest *et al.*'s 4D model are considered in existing health literacy instruments for children and adolescents.

To fill these knowledge gaps, this systematic review aimed to examine the quality of health literacy instruments used in the young population and to identify the best instrument for field use. We expect the findings will assist researchers in identifying and selecting the most appropriate instrument for different purposes when measuring childhood and adolescent health literacy.

METHODS

Following the methods for conducting systematic reviews outlined in the Cochrane Handbook,[22] we developed a review protocol (See **Appendix 1**, PROSPERO registered ID: CRD42018013759) prior to commencing the study. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement[23] (See **Research Checklist**) was used to ensure the reporting quality of this review.

Literature search

The review took place over two time periods: The initial systematic review covered the period between 1 January 1974 and 16 May 2014 (period 1). The start date of 1974 was chosen because this was the date from which the term '*health literacy*' was first used.[24] A second search was used to update the review in February 2018. It covered the period 17 May 2014 to 31 Jan 2018 (period 2). The databases searched were: Medline, PubMed, Embase, PsycINFO, CINAHL, ERIC and the Cochrane Library. The search strategy was designed on the basis of previous reviews[5 10 25 26] and in consultation with two librarian experts. Three types of search terms were used: (1)

construct-related terms: '*health literacy*' OR '*health and education and literacy*'; (2) outcome-related terms: '*health literacy assess**' OR '*health literacy measure**' OR '*health literacy evaluat**' OR '*health literacy instrument**' OR '*health literacy tool**'; and (3) age-related terms: '*child**' OR '*adolescent**' OR '*student**' OR '*youth*' OR '*young people*' OR '*teen**' OR '*young adult.*'

No language restriction was applied. The detailed search strategy for each database is available in **Appendix 2**. As per the PRISMA flow diagram,[23] the references from included studies and from six previously published systematic reviews on health literacy[5 10 25-28] were also included.

Eligibility criteria

Studies had to fulfil the following criteria to be included: (1) the stated aim of the study was to develop or validate a health literacy instrument; (2) participants were children or adolescents aged 6 to 24. This broad age range was used because the age range for '*children*' (under the age of 18) and '*adolescents*' (aged 10 to 24) overlap[29] and also because children aged over 6 are able to learn and develop their own health literacy[30]; (3) the term '*health literacy*' was explicitly defined, although studies assessing health numeracy (the ability to understand and use numbers in healthcare settings) were also considered; and (4) at least one measurement property (reliability, validity and responsiveness) was reported in the outcomes.

Studies were excluded if: (a) the full paper was not available (i.e. only a conference abstract or protocol was available); (b) they were not peer-reviewed (e.g. dissertations, government reports); or (c) they were qualitative studies.

Selection process

All references were imported into EndNote X7 software (Thomson Reuters, New York, NY) and duplicate records were initially removed before screening. Next, one author (GS) screened all studies based on title and abstract. Full-text papers of the remaining titles and abstracts were then obtained separately for each review round (period 1 and period 2). All papers were screened by two independent authors (GS

and SA). At each major step of this systematic review, discrepancies between authors were resolved through discussion.

Data extraction

The data that were extracted from papers were: characteristics of included studies (e.g. first author, published year and country), general characteristics of instruments (e.g. health topics, components and scoring systems), methodological quality of the study (e.g. internal consistency, reliability and measurement error) and ratings of measurement properties of included instruments (e.g. internal consistency, reliability and measurement error). Data extraction from full-text papers published during period 1 was performed by two independent authors (GS and TS), whereas data extraction from full-text papers published during period 2 was conducted by one author (GS) and then checked by a second author (TS).

Methodological quality assessment of included studies

The methodological quality of included studies was assessed using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist.[31] The COSMIN checklist is a critical appraisal tool containing standards for evaluating the methodological quality of studies on measurement properties of health measurement instruments.[32] Specifically, nine measurement properties (internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity and responsiveness) were assessed.[32] Since there is no agreed-upon '*gold standard*' for health literacy measurement,[33 34] criterion validity was not assessed in this review. Each measurement property section contains 5 to 18 evaluating items. For example, '*internal consistency*' is evaluated against 11 items. Each item is scored using a four-point scoring system ('*excellent*', '*good*', '*fair*' or '*poor*'). The overall methodological quality of a study is obtained for each measurement property separately, by taking the lowest rating of any item in that section (i.e. '*worst score counts*'). Two authors (GS and TS) independently assessed the methodological quality of included studies published during period 1, whereas the quality of included

studies published during period 2 was assessed by one author (GS) and then checked by another (TS).

Evaluation of measurement properties for included instruments

The quality of each measurement property of an instrument was evaluated using quality criteria proposed by Terwee *et al.*[35], who are members of the group that developed the COSMIN checklist (See **Appendix 3**). Each measurement property was given a rating result ('+' positive, '-' negative, '*?*' indeterminate and '*na*' no information available).

Best evidence synthesis: levels of evidence

As recommended by the COSMIN checklist developer group,[32] '*a best evidence synthesis*' was used to synthesise all the evidence on measurement properties of different instruments. The procedure used was similar to the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework[36], a transparent approach to rating quality of evidence that is often used in reviews of clinical trials.[37] Given that this review did not target clinical trials, the GRADE framework adapted by the COSMIN group was used.[38] Under this procedure, the possible overall rating for a measurement property is '*positive*', '*negative*', '*conflicting*' or '*unknown*', accompanied by levels of evidence ('*strong*', '*moderate*' or '*limited*') (See **Appendix 4**). Three steps were taken to obtain the overall rating for a measurement property. First, the methodological quality of a study on each measurement property was assessed using the COSMIN checklist. Measurement properties from '*poor*' methodological quality studies did not contribute to '*the best evidence synthesis*'. Second, the quality of each measurement property of an instrument was evaluated using Terwee's quality criteria.[35] Third, the rating results of measurement properties in different studies on the same instrument were examined whether consistent or not. This best evidence synthesis was performed by one author (GS) and then checked by a second author (TS).

Patient and Public Involvement

Children and adolescents were not involved in setting the research question, the outcome measures, or the design or implementation of this study.

Results

The initial search identified 2790 studies. After duplicates and initial title/abstract screening, 361full-text articles were identified and obtained. As per the eligibility criteria, 29 studies were included,[39-53] yielding 29 unique health literacy instruments used in children and adolescents (See **Figure 1**).

Characteristics of included studies

Of the 29 studies identified, 25 were published between 2010 and 2017 (See **Table 1**). Most included studies were conducted in Western countries (n=20), with eleven studies carried out in the USA. The target population (aged 7 to 25) could be roughly classified into three subgroups: children aged 7 to 12 (n=5), adolescents aged 13 to 17 (n=20) and young adults aged 18 to 25 (n=4). Schools (n=17) were the most common recruitment settings, compared to clinical settings (n=8) and communities (n=4).

Table 1. Characteristics of included studies

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
1	Davis *et al.*[41] (2006)	USA	Adolescents aged 10-19 years (mean age=14.8±1.9)	REALM-Teen	1533 (47.4)	na	Middle schools, high schools, paediatric primary care clinic and summer programs
2	Norman and Skinner[43] (2006)	Canada	Adolescents aged 13-21 years (mean age=14.95±1.24)	eHEALS	664 (55.7)	Sampling from one arm of a randomized controlled trial	Secondary schools
3	Chisolm and Buchanan[48] (2007)	USA	Young people aged 13-17 years (mean age=14.7)	TOFHLA	50 (48.0)	na	Children's hospital
4	Steckelberg *et al.*[47] (2009)	Germany	Students in Grade 10-11 and university	CHC Test	Sample 1: 322 (36.6) Sample 2: 107 (32.7)	na	Secondary schools, university
5	Schmidt *et al.*[46] (2010)	Germany	Children aged 9-13 years (mean age=10.4)	HKACSS	852 (52.9)	na	Primary school
6	Wu *et al.*[40] (2010)	Canada	Students in Grade 8-12	HLAB	275 (48.0)	Convenience sampling	Secondary schools
7	Levin-Zamir *et al.*[49] (2011)	Israel	Adolescents in Grade 7, 9, 11 (approximately age 13, 15 and 17)	MHL	1316 (52.0)	Probability sampling and random cluster sampling	Public schools
8	Chang *et al.*[51] (2012)	Taiwan	Students in high school (mean age=16.01±1.02)	c-sTOFHLAd	300 (52.6)	Multiple-stage stratified random sampling	High schools
9	Hoffman *et al.*[50] (2013)	USA	Youth aged 14-19 years (mean age=17)	REALM-Teen; NVS; s-TOFHLA	229 (61.6)	na	Private high school
10	Massey *et al.*[44] (2013)	USA	Adolescents aged 13-17 years (mean age=14.8)	MMAHL	1208 (37.6)	Sampling from a large health insurance network	Publicly health insurance network
11	Mulvaney *et al.*[53] (2013)	USA	Adolescents aged 12-17 years (Sample 1: mean age=13.92; Sample 2: mean age=15.10)	DNT-39 and DNT-14	Sample 1: 61 (52.5) Sample 2: 72 (55.6)	na	Diabetes clinics

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
12	Abel *et al.*[45] (2014)	Switzerland	Young adults aged 18-25 years (male mean age: 19.6; female mean age=18.8)	HLAT-8	7428 (95.5)	Sampling from compulsory military service for males and two-stage random sampling for females	Compulsory military service, communities
13	Driessnack *et al.*[54] (2014)	USA	Children aged 7-12 years	NVS	47 (53.0)	Convenience sampling	The science centre
14	Harper[42] (2014)	New Zealand	Students aged 18-24 years	HLAT-51	144 (41.0)	Purposeful sampling	College
15	Warsh *et al.*[39] (2014)	USA	Children aged 7-17 years (median age=11)	NVS	97 (46.0)	Convenience sampling	Paediatric clinics
16	Liu et al.[55] (2014)	Taiwan	Children in grade six	CHLT	162609 (51.1)	National sampling	Primary schools
17	Ueno *et al.*[56] (2014)	Japan	Students in high school Grade 1 (age range: 15-16 years)	VOHL	162 (46.3)	Convenience sampling	A senior high school
18	Manganello *et al.*[57] (2015)	USA	Youth aged 12-19 years (mean age=15.6)	HAS-A	272 (37.0)	Convenience sampling	A paediatric clinic and the community
19	Naigaga *et al.*[58] (2015)	Uganda	Pregnant adolescents aged 15-19 years	MaHeLi	384 (0)	Random sampling	Health centres
20	de Jesus Loureiro *et al.*[59] (2015)	Portugal	Adolescents and young people aged 14-24 years (mean age=16.75±1.62)	QuALiSMental	4938 (43.3)	Multi-stage cluster random sampling	Schools
21	McDonald *et al.*[60] (2016)	Australia	Adolescents and young adults diagnosed with cancer (age range: 12-24 years)	FCCHL-AYAC	105 (33.3)	Sampling from a support organisation	An organisation for young people living with cancer
22	Smith *et al.*[61] (2016)	USA	Deaf/hard-of hearing and hearing adolescents in high school (mean age=17.0±0.84 and 15.8±1.1)	ICHL	Sample 1: 154 (53.2) Sample 2: 89 (33.0)	Convenience sampling	Medical centre summer programs
23	Ghanbari *et al.*[62] (2016)	Iran	Adolescents aged 15-18 years (mean age=16.2±1.03)	HELMA	582 (48.8)	Multi-stage sampling	High schools
24	Paakkari *et al.*[63] (2016)	Finland	Pupils (7th graders aged 13 years: n=1918; 9th graders aged 15 years: n=1935)	HLSAC	3853 (na)	Cluster sampling	Secondary schools
25	Manganello *et*	USA	Adolescents aged 14-19 years	REALM-TeenS	174 (na)	na	Adolescent medicine

Study no	Author (Year)	Country	Target population	Health literacy instrument	Sample size (% male)	Sampling method	Recruitment setting
	al.[64] (2017)		(mean age=16.6)				clinics
26	Tsubakita *et al.*[65] (2017)	Japan	Young adults aged 18-26 years (mean age=19.65±1.34)	funHLS-YA	1751 (76.8)	Convenience sampling	A private university
27	Intarakamhang *et al.*[66] (2017)	Thailand	Overweight children aged 9-14 years	HLS-TCO	2000 (na)	Quota-stratified random sampling	Schools
28	Bradley-Klug *et al.*[67] (2017)	USA	Youth and young adults with chronic health conditions aged 13-21 years (mean age=17.6)	HLRS-Y	204 (24.3)	National sampling	Community-based agencies and social media outlets
29	Quemelo *et al.*[68] (2017)	Brazil	University students (mean age=22.7±5.3)	p_HLAT-8	472 (33.9)	na	A university

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

General characteristics of included instruments

Compared to previous systematic reviews,[10 11 13] this review identified 19 additional new health literacy instruments (eHEALS, s-TOFHLA, DNT-39, DNT-14, HLAT-51, HLAT-8, CHLT, VOHL, HAS-A, QuALiSMental, FCCHL-AYAC, ICHL, HELMA, HLSAC, REALM-TeenS, funHLS-YA, HLS-TCO, HLRS-Y and p_HLAT-8). The 29 health literacy instruments were classified into three groups based on whether the instrument was developed bespoke for the study or not (See **Table 2**).[10] The three groups were: (1) newly-developed instruments for childhood, adolescent and youth health literacy (n=20);[40-47 49 50 55-58 61-63 65-67] (2) adapted instruments that were based on previous instruments for adult/adolescent health literacy (n=6);[51 53 59 60 64 68] and (3) original instruments that were developed for adult health literacy (n=3).[39 48 50 52]

Health literacy domains and components

Next, Nutbeam's three-domain health literacy model[17] was used to classify the 29 instruments according to which of the commonly-used components of health literacy were included. Results showed that ten instruments measured only functional health literacy[39 41 48 50-53 56 64 65] and one instrument measured only critical health literacy.[47] There was one instrument measuring functional and interactive health literacy[46], one measuring functional and critical health literacy[40], and one measuring interactive and critical health literacy.[61] Fifteen instruments measured health literacy by all three domains (functional, interactive and critical).[42-45 49 55 57-60 62 63 66-68]

Consideration of participants' characteristics

As per Forrest *et al.*'s 4D model,[14 15] the 29 included instruments were examined for whether participant characteristics were considered when developing a new instrument or validating an existing instrument. Results showed most of the health literacy instruments considered developmental change, dependency and demographic patterns. In contrast, only seven instruments considered differential epidemiology.[53 58 60 61 66 67]

Health topics, contents and readability levels

Health literacy instruments for children and adolescents covered a range of health topics such as nutrition and sexual health. Most instruments (n=26) measured health literacy in healthcare settings or health promotion contexts (e.g. general health topics, oral health, or mental health), while only three instruments measured health literacy in the specific context of eHealth or media health.[42 43 49] In relation to the readability of tested materials, only eight health literacy instruments reported their readability levels, ranging from 2th to 19.5th grade.

Burden and forms of administration

The time to administer was reported in seven instruments, ranging from 3 to 90 minutes. There were three forms of administration: self-administered instruments (n=19), interviewer-administered instruments (n=9), and video-assisted, interviewer-administered instruments (n=1). Regarding the method of assessment, fifteen instruments were performance-based, eleven instruments were self-report, and three included both performance-based and self-report items.

Table 2. General and important characteristics of included instruments used in children and adolescents

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
1	NVS[50 54 39]	***Functional HL*** 1. Reading comprehension (2) 2. Numeracy (4)	Demographic patterns	Nutrition-related information about the label of an ice cream container (na)	Open-ended	Score range: 0-6; Ordinal category: 0-1: high likelihood of limited literacy; 2-3: possibility of limited literacy; 4-6: adequate literacy	No longer than 3 minutes	Interviewer-administered & Performance-based
2	TOFHLA[48]	***Functional HL*** 1. Reading comprehension (50) 2. Numeracy (17)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4.3 grade), a standard informed consent form (10.4 grade), patients' rights and responsibilities section of a Medicaid application form (19.5 grade), actual hospital forms & labelled prescription vials (9.4 grade)	4 response options	Score range: 0-100; Ordinal category: 0-59: inadequate health literacy; 60-74: marginal health literacy; 75-100: adequate health literacy	12.9 minutes (8.9-17.3 minutes)	Interviewer-administered & Performance-based
3	s-TOFHLA[50]	***Functional HL*** 1. Reading comprehension (36)	Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score range: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	na	Interviewer-administered & Performance-based
4	c-sTOFHLAd[51]	***Functional HL*** 1. Reading comprehension (36)	Developmental change Demographic patterns	Instruction for preparation for an upper gastrointestinal series (4th grade), patients' rights and responsibilities section of a Medicaid application form (10th grade)	4 response options	Score range: 0-36; Ordinal category: 0-16: inadequate literacy; 17-22: marginal literacy; 23-36: adequate literacy	20-minute class period	Self-administered & Performance-based
5	REALM-Teen[41 50]	***Functional HL*** 1. Reading recognition (66)	Developmental change Demographic patterns	66 health-related words such as weight, prescription and tetanus (6th grade)	Open-ended	Score range: 0-66; Ordinal category: 0-37: ≤ 3rd; 38-47: 4th-	2-3 minutes	Interviewer-administered & Performance-

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
						5th; 48-58: 6th-7th; 59-62: 8th-9th; 63-66: ≥ 10th		based
6	HLAB[40]	***Functional and critical HL*** 1. Understanding health information (30) 2. Evaluating health information (17)	Developmental change Demographic patterns	A range of topics such as nutrition and sexual health (pilot-tested)	Open-ended	Score range: 0-107; Continuous score	Two regular classroom sessions	Self-Administered & Performance-based
7	MMAHL[44]	***Functional, interactive and critical HL*** 1. Patient-provider encounter (4) 2. Interaction with the health care system (5) 3. Rights and responsibilities (7) 4. Confidence in using health information from personal source (3) 5. Confidence in using health information from media source (3) 6. Health information seeking competency using the Internet (2)	Developmental change Demographic patterns Dependency	Experiences of how to access, navigate and manage one's health care and preventive health needs (6th grade)	5-point Likert scale	Score range: na; Continuous score	na	Self-administered & Self-reported
8	MHL[49]	***Functional, interactive and critical HL*** 1. Content identification (6) 2. Perceived influence on behaviour (6) 3. Critical analysis and intended (6)	Dependency Demographic patterns	Nutrition/dieting, physical activity, body image, sexual activity, cigarette smoking, alcohol consumption, violent behaviour, safety habits and/or friendship and family connectedness (pilot-tested)	Open-ended & multiple choice	Score range: 0-24; Continuous score	na	Video-assisted interviewer-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
		4. Action/reaction (6)						
9	DNT-39[53]	***Functional health literacy*** 1. Health numeracy (39)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score range: 0-100; Continuous score	na	Interviewer-administered & Performance-based
10	DNT-14[53]	***Functional health literacy*** 1. Health numeracy (14)	Differential epidemiology Demographic patterns	Nutrition, exercise, blood glucose monitoring and insulin administration (na)	Open-ended	Score range: 0-100; Continuous score	na	Interviewer-administered & Performance-based
11	eHEALS[43]	***Functional, interactive and critical HL*** 1. Accessing health information (4) 2. Evaluating health information (2) 3. Applying health information (2)	Developmental change Dependency Demographic patterns	General health topics about online health information (pilot-tested)	5-point Likert scale	Score range: na; Continuous score	na	Self-Administered & Self-reported
12	CHC Test[47]	***Critical HL*** 1. Understanding medical concepts (15) 2. Searching literature skills (22) 3. Basic statistics (18) 4. Design of experiments and sampling (17)	Developmental change Demographic patterns	Echinacea and common cold, magnetic resonance imaging in knee injuries, treatment of acne, breast cancer screening (pilot-tested)	Open-ended & multiple choice	na	Less than 90 minutes	Interviewer-administered & Performance-based
13	HKACSS[46]	***Functional and interactive HL*** 1. Health knowledge (3) 2. Health attitudes (4) 3. Health communication (3) 4. Self-efficacy (3)	Developmental change Dependency Demographic patterns	Physical activities, nutrition, smoking, vaccination, tooth health and general health (na)	2 response options; 5-point Likert scale; 4-point Likert	Score range: na; Continuous score	na	Self-Administered & Performance-based & Self-reported

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
					scale			
14	HLAT-51[42]	***Functional, interactive and critical HL*** 1. Comprehension skill (20) 2. Health numeracy (11) 3. Media literacy (8) 4. Digital literacy (12)	Developmental change Dependency Demographic patterns	Health topics such as gout and uric acid, high cholesterol and triglyceride levels, health-information-seeking skills (na)	Yes/no; multiple choice	na	30-45 minutes	Self-administered & Performance-based & Self-reported
15	HLAT-8[45]	***Functional, interactive and critical HL*** 1. Understanding health information (2) 2. Finding health information (2) 3. Interactive health literacy (2) 4. Critical health literacy (2)	Dependency Demographic patterns	General health topics in people's daily life (na)	5-point Likert scale; 4-point Likert scale	Score range: 0-37; Continuous score	na	Self-administered & Self-reported
16	CHLT[55]	***Functional, interactive and critical HL*** 1. Health knowledge (11) 2. Health attitude (16) 3. Health skills (5)	Developmental change Dependency Demographic patterns	Personal hygiene, growth and aging, sexual education and mental health, healthy eating, safety and first aid, medicine safety, substance abuse prevention, health promotion and disease prevention, consumer health, health and environment (pilot-tested)	Multiple choice	Score range: 0-32; Continuous score	na	Self-administered & Performance-based
17	VOHL[56]	***Functional HL*** 1. Health knowledge (2)	Developmental change	Oral health for tooth & gingiva (na)	Visual drawing	Score range: 0-6; Continuous score	na	Self-administered & Performance-based
18	HAS-A[57]	***Functional, interactive and***	Developmental change	General health topics in daily	5-point	Score range: 0-24	na	Self-

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
		critical HL 1. Understanding health information (6) 2. Communication health information (5) 3. Confusion about health information (4)	Demographic patterns Dependency	life (pilot-tested)	Likert scale	(understanding), 0-20 (communication), 0-16 (confusion); Continuous score		administered & Self-reported
19	MaHeLi[58]	***Functional, interactive and critical HL*** 1. Health seeking-behaviour (1) 2. Competence and coping skills (6) 3. Appraisal of health information (5)	Developmental change Demographic patterns Dependency Differential epidemiology	General health and maternal health topics (na)	6-point Likert scale	na	na	Interviewer-administered & Self-reported
20	QuALiSMental[59]	***Functional, interactive and critical HL*** 1. Recognition disorders (14) 2. Knowledge about the professionals and treatments available (16) 3. Knowledge of the effectiveness of self-help strategies (12) 4. Knowledge and skills needed to provide support and first aid to others (10) 5. Knowledge of how to prevent mental disorders (8)	Developmental change Demographic patterns Dependency	Mental health vignettes (na)	Yes/no; Multiple choice	na	40-50 minutes	Self-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
21	FCCHL-AYA[60]	***Functional, interactive and critical HL*** 1. Functional HL (6) 2. Communicative HL(3) 3. Critical HL (4)	Developmental change Dependency Differential epidemiology	Health topics regarding cancer in daily life (2nd grade)	4-point Likert scale	Score range: 13-52; Continuous score	na	Self-administered & Self-reported
22	ICHL[61]	***Interactive and critical HL*** 1. Interactive HL (2) 2. Critical HL (7)	Developmental change Demographic patterns Dependency Differential epidemiology	General health topics in daily life (pilot-tested)	5-point Likert scale	na	40-55 minutes (together with other measures)	Self-administered & Self-reported
23	HELMA[62]	***Functional, interactive and critical HL*** 1. Access (5) 2. Reading (5) 3. Understanding (10) 4. Appraise (5) 5. Use (4) 6. Communication (8) 7. Self-efficacy (4) 8. Numeracy (3)	Developmental change Dependency	General health topics in daily life (pilot-tested)	5-point Likert scale	Score range: 0-100; Ordinal category: 0-50: inadequate; 50.1-66: problematic; 66.1-84: sufficient; 84.1-100: excellent	15 minutes	Self-administered & Self-reported
24	HLSAC[63]	***Functional, interactive and critical HL*** 1. Theoretical knowledge (2) 2. Practical knowledge (2) 3. Individual critical thinking (2) 4. Self-awareness (2) 5. Citizenship (2)	Developmental change Dependency	General health topics in daily life (7th grade)	4-point Likert scale	na	45 minutes (together with the HBSC survey)	Self-administered & Self-reported
25	REALM-TeenS[64]	***Functional HL*** 1. Reading recognition (10)	Developmental change Demographic patterns	10 health-related words such as diabetes (6th grade)	Open-ended	Score range: 0-10; Ordinal category: 0-2: ≤ 3rd; 3-4: 4th-5th; 5-6: 6th-7th; 7-8: 8th-	13.6 seconds (range: 7.8-23.0)	Interviewer-administered & Performance-based

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
26	funHLS-YA[65]	***Functional HL*** 1. Word recognition and comprehension (19)	Developmental change	Diseases and symptoms, nutrition and diet, biology of the human body (na)	Multiple choice	9^{th}; 9-10: $\geq 10^{th}$ Score range: na; Continuous score	5 minutes	Self-administered & Performance-based
27	HLS-TCO[66]	***Functional, interactive and critical HL*** 1. Health knowledge and understanding (10) 2. Accessing information and services (5) 3. Communicating skills (6) 4. Managing health conditions (5) 5. Media literacy (5) 6. Making decisions (4)	Developmental change Demographic patterns Dependency Differential epidemiology	Health information for obesity preventive behaviours (pilot-tested)	Multiple choice; Likert scale	Score range: 0-135; Ordinal category: low: <21 for FHL, <33 for IHL, <27 for CHL; fair: 21-27.99 for FHL, 33-43.99 for IHL, 27-35.99 for CHL; high: 28-35 for FHL, 44-54.9 for IHL, 36-45 for CHL	na	Self-administered & Performance-based & Self-reported
28	HLRS-Y[67]	***Functional, interactive and critical HL*** 1. Knowledge (10) 2. Self-advocacy/support (14) 3. Resiliency (13)	Developmental change Demographic patterns Differential epidemiology	Health information about chronic health conditions (pilot-tested)	4-point Likert scale	Score range: na; Continuous score	15-20 minutes	Self-administered & Self-reported
29	p_HLAT-8[68]	***Functional, interactive and critical HL*** 1. Understanding health information (2) 2. Searching health information (2) 3. Communicating health information (2) Appraising health information (2)	Dependency	General health topics in daily life (pilot-tested)	5-point Likert scale; 4-point Likert scale	Score range: 0-37; Continuous score	na	Self-administered & Self-reported

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HBSC, Health Behaviour in School-aged Children; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HL, Health Literacy; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

Evaluation of methodological quality of included studies

According to the COSMIN checklist, the methodological quality of each instrument as assessed by each study is presented in **Table 3**. Almost all studies (n=28) examined content validity, 24 studies assessed internal consistency and hypotheses testing, 17 studies examined structural validity, eight studies assessed test-retest/inter-rater reliability, two studies assessed cross-cultural validity and only one study assessed responsiveness.

Evaluation of instruments' measurement properties

After the methodological quality assessment of included studies, measurement properties of each health literacy instrument were examined according to Terwee's quality criteria (See **Appendix 5**).[35] The rating results of measurement properties of each instrument are summarised in **Table 4**.

The synthesised evidence for the overall rating of measurement properties

Finally, a synthesis was conducted for the overall rating of measurement properties for each instrument according to '*the best evidence synthesis*' guidelines recommended by the COSMIN checklist developer group.[32] This synthesis result was derived from information presented in **Table 3** and **Table 4**. The overall rating of each measurement property for each health literacy instrument is presented in **Table 5**. In summary, most information (62.9%, 146/232) on measurement properties was unknown due to either poor methodological quality of studies or a lack of information on reporting or assessment.

Table 3. Methodological quality of each study for each measurement property according to the COSMIN checklist

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Fair	na	na
NVS (Driessnack *et al.*, 2014) [54]	Poor	na	na	Poor	na	Poor	na	na
NVS (Warsh *et al.*, 2014) [39]	na	na	na	Poor	na	Fair	na	na
TOFHLA (Chisolm and Buchanan, 2007) [48]	na	na	na	Poor	na	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Fair	na	na
c-sTOFHLAd (Chang *et al.*, 2012) [51]	Fair	Fair	na	Good	Fair	Fair	Fair	na
REALM-Teen (Davis *et al.*, 2006) [41]	Poor	Fair	na	Good	na	Fair	na	na
REALM-Teen (Hoffman *et al.*, 2013) [50]	Poor	na	na	Poor	na	Poor	na	na
HLAB (Wu *et al.*, 2010) [40]	Fair	Poor	na	Good	na	Fair	na	na
MMAHL (Massey *et al.*, 2013) [44]	Good	na	na	Good	Good	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [49]	Poor	na	na	Good	na	Good	na	na
DNT-39 (Mulvaney *et al.*, 2013) [53]	Fair	na	na	Poor	na	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013) [53]	Fair	na	na	Poor	na	Fair	na	na
eHEALS (Norman *et al.*, 2006) [43]	Fair	Fair	na	Good	Fair	Fair	na	na
CHC Test (Steckelberg *et al.*, 2009) [47]	na	Poor	na	Good	Poor	na	na	na
HKACSS (Schmidt *et al.*, 2010) [46]	**Excellent**	na	na	Good	na	Good	na	na
HLAT-51 (Harper, 2014) [42]	Poor	na	na	Good	Poor	na	na	na
HLAT-8 (Abel *et al.*, 2014) [45]	**Excellent**	na	na	Poor	**Excellent**	Good	na	na
CHLT (Liu *et al.*, 2014) [55]	Fair	na	na	Good	Fair	Fair	na	na
VOHL (Ueno *et al.*, 2014) [56]	na	Fair	na	na	na	Fair	na	Fair
HAS-A (Manganello *et al.* 2015) [57]	Fair	na	na	Good	Fair	Fair	na	na
MaHeLi (Naigaga *et al.* 2015) [58]	Fair	na	na	Poor	Fair	na	na	na
QuALiSMental (de Jesus Loureiro *et al.*, 2015) [59]	Fair	na	na	**Excellent**	Fair	Fair	na	na
FCCHL-AYAC (McDonald *et al.*, 2016) [60]	Fair	na	na	Good	Fair	Fair	na	na

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive -ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
ICHL (Smith *et al.*, 2016) [61]	na	na	na	Good	na	Fair	na	na
HELMA (Ghanbari *et al.*, 2016) [62]	Good	Good	na	Good	Good	na	na	na
HLSAC (Paakkari *et al.*, 2016) [63]	Fair	Fair	na	Good	Fair	Fair	na	na
REALM-TeenS (Manganello *et al.*, 2017) [64]	Good	na	na	Good	na	Good	na	na
funHLS-YA (Tsubakita *et al.*, 2017) [65]	Fair	na	na	Poor	Fair	Fair	na	na
HLS-TCO (Intarakamhang *et al.*, 2017) [66]	Fair	na	na	Good	Fair	Fair	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017) [67]	Fair	na	na	**Excellent**	Fair	Fair	na	na
p_HLAT-8 (Quemelo *et al.*, 2017) [68]	Fair	na	na	Good	Fair	Fair	Fair	na

Note: na, no information available. CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

Table 4. Evaluation of measurement properties for included instruments according to Terwee's quality criteria

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS (Hoffman *et al.*, 2013) [50]	-	na	na	?	na	-	na	na
NVS (Driessnack *et al.*, 2014) [54]	+	na	na	?	na	-	na	na
NVS (Warsh *et al.*, 2014) [39]	na	na	na	?	na	+	na	na
TOFHLA (Chisolm and Buchanan, 2007) [48]	na	na	na	?	na	-	na	na
s-TOFHLA (Hoffman *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
c-sTOFHLAd (Chang *et al.*, 2012) [51]	+	+	na	+	?	+	?	na
REALM-Teen (Davis *et al.*, 2006) [41]	+	+	na	+	na	+	na	na
REALM-Teen (Hoffman *et al.*, 2013) [50]	+	na	na	?	na	-	na	na
HLAB (Wu *et al.*, 2010) [40]	+	+	na	+	na	-	na	na
MMAHL (Massey *et al.*, 2013) [44]	+	na	na	+	-	na	na	na
MHL (Levin-Zamir *et al.*, 2011) [49]	+	na	na	+	na	+	na	na
DNT-39 (Mulvaney *et al.*, 2013) [53]	+	na	na	?	na	-	na	na
DNT-14 (Mulvaney *et al.*, 2013) [53]	+	na	na	?	na	-	na	na
eHEALS (Norman and Skinner, 2006) [43]	+	-	na	+	+	-	na	na
CHC Test (Steckelberg *et al.*, 2009) [47]	na	+	na	+	+	na	na	na
HKACSS (Schmidt *et al.*, 2010) [46]	+ (HC) - (HA)	na	na	+	na	+	na	na
HLAT-51 (Harper, 2014) [42]	?	na	na	+	?	na	na	na
HLAT-8 (Abel *et al.*, 2014) [45]	-	na	na	?	+	+	na	na
CHLT (Liu *et al.*, 2014) [55]	+	na	na	+	+	+	na	na
VOHL (Ueno *et al.*, 2014) [56]	na	- (TS) + (GS)	na	na	na	-	na	+
HAS-A (Manganello *et al.* 2015) [57]	+	na	na	+	+	-	na	na
MaHeLi (Naigaga *et al.* 2015) [58]	+	na	na	?	+	na	na	na
QuALiSMental (de Jesus Loureiro *et al.*, 2015) [59]	-	na	na	+	+	+	na	na
FCCHL-AYAC (McDonald *et al.*, 2016) [60]	+ (FHL) - (IHL) + (CHL)	na	na	+	+	-	na	na
ICHL (Smith *et al.*, 2016) [61]	na	na	na	+	na	+	na	na
HELMA (Ghanbari *et al.*, 2016) [62]	+	+	na	+	+	na	na	na
HLSAC (Paakkari *et al.*, 2016) [63]	+	+	na	+	-	+	na	na
REALM-TeenS (Manganello *et al.*,	+	na	na	+	na	+	na	na

Health literacy instrument (Author, year)	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsive-ness
					Structural validity	Hypotheses testing	Cross-cultural validity	
2017) [64]								
funHLS-YA (Tsubakita *et al.*, 2017) [65]	+	na	na	?	+	-	na	na
HLS-TCO (Intarakamhang *et al.*, 2017) [66]	+	na	na	+	+	+	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017) [67]	+	na	na	+	+	+	na	na
p_HLAT-8 (Quemelo *et al.*, 2017) [68]	+	na	na	+	+	-	+	na

Note: na, no information available; + positive rating; ? indeterminate rating; - negative rating. CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; GS, Gingiva Score; HA, Health Attitude; HAS-A, Health Literacy Assessment Scale for Adolescents; HC, Health Communication; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; TS, Tooth Score; VOHL, Visual Oral Health Literacy.

Table 5. The overall quality of measurement properties for each health literacy instrument used in children and adolescents

Health literacy instrument	Internal consistency	Reliability	Measurement error	Content validity	Construct validity			Responsiveness
					Structural validity	Hypotheses testing	Cross-cultural validity	
NVS [50 54 39]	?	na	na	?	na	±	na	na
TOFHLA [48]	na	na	na	?	na	-	na	na
s-TOFHLA [50]	?	na	na	?	na	-	na	na
c-sTOFHLAd [51]	+	+	na	++	?	+	?	na
REALM-Teen [41 50]	?	+	na	++	na	+	na	na
HLAB [40]	+	?	na	++	na	-	na	na
MMAHL [44]	++	na	na	++	--	na	na	na
MHL [49]	?	na	na	++	na	++	na	na
DNT-39 [53]	+	na	na	?	na	-	na	na
DNT-14 [53]	+	na	na	?	na	-	na	na
eHEALS [43]	+	-	na	++	+	-	na	na
CHC Test [47]	na	?	na	++	?	na	na	na
HKACSS [46]	+++ (HC) --- (HA)	na	na	++	na	++	na	na
HLAT-51 [42]	?	na	na	++	?	na	na	na
HLAT-8 [45]	---	na	na	?	+++	++	na	na
CHLT [55]	+	na	na	++	+	+	na	na
VOHL [56]	na	- (TS) + (GS)	na	na	na	-	na	+
HAS-A [57]	+	na	na	++	+	-	na	na
MaHeLi [58]	+	na	na	?	+	na	na	na
QuALiSMental [59]	-	na	na	+++	+	+	na	na
FCCHL-AYAC [60]	+ (FHL) - (IHL) + (CHL)	na	na	++	+	-	na	na
ICHL [61]	na	na	na	++	na	+	na	na
HELMA [62]	++	++	na	++	++	na	na	na
HLSAC [63]	+	+	na	++	-	+	na	na
REALM-TeenS [64]	++	na	na	++	na	++	na	na
funHLS-YA [65]	+	na	na	?	+	-	na	na
HLS-TCO [66]	+	na	na	++	+	+	na	na
HLRS-Y [67]	+	na	na	+++	+	+	na	na
p_HLAT-8 [68]	+	na	na	++	+	-	+	na

Note: na, no information available; +++ or --- strong evidence and positive/negative result; ++ or -- moderate evidence and positive/negative result; + or - limited evidence and positive/negative

result; ± conflicting evidence; ? unknown, due to poor methodological quality or indeterminate rating of a measurement property. CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; GS, Gingiva Score; HA, Health Attitude; HAS-A, Health Literacy Assessment Scale for Adolescents; HC, Health Communication; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; TS, Tooth Score; VOHL, Visual Oral Health Literacy.

Discussion

Summary of main results

This study identified and examined 29 health literacy instruments used in children and adolescents and exemplified the large variety of methods used. Compared to previous three systematic reviews,[10 11 13] this review identified 19 additional new health literacy instruments and critically appraise measurement properties of each instrument. It showed that to date, only half of included health literacy instruments (15/29) measure all three domains (functional, interactive and critical) and that the functional domain is still the focus of attention when measuring health literacy in children and adolescents. Additionally, researchers mainly focus on participant characteristics of developmental change (of cognitive ability), dependency (on parents) and demographic patterns (e.g. racial/ethnic backgrounds), and less so on differential epidemiology (of health and illness). The methodological quality of included studies as assessed via measurement properties varied from poor to excellent. Most information (62.9%) on measurement properties was unknown due to either the poor methodological quality of studies or a lack of reporting or assessment. It is therefore difficult to draw a robust conclusion about which instrument is the best.

Health literacy measurement in children and adolescents

This review found that health literacy measurement in children and adolescents tends to include Nutbeam's three-domain health literacy construct (i.e. functional, interactive and critical), especially in the past five years. However, almost one third of included instruments focused only on the functional domain (n=10). Unlike health literacy research for patients in clinics, health literacy research for children and adolescents (a comparatively healthy population) should be considered from a health promotion perspective,[69] rather than a health care or disease management perspective. Integrating interactive and critical domains into health literacy measurement is aligned with the rationale of emphasising empowerment in health promotion for children and adolescents.[70] The focus of health literacy for this population group

should therefore include all three domains and so there is a need for future research to integrate the three domains within health literacy instruments.

Similar to previous findings by Ormshaw *et al.*[10] and Okan *et al.*,[13] this review also revealed that childhood and adolescent health literacy measurement varied by its dimensions, health topics, forms of administration, and by the level to which participant characteristics were considered. There are likely four main reasons for these disparities. First, definitions of health literacy were inconsistent. Some researchers measured general health literacy,[40 45] while others measured eHealth literacy or media health literacy.[43 49] Second, researchers had different research purposes for their studies. Some researchers used what were originally adult instruments to measure adolescent health literacy,[39 48 52] whereas others developed new or adapted instruments.[40-42 53] Third, the research settings affected the measurement process. As clinical settings were busy, short surveys were more appropriate than long surveys.[39 41 44] On the other hand, health literacy in school settings was often measured using long and comprehensive surveys.[40 42 47] Fourth, researchers considered different participant characteristics when measuring health literacy in children and adolescents. For example, some researchers took considerations of students' cognitive development,[40 41 44 46 51] some focused on adolescents' resources and environments (e.g. friends and family contexts, eHealth contexts, media contexts),[43 45 49] and others looked at the effect of different cultural backgrounds and socio-economic status.[40 41 43 44 46 47 49-52] Based on Forrest *et al.*'s 4D model,[14 15] this review showed that most health literacy instruments considered participants' development, dependency and demographic patterns, with only seven instruments considering differential epidemiology.[53 58 60 61 66 67] Although the '4D' model cannot be used to reduce the disparities in health literacy measurement, it does provide an opportunity to identify gaps in current research and assist researchers to consider participants' characteristics comprehensively in future research.

The methodological quality of included studies

This review included a methodological quality assessment of included studies, which was absent from previous reviews on this subject.[10 11] Methodological quality assessment is important because strong conclusions about the measurement properties

of instruments can only be drawn from high-quality studies. In this review, the COSMIN checklist was shown to be a useful framework for critically appraising the methodological quality of studies via each measurement property. Findings suggested that there was wide variation in the methodological quality of studies for all instruments. Poor methodological quality of studies was often seen in the original or adapted health literacy instruments (the NVS, the TOFHLA, the s-TOFHLA, the DNT-39 and the DNT-14) for two main reasons. The first reason was the vague description of the target population involved. This suggested that researchers were less likely to consider an instrument's content validity when using the original, adult instrument for children and/or adolescents. Given that children and adolescents have less well-developed cognitive abilities, in future it is essential to assess whether all items within an instrument are understood. The second reason was a lack of uni-dimensionality analysis for internal consistency. As explained by the COSMIN group,[71] a set of items can be inter-related and multi-dimensional, whereas uni-dimensionality is a prerequisite for a clear interpretation of the internal consistency statistics (Cronbach's alpha). Future research on the use of health literacy instruments therefore needs to assess and report both internal consistency statistics and uni-dimensionality analysis (e.g. factor analysis).

Critical appraisal of measurement properties for included instruments

This review demonstrated that of all instruments reviewed three instruments (the c-sTOFHLAd, the HELMA and the HLSAC) showed satisfactory evidence about internal consistency and test-retest reliability. Based on the synthesised evidence, the HELMA showed moderate evidence and positive results of internal consistency (α=0.93) and test-retest reliability (intraclass correlation coefficient (ICC)=0.93), whereas the HLSAC (α=0.93; standardised stability estimate=0.83) and the c-sTOFHLAd (α=0.85; ICC=0.95) showed limited evidence and positive results. Interestingly, compared to the overall reliability rating of the s-TOFHLA,[50] the c-sTOFHLAd showed better results.[72] The reason for this was probably the different methodological quality of the studies that examined the s-TOFHLA and the c-sTOFHLAd. The c-sTOFHLAd study had fair methodological quality in terms of

internal consistency and test-retest reliability, whereas the original s-TOFHLA study had poor methodological quality for internal consistency and unknown information for test-retest reliability. Given the large disparity of rating results between the original and translated instrument, further evidence is needed to confirm whether the s-TOFHLA has the same or a different reliability within different cultures, thus assisting researchers to understand the generalisability of the s-TOFHLA's reliability results.

Four instruments were found to show satisfactory evidence about both content validity and construct validity (structural validity and hypotheses testing). Construct validity is a fundamental aspect of psychometrics and was examined in this review for two reasons. First, it enables an instrument to be assessed for the extent to which operational variables adequately represent underlying theoretical constructs.[73] Second, the overall rating results of content validity for all included instruments were similar (i.e. unknown or moderate/strong evidence and positive result). The only difference was that the target population was involved or not. Given that all instruments' items reflected the measured construct, in this review, construct validity was determined to be key to examining the overall validity of included instruments. In this context, only the HLAT-8 showed strong evidence and positive result for structural validity (CFI=0.99; TLI=0.97; RMSEA=0.03; SRMR=0.03) and moderate evidence on hypotheses testing (known-group validity results showed differences of health literacy by gender, educational status and health valuation). However, in the original paper,[45] the HLAT-8 was only tested for its known-group validity, not for convergent validity. Examination of convergent validity is important because it assists researchers in understanding the extent to which two examined measures' constructs are theoretically and practically related.[74] Therefore, future research on the convergent validity of the HLAT-8 would be beneficial for complementing that which exists for its construct validity.

Similar to a previous study by Jordan *et al.*,[26] this review demonstrated that only one included study contained evidence of responsiveness. Ueno *et al.*[56] developed a visual oral health literacy instrument and examined responsiveness by comparing changes in health literacy before and after oral health education. Their results showed students' health literacy scores increased significantly after health education. Responsiveness is

the ability of an instrument to detect change over time in the construct being measured, and it is particularly important for longitudinal studies.[31] However, most studies included in this review were cross-sectional studies, and only one study (on the MMAHL[44]) discussed the potential to measure health literacy over time. Studies that measure health literacy over time in populations are needed, not only because this is a prerequisite for longitudinal studies, but also so that the responsiveness of instruments can be monitored and improved.

Feasibility issues for included instruments

This review showed that the feasibility aspects of instruments varied markedly. In relation to forms of administration, this review identified 19 self-administered instruments and 10 interviewer-administered instruments. This suggests that self-administered instruments are more commonly used in practice than interviewer-administered instruments. However, both administration modes have limitations. Self-administered instruments are cost-effective and efficient, but may bring about respondent bias, whereas interviewer-administered instruments, while able to ensure high response rates, are always resource intensive and expensive to administer.[75] Although the literature showed that there was no significant difference in scores outcome between these two administration modes,[76 77] the relevant studies mostly concerned health-related quality of life instruments. It is still unknown whether the same is true for health literacy instruments. Among children and adolescents, health literacy research is more likely to be conducted through large-scale surveys in school settings. Therefore, the more cost-effective, self-administered mode seems to have great potential for future research. To further support the wide use of self-administered instruments, there is a need for future research to confirm the same effect of administration between self-administered and interviewer-administered instruments.

With regard to the type of assessment method, this review revealed that performance-based health literacy instruments (n=15) are more preferable than self-report instruments (n=11). There might be two reasons for this. First, it is due to participant characteristics. Compared with adults, children and adolescents are more dependent on their parents for health-related decisions.[15] Measurement error is more likely to

occur when children and adolescents answer self-report items.[78] Therefore, performance-based assessment is often selected to avoid such inaccuracy. Second, performance-based instruments are objective, whereas self-report instruments are subjective and may bring about over-estimated results.[79] However, the frequent use of performance-based instruments does not mean that they are more appropriate than self-report instruments when measuring childhood and adolescent health literacy. Compared with performance-based instruments, self-report instruments are always time-efficient and help to preserve respondents' dignity.[21] The challenge in using self-report instruments is to consider the readability of tested materials. If children and adolescents can understand what a health literacy instrument measures, then they are more able to accurately self-assess their own health literacy skills.[70] The difference between self-report and performance-based instruments of health literacy has been discussed in the literature,[80] but the evidence about the difference is still limited due to a lack of specifically-designed studies for exploring the difference. Further studies are needed to fill this knowledge gap.

Recommendations for future research

This review identified 18 instruments (the REALM-Teen, the NVS, the s-TOFHLA, the c-sTOFHLAd, the eHEALS, the CHC Test, the HKACSS, the HLAB, the MHL, the HLAT-51, the CHLT, the VOHL, the QuALiSMental, the HELMA, the HLSAC, the funHLS-YA, the HLS-TCO and the p_HLAT-8) that were used to measure health literacy in school settings. Although it is difficult to categorically state which instrument is the best, this review provides useful information that will assist researchers to identify the most suitable instrument to use when measuring health literacy in children and adolescents in school contexts.

Among the 18 instruments, six tested functional health literacy (the REALM-Teen, the NVS, the s-TOFHLA, the c-sTOFHLAd, the VOHL and the funHLS-YA); one examined critical health literacy (the CHC Test); one measured functional and interactive health literacy (the HKACSS); one examined functional and critical health literacy (the HLAB); and nine tested health literacy comprehensively focusing on functional, interactive and critical domains (the eHEALS, the MHL, the HLAT-51, the CHLT, the QuALiSMental, the HELMA, the HLSAC, the HLS-TCO and the

p_HLAT-8). However, only one of these three-domain instruments (the HLSAC) was considered appropriate for use in schools because of its quick administration, satisfactory reliability and one-factor validity. Eight three-domain instruments were excluded due to the fact that they focused on non-general health literacy (the eHEALS, the MHL, the QuALiSMental, the HLS-TCO) or were burdensome to administer (the HLAT-51, the HELMA-44) or were not published in English (the CHLT and the p_HLAT-8).

Compared with the HLSAC, the HLAT-8 examines the construct of health literacy via three domains rather than one-factor structure, thus enabling a more comprehensive examination of the constrcut. Meanwhile, although the p_HLAT-8 (Portuguese version) is not available in English, the original HLAT-8 is. After comparing measurement domains and measurement perperties, the HLAT-8 was deemed to be more suitable for measuring health literacy in school settings for four reasons: (1) it measures health literacy in the context of family and friends,[45] a highly important attribute because children and adolescents often need support for health decisions from parents and peers;[7 15] (2) it is a short but comprehensive tool that captures Nutbeam's three-domain nature of health literacy;[17] (3) it showed satisfactory structural validity (RMSEA=0.03; CFI=0.99; TLI=0.97; SRMR=0.03);[45] and (4) it has good feasibility (e.g. the p_HLAT-8 is self-administered and time-efficient) in school-based studies. However, there are still two main aspects that need to be considered in future. One aspect is its use in the target population. Given the HLAT-8 has not been tested for children and adolescents under 18, its readability and measurement properties need to be evaluated. The other aspect is that its convergent validity (the strength of association between two measures of a similar construct, an essential part of construct validity) has not been examined. Testing convergent validity of the HLAT-8 is important because high convergent validity assists researchers to understand the extent to which two examined measures' constructs are theoretically and practically related.

Limitations

This review was not without limitation. First, we restricted the search to studies aiming to develop or validate a health literacy instrument. Thus we may have missed

relevant instruments in studies that were not aiming to develop instruments.[81 82] Second, although the COSMIN checklist provided us with strong evidence of the methodological quality of a study via an assessment of each measurement property, it cannot evaluate a study's overall methodological quality. Third, criterion validity was not examined due to lack of 'gold standard' for health literacy measurement. However, we examined convergent validity under the domain of 'hypotheses testing'. This can ascertain the validity of newly-developed instruments against existing commonly-used instruments. Finally, individual subjectivity inevitably played a part in the screening, data extraction and synthesis stage of the review. To reduce this subjectivity, two authors independently managed the major stages.

CONCLUSION

This review updated previous reviews of childhood and adolescent health literacy measurement (c.f. Ormshaw *et. al.*, Perry & Okan *et al.*) to incorporate a quality assessment framework. It showed that most information on measurement properties was unknown due to either the poor methodological quality of studies or a lack of assessment and reporting. Rigorous and high-quality studies are needed to fill the knowledge gap in relation to health literacy measurement in children and adolescents. Although it is challenging to draw a robust conclusion about which instrument is the best, this review provides important evidence that supports the use of the HLAT-8 to measure childhood and adolescent health literacy in future research.

ACKNOWLEDGMENTS

This paper is part of Shuaijun Guo's PhD research project, which is supported by the Melbourne International Engagement Award. The authors also appreciate the helpful comments received from the reviewers (Martha Driessnack and Debi Bhattacharya) at *BMJ Open*.

CONTRIBUTORS

SG conceived the review approach. RA and EW provided general guidance for the drafting of the protocol. SG and SA screened the literature. SG and TS extracted data. SG drafted the manuscript. SG, GB, RA, EW, XY, SA and TS reviewed and revised the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

None.

DATA SHARING STATEMENT

There are no additional data available.

REFERENCES

1. Nutbeam D. The evolving concept of health literacy. *Soc Sci Med* 2008;67(12):2072-78.
2. Nutbeam D. Health promotion glossary. *Health Promot Int* 1998;13(4):349-64.
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav* 2007;31(Supplement 1):S19-S26.
4. Squiers L, Peinado S, Berkman N, et al. The health literacy skills framework. *J Health Commun* 2012;17(sup3):30-54.
5. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155(2):97-107.
6. Dodson S, Good S, Osborne R. Health literacy toolkit for low-and middle-income countries: a series of information sheets to empower communities and strengthen health systems. New Delhi: World Health Organization, Regional Office for South-East Asia, 2015.
7. Manganello J. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res* 2008;23(5):840-47.
8. Diamond C, Saintonge S, August P, et al. The development of building wellness™, a youth health literacy program. *J Health Commun* 2011;16(sup3):103-18.
9. Robinson LD, Jr., Calmes DP, Bazargan M. The impact of literacy enhancement on asthma-related outcomes among underserved children. *J Natl Med Assoc* 2008;100(8):892-96.

10. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. *Health Educ* 2013;113(5):433-55.
11. Perry EL. Health literacy in adolescents: an integrative review. *J Spec Pediatr Nurs* 2014;19(3):210-18.
12. Bröder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017;17:361.
13. Okan O, Lopes E, Bollweg TM, et al. Generic health literacy measurement instruments for children and adolescents: a systematic review of the literature. *BMC public health* 2018;18(1):166.
14. Forrest CB, Simpson L, Clancy C. Child health services research: challenges and opportunities. *JAMA* 1997;277(22):1787-93.
15. Rothman RL, Yin HS, Mulvaney S, et al. Health literacy and quality: focus on chronic illness care and patient safety. *Pediatrics* 2009;124(Supplement 3):S315-S26.
16. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12(1):80.
17. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15(3):259-67.
18. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? *J Health Commun* 2010;15(S2):9-19.
19. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies? *Int J Public Health* 2009;54(5):303-05.
20. Abel T. Measuring health literacy: moving towards a health-promotion perspective. *Int J Public Health* 2008;53(4):169-70.
21. Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. *J Health Commun* 2014;19(12):1481-96.
22. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. In: Higgins JP, Green S, eds. London, UK: The Cochrane Collaboration, 2011.
23. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151(4):264-69.
24. Simonds SK. Health education as social policy. *Health Educ Monogr* 1974;2(1_suppl):1-10.
25. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. *Nurs Health Sci* 2009;11(1):77-89.
26. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol* 2011;64(4):366-79.
27. Sanders LM, Federico S, Klass P, et al. Literacy and child health: a systematic review. *Arch Pediatr Adolesc Med* 2009;163(2):131-40.
28. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics* 2009;124(Supplement 3):S265-S74.
29. Sawyer S, Sawyer R, Afifi L, et al. Adolescence: a foundation for future health. *Lancet (London, England)* 2012;379(9826):1630-40.
30. Fok TKS, Wong. What does health literacy mean to children? *Contemporary Nurse : a Journal for the Australian Nursing Profession* 2002;13(2-3):249-58.

31. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):651-57.
32. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13.
33. McCormack L, Haun J, Sørensen K, et al. Recommendations for Advancing Health Literacy Measurement. *J Health Commun* 2013;18(sup1):9-14.
34. Pleasant A, McKinney J, Rikard RV. Health Literacy Measurement: A Proposed Research Agenda. *J Health Commun* 2011;16(sup3):11-21.
35. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
36. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490-94.
37. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-26.
38. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012;21(4):659-70.
39. Warsh J, Chari R, Badaczewski A, et al. Can the newest vital sign be used to assess health literacy in children and adolescents? *Clin Pediatr* 2014;53(2):141-44.
40. Wu A, Begoray D, Macdonald M, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. *Health Promot Int* 2010;25(4):444-52.
41. Davis TC, Wolf MS, Arnold CL, 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(6):e1707-e14.
42. Harper R. Development of a health literacy assessment for young adult college students: A pilot study. *J Am Coll Health* 2014;62(2):125-34.
43. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006;8(4):e27.
44. Massey P, Prelip M, Calimlim B, et al. Findings Toward a Multidimensional Measure of Adolescent Health Literacy. *Am J Health Behav* 2013;37(3):342-50.
45. Abel T, Hofmann K, Ackermann S, et al. Health literacy among young adults: a short survey tool for public health and health promotion research. *Health Promot Int* 2014;30(3):725-35.
46. Schmidt CO, Fahland RA, Franze M, et al. Health-related behaviour, knowledge, attitudes, communication and social status in school children in Eastern Germany. *Health Educ Res* 2010;25(4):542-51.
47. Steckelberg A, Hülfenhaus C, Kasper J, et al. How to measure critical health competences: development and validation of the Critical Health Competence Test (CHC Test). *Adv Health Sci Educ* 2009;14:11-22.

48. Chisolm DJ, Buchanan L. Measuring adolescent functional health literacy: a pilot validation of the test of functional health literacy in adults. *J Adolescent Health* 2007;41(3):312-14.
49. Levin-Zamir D, Lemish D, Gofin R. Media Health Literacy (MHL): development and measurement of the concept among adolescents. *Health Educ Res* 2011;26(2):323-35. doi: 10.1093/her/cyr007
50. Hoffman S, Trout A, Nelson T, et al. A psychometric assessment of health literacy measures among youth in a residential treatment setting. *J Stud Soc Sci* 2013;5(2):288-300.
51. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese version of short-form Test of Functional Health Literacy in Adolescents. *J Clin Nurs* 2012;21(17-18):2429-37. doi: 10.1111/j.1365-2702.2012.04147.x
52. Driessnack M, Chung S, Perkhounkova E, et al. Using the "newest vital sign" to assess health literacy in children. *J Pediatr Health Care* 2014;28(2):165-71.
53. Mulvaney SA, Lilley JS, Cavanaugh KL, et al. Validation of the diabetes numeracy test with adolescents with type 1 diabetes. *J Health Commun* 2013;18(7):795-804.
54. Driessnack M, Chung S, Perkhounkova E, et al. Using the "Newest Vital Sign" to assess health literacy in children. *Journal of Pediatric Health Care* 2014;28(2):165-71.
55. Liu CH, Liao LL, Shih SF, et al. Development and implementation of Taiwan's child health literacy test. *Taiwan Journal of Public Health* 2014;33(3):251-70.
56. Ueno M, Takayama A, Adiatman M, et al. Application of visual oral health literacy instrument in health education for senior high school students. *International Journal of Health Promotion and Education* 2014;52(1):38-46. doi: http://dx.doi.org/10.1080/14635240.2013.845412
57. Manganello JA, DeVellis RF, Davis TC, et al. Development of the Health Literacy Assessment Scale for Adolescents (HAS-A). *Journal of communication in healthcare* 2015;8(3):172-84. doi: 10.1179/1753807615y.0000000016 [published Online First: 2015/01/01]
58. Naigaga MDAS, Pettersen KS. Measuring Maternal Health Literacy in Adolescents Attending Antenatal Care in Uganda: Exploring the Dimensionality of the Health Literacy Concept Studying a Composite Scale. *Journal of nursing measurement* 2015;23(2):E50.
59. de Jesus Loureiro LM. Questionnaire for Assessment of Mental Health Literacy--QuALiSMental: study of psychometric properties. *Revista de Enfermagem Referência* 2015;4(4):79-88. doi: 10.12707/riv14031
60. McDonald FE, Patterson P, Costa DS, et al. Validation of a Health Literacy Measure for Adolescents and Young Adults Diagnosed with Cancer. *Journal of adolescent and young adult oncology* 2016;5(1):69-75. doi: 10.1089/jayao.2014.0043 [published Online First: 2016/01/27]
61. Smith SR, Samar VJ. Dimensions of Deaf/Hard-of-Hearing and Hearing Adolescents' Health Literacy and Health Knowledge. *Journal of health communication* 2016;21:141-54. doi: 10.1080/10810730.2016.1179368
62. Ghanbari S, Ramezankhani A, Montazeri A, et al. Health Literacy Measure for Adolescents (HELMA): Development and Psychometric Properties. *PloS one* 2016;11(2):e0149202. doi: 10.1371/journal.pone.0149202 [published Online First: 2016/02/18]

63. Paakkari O, Torppa M, Kannas L, et al. Subjective health literacy: Development of a brief instrument for school-aged children. *Scandinavian journal of public health* 2016;44(8):751-57. doi: 10.1177/1403494816669639
64. Manganello JA, Colvin KF, Chisolm DJ, et al. Validation of the Rapid Estimate for Adolescent Literacy in Medicine Short Form (REALM-TeenS). *Pediatrics* 2017;139(5) doi: 10.1542/peds.2016-3286 [published Online First: 2017/05/31]
65. Tsubakita T, Kawazoe N, Kasano E. A New Functional Health Literacy Scale for Japanese Young Adults Based on Item Response Theory. *Asia-Pacific journal of public health* 2017;29(2):149-58. doi: 10.1177/1010539517690226 [published Online First: 2017/02/17]
66. Intarakamhang U, Intarakamhang P. Health Literacy Scale and Causal Model of Childhood Overweight. *Journal of research in health sciences* 2017;17(1):e00368.
67. Bradley-Klug K, Shaffer-Hudkins E, Lynn C, et al. Initial development of the Health Literacy and Resiliency Scale: Youth version. *Journal of communication in healthcare* 2017;10(2):100-07. doi: 10.1080/17538068.2017.1308689
68. Quemelo PRV, Milani D, Bento VF, et al. Health literacy: translation and validation of a research instrument on health promotion in Brazil. *Cadernos de saude publica* 2017;33(2):e00179715. doi: 10.1590/0102-311x00179715
69. Kilgour L, Matthews N, Christian P, et al. Health literacy in schools: prioritising health and well-being issues through the curriculum. *Sport Educ Soc* 2015;20(4):485-500.
70. Velardo S, Drummond M. Emphasizing the child in child health literacy research. *J Child Health Care* 2017;21(1):5-13.
71. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist manual. Amsterdam: VU University Medical Centre, 2012.
72. Chang LC, Hsieh PL, Liu CH. Psychometric evaluation of the Chinese version of short-form Test of Functional Health Literacy in Adolescents. *Journal of clinical nursing* 2012;21(17-18):2429-37. doi: 10.1111/j.1365-2702.2012.04147.x
73. Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008;98(1):9-10.
74. Akpa OM, Bamgboye EA, Baiyewu O. The Adolescents' Psychosocial Functioning Inventory (APFI): scale development and initial validation using Exploratory and Confirmatory Factor Analysis. *Afr J Psychol Study Soc Issues* 2015;18(1):1-21.
75. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health* 2005;27(3):281-91.
76. Lozano F, Lobos JM, March JR, et al. Self-administered versus interview-based questionnaires among patients with intermittent claudication: Do they give different results? A cross-sectional study. *Sao Paulo Med J* 2016;134(1):63-69.
77. Dujaili JA, Sulaiman SAS, Awaisu A, et al. Comparability of Interviewer-Administration Versus Self-Administration of the Functional Assessment of Chronic Illness Therapy-Tuberculosis (FACIT-TB) Health-Related Quality of Life Questionnaire in Pulmonary Tuberculosis Patients. *Pulm Ther* 2016;2(1):127-37.

78. Vaz S, Parsons R, Passmore AE, et al. Internal consistency, test–retest reliability and measurement error of the self-report version of the Social Skills Rating System in a sample of Australian adolescents. *PloS one* 2013;8(9):e73924.
79. Altin SV, Finke I, Kautz-Freimuth S, et al. The evolution of health literacy assessment tools: a systematic review. *BMC Public Health* 2014;14:1207.
80. Kiechle ES, Bailey SC, Hedlund LA, et al. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. *J Gen Intern Med* 2015;30(10):1538-46.
81. Paek H-J, Reber BH, Lariscy RW. Roles of interpersonal and media socialization agents in adolescent self-reported health literacy: a health socialization perspective. *Health Educ Res* 2011;26(1):131-49.
82. Hubbard B, Rainey J. Health Literacy Instruction and Evaluation among Secondary School Students. *Am J Health Educ* 2007;38(6):332-37.

FIGURE

Figure 1. Flowchart of search and selection process according to PRISMA flow diagram

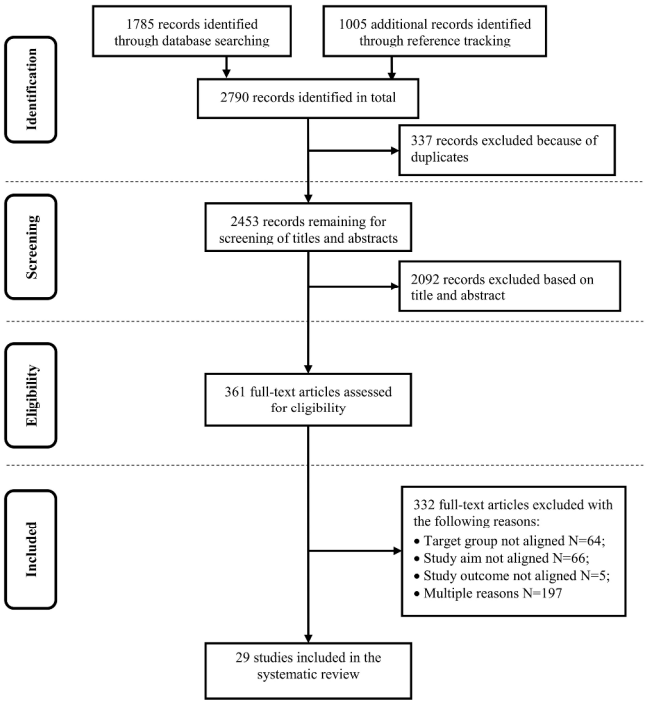

Figure 1. Flowchart of search and selection process according to PRISMA flow diagram

297x420mm (300 x 300 DPI)

Appendix 1. A systematic review protocol

Measuring the Quality of Child and Adolescent Health Literacy Instruments: A Systematic Review

Shuaijun Guo[1*], Rebecca Armstrong[1], Elizabeth Waters[1], Thirunavukkarasu Sathish[1], Sheikh M Alif[1], Geoffrey R Browne[1], Xiaoming Yu[2*]

[1] *School of Population and Global Health, The University of Melbourne, Melbourne, Australia*
[2] *Institute of Child and Adolescent Health, School of Public Health, Peking University, Beijing, China*

* Corresponding author email: gshj1986@gmail.com yxm@bjmu.edu.cn

Background

Health literacy research has been a growing interest by researchers across the globe. The term *'health literacy'* was first used in 1974 in the proceedings of a health education conference discussing health education as a social policy issue affecting the healthcare system, mass communication and the education system (1, 2). However, few references were found regarding health literacy in the literature until 1992 (3). Since 1992, health literacy has been broadly studied both in clinical and public health contexts. In clinical settings, health literacy is typically 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*' by the Institute of Medicine (IOM) in America (4). In such circumstances, health literacy is a derivative concept from literacy and numeracy skills, which is often used as a risk factor that needs to be identified and appropriately managed for patients and health professionals (5). Accordingly, health literacy measurement tools and 'screening aids' for clinicians are developed to assess patient literacy levels, and help health professionals to tailor health information for better communication with their patients (6). From the public health perspective, health literacy is defined and accepted by World Health Organization (WHO) as '*the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health*' (7). This understanding of health literacy identifies it as a broad concept, which is seen as a personal asset to enable individuals to take more control over their health and determinants of health (5). With a different understanding of the concept, health literacy measures vary in a different way. Although health literacy measurement varies and is still

being debated (1, 8-10), there is consistent evidence showing health literacy is of potential importance and considered as a public health goal internationally. A recent WHO report pointed out that poor health literacy skills were associated with riskier behaviours, poorer health status, less self-management and longer hospitalization and more health costs (11).

Based on a preliminary search of health literacy, there were more interests in studies focusing on adult health literacy than adolescent health literacy. However, previous research studies suggested that poor health literacy was a prevalent problem in adolescents. In Australia, the 2006 National Health Literacy Survey reported that 67.6% of adolescents aged 15 to 19 years old did not attain the minimum skills required to deal with health information and service in everyday life (12). Compared with adult health literacy, there are several reasons for the potential importance of adolescent health literacy: 1) adolescents are future mainstream and independent healthcare consumers, a health literate person can contribute to less health care costs, better health status compared to that is not health literate (13); 2) adolescents are at a critical stage of development characterised by physical, emotional and cognitive changes, attempting to prepare for independence but lacking the adequate ability of reasoning and decision-making. Therefore, improving their health literacy skills could support sound health decisions in future (14, 15); 3) low health literacy has been demonstrated to associate with high levels of health-risk behaviors (16, 17) and low levels of health-promoting behaviors for adolescents (18); 4) enhancing health literacy through school-based interventions has great potential for improving students' access to and interpretation of health information (19). Adolescents spend most of their daily time in school, which means they can receive health education and learn how to improve healthy lifestyles and related skills through this setting (20, 21).

Health literacy is more challenging to understand for adolescents than that for adults. Researchers may have different understandings and underlying constructs when using the same definition. That is why there are such a large number of measurement tools of health literacy currently (22, 23), along with some newly-developed health literacy instruments (24). According to Mancuso (1), it is recommended to use specific assessment tools for a specific age group in a specific context. Studies measuring childhood and adolescent health literacy have been a research focus, particularly in the past five years (23). Ormshaw *et al.* (23) conducted a systematic review on measuring childhood and adolescent health literacy in 2011. They found 16 studies that were involved with health literacy measures in children and adolescents. The authors also identified 13 health topics and nine underlying components from existing health literacy instruments. However, the authors did not critically appraise health literacy indices explicitly regarding their validity and reliability. More importantly, the

authors did not assess the methodological quality of each included study. This may undermine the persuasiveness of its conclusion. To fill this knowledge gap, we aim to conduct a systematic review that examines studies' methodological quality and examine reliability and validity of each health literacy instrument, thus providing researchers with unbiased information about which instruments have good psychometric properties. The '*COnsensus-based Standards for the selection of health status Measurement INstruments*' (COSMIN) group has recently developed as a critical appraisal tool (a checklist) to evaluate the methodological quality of studies on measurement properties of health measurement instruments (25). These measurement properties are divided into three domains: reliability, validity, and responsiveness (26). According to the COSMIN checklist, it is possible and scientific to critically appraise and compare psychometric properties of health literacy instruments for children and adolescents.

In this protocol, our target population is adolescent. According to the definition of the WHO, adolescents are those people aged 10 to 19 years and young people aged 10-24 years (27, 28). Given that the term '*adolescent*', '*child*', '*youth*' and '*young people*' is closely related, and Erikson (29) reckoned that children between the ages of 6 and 12 years could learn, compete and co-operate with others, we define our target group as those aged 6-24 years old.

Objectives of the review

This review aims to identify which health literacy instruments have good psychometric properties for children and adolescents. Specifically, there are three objectives:

1) To examine the methodological quality of included studies that aim to measure health literacy in children and adolescents;

2) To examine the measurement properties (i.e. reliability; validity; responsiveness) of health literacy instruments in children and adolescents;

3) To compare the overall rating of measurement properties between each health literacy instrument used in children and adolescents.

Search strategy

Database and search terms

As the term '*health literacy*' was first coined in 1974, articles published from 1st,January 1974 to 30th May 2014 in all languages will be searched. Search strategies will be first designed and then be consulted with two librarian experts. Articles indexed in the following seven databases: Medline, Pubmed, Embase, PsycINFO, CINAHL, ERIC and Cochrane

Library will be searched. The search key terms are '*health literacy*' and '*assessment*' according to previously published studies (1, 23, 30, 31). Age group for '*child, adolescent and young adult*' will be defined in the database settings. The synonyms are listed in **Appendix Table 1**. These synonyms are connected by '*or*' and search strategies are completed by '*and*'.

Appendix Table 1 Searching terms in databases

Key term (1)	Key term (2)
health literacy	health literacy measur*
health AND literacy AND education	health literacy assess*
	health literacy evaluat*
	health literacy instrument*
	health literacy tool*

Other sources of literature

Searching other sources to identify relevant research including:

- Reference lists of identified studies;
- Reference lists of previous systematic reviews on health literacy (1, 23, 30-33).

Eligibility criteria for inclusion and exclusion

According to the guidelines recommended by Cochrane Handbook for systematic reviews (34), inclusion criteria will be addressed regarding population, intervention, comparison, outcome and study design (PICOS):

Inclusion criteria-Participants

The target group should be children and/or adolescents, any age from 6 to 24 years of age.

Inclusion criteria-Interventions and Comparators

As interventional studies are not our interest in this review, it is not applicable to set out guidelines for interventions and comparators

Inclusion criteria-Outcomes

The included studies must be involved with health literacy assessment for children and adolescents, that is, the study should specify the term '*health literacy*', and studies are included if they report on at least one or more attributes of the three measurement properties: 1) reliability; 2) validity; and 3) responsiveness.

Inclusion criteria-Study design

The article should be research-based and peer-reviewed paper including study aim, methods, and results. Also, the study aim should focus on health literacy instrument development or validation.

Exclusion criteria

Studies will be excluded if they are: 1) not focusing on the target group; 2) not focusing on the health literacy instrument development or tool validation; 3) not research-based and peer-reviewed papers including editorials, comments and letters; 4) not reporting findings or results regarding any one of the measurement properties.

Study selection

Search records will be kept including the names of databases searched, keywords, search timeframe, and the search results. All the electronic search results will be initially inputted into the bibliography software of EndNote X7 (Thomson Reuters, New York, NY), and other sources of literature results will be summarised in the print paper. This screening process will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (35). One reviewer will screen studies by titles and abstracts. Secondly, full copies of articles identified will be obtained for thorough screening according to the inclusion criteria by two reviewers independently. Any disagreements in reviewer selections will be resolved at a meeting.

Quality assessment

The methodological quality of each included study will be assessed by two reviewers independently using the COSMIN checklist (25). The checklist consists of nine boxes with 5-18 items concerning methodological standards for how each measurement property should be assessed. Four response options for each item of the COSMIN checklist are defined, representing '*excellent*', '*good*', '*fair*' and '*poor*' quality. An overall score for the methodological quality of a study will be determined for each measurement property separately, by taking the lowest rating of any items in a box ('*worst score counts*') (36). Discrepancies arise between the reviewers will be resolved through discussion, if necessary with a third independent person.

Data extraction

Data extraction will be performed along with the assessment of methodological quality using

the COSMIN checklist (25). In addition, information on the interpretability (e.g. norm scores, floor-ceiling effects, minimal important change of the instruments), generalisability (e.g. characteristics of the study population and sampling procedure), respondent and administrative burden, and forms of administration will be also collected because they are important characteristics of a measurement instrument (26, 37). The data will be entered in an electronic form. Where possible, authors of the original studies will be contacted to obtain essential missing or additional data. Two reviewers will independently extract the data. Consensus should be reached afterward, if necessary with a third independent person.

Data synthesis

The results of the quality of health literacy instruments will be assessed using Terwee's quality criteria (38), to see whether the results of the measurement attributes are '*positive*', '*negative*', or '*indeterminate*'. To summarise the overall ratings of the measurement properties of one health literacy instruments by different authors, the synthesis will be performed by combining the results of the quality of health literacy instruments, the results of methodological quality of health literacy measurement studies and the consistency of their results. The possible overall rating for a measurement property is '*positive*', '*indeterminate*', or '*negative*', accompanied by levels of evidence, similarly as was proposed by the Cochrane Back Review Group (39, 40). One reviewer will perform the data synthesis and a second reviewer will check the synthesised results. Discrepancies of the results will be resolved by discussion.

References

1. Mancuso JM. Assessment and measurement of health literacy: An integrative review of the literature. Nurs Health Sci 2009;11(1):77-89.
2. Simonds SK. Health education as social policy. Health Educ Monogr. 1974;2(1_suppl):1-10.
3. Speros C. Health literacy: concept analysis. J Adv Nurs. 2005;50(6):633-40.
4. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health literacy: a prescription to end confusion. Washington DC, USA: The National Academies Press; 2004.
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med. 2008;67(12):2072-8.
6. Parker RM, Williams MV, Weiss BD, Baker DW, Davis TC, Doak CC, et al. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA. 1999;281(6):552-7.
7. Nutbeam D. Health promotion glossary. Health Promot Int. 1998;13(4):349-64.
8. Mancuso JM. Health literacy: a concept/dimensional analysis. Nurs Health Sci. 2008;10(3):248-55.
9. Higgins JW, Begoray D, MacDonald M. A Social Ecological Conceptual Framework for Understanding Adolescent Health Literacy in the Health Education Classroom. Am J Commun Psychol. 2009;44(3-4):350-62.
10. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int. 2005;20(2):195-203.
11. Kickbusch I, Pelikan JM, Apfel F, Tsouros A. Health literacy: the solid facts. Copenhagen, Denmark: WHO Regional Office for Europe, 2013.
12. Australian Bureau of Statistics. 4233.0-Health Literacy, Australia, 2006. Canberra: Australian Bureau of Statistics; 2008 Jun 25 [cited 2014 Mar 23]; Available from:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4233.02006?OpenDocument.

13. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health. 2012;82(1):28-36.
14. Bacon JL. Adolescent sexuality and pregnancy. Current Opinion in Obstetrics and Gynecology. 2000;12(5):345-7.
15. Paus T. Mapping brain maturation and cognitive development during adolescence. Trends Cogn Sci. 2005;9(2):60-8.
16. Miles, SB, Stipek, D. Contemporaneous and Longitudinal Associations Between Social Behavior and Literacy Achievement in a Sample of Low-Income Elementary School Children. Child Dev. 2006;77(1):103-17.
17. Conwell, LS, O'Callaghan, MJ, Andersen, MJ, Bor, W, Najman, JM, Williams, G. Early adolescent smoking and a web of personal and social disadvantage. J Paediatr Child Health. 2003;39(8):580-5.
18. Chang LC. Health literacy, self-reported status and health promoting behaviours for adolescents in Taiwan. J Clin Nurs. 2011;20(1-2):190-6.
19. Gazmararian, JA, Curran, JW, Parker, RM, Bernhardt, JM, DeBuono, BA. Public health literacy in America: an ethical imperative. Am J Prev Med. 2005;28(3):317-22.
20. Brey RA, Clark SE, Wantz MS. Enhancing health literacy through accessing health information, products, and services: An exercise for children and adolescents. J Sch Health. 2007;77(9):640-4.
21. Kickbusch I. Health literacy: an essential skill for the twenty-first century. Health Educ. 2008;108(2):101-4.
22. Wu A, Begoray D, Macdonald M, Wharf Higgins J, Frankish J, Kwan B, et al. Developing and evaluating a relevant and feasible instrument for measuring health literacy of Canadian high school students. Health Promot Int. 2010;25(4):444-52.
23. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. Health Educ. 2013;113(5):433-55.
24. National Institutes of Health (NIH). Office of Behavior and Social Sciences Reasearch. Research Underway in Health Literacy Supported by NIH. 2014. Bethesda, Maryland: NIH; 2014 [cited 2014 Mar 3]; Available: https://obssr-archive.od.nih.gov/scientific_areas/social_culture_factors_in_health/health_literacy/research-underway-health.aspx.
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):539-49
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
27. World Health Organization. Health topics: Adolescent health. Geneva, Switzerland: WHO; 2014 [cited 2014 Mar 7]; Available from: http://www.who.int/topics/adolescent_health/en/.
28. World Health Organization Media Centre. Young people: health risks and solutions, Fact sheet N°345. Geneva, Switzerland: WHO Media Centre; 2014 [cited 2014 Mar 7]; Available: http://www.who.int/mediacentre/factsheets/fs345/en/.
29. Erikson EH. Childhood and society. 2nd ed. New York, USA: W. W. Norton & Company; 1963.
30. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol. 2011;64(4):366-79.
31. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155(2):97-107.
32. Sanders LM, Federico S, Klass P, Abrams MA, Dreyer B. Literacy and child health: a systematic review. Arch Pediatr Adolesc Med 2009;163(2):131-40.
33. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics. 2009;124(Supplement 3):S265-S74.
34. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated Mar 2011]. London, UK: The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org.
35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.
36. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2012;21(4):651-7.
37. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2002;11(3):193-205.
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.

39. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine. 2009;34(18):1929-41.
40. Van Tulder, M, Furlan, A, Bombardier, C, Bouter, L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine. 2003;28(12):1290-9.

Appendix 2. Search strategy for seven databases

This section has two parts for SEARCH STRATEGY. The first part focuses on the timeline of 1974 to 2014. The second part focuses on the timeline of May 2014 to Jan 2018.

Part 1:

1 MEDLINE (Web of Science) search strategy

MEDLINE database was searched using the Web of Science interface on 16/05/2014 for the period 1974 to 2014.

Basic search:

Set	Results	
# 1	500	MeSH HEADING: (health literacy) OR ((TITLE: (health literacy) OR MeSH HEADING:exp: (Health Literacy)) AND (TITLE: (education) OR MeSH HEADING:exp: (Educational Status) OR MeSH HEADINGS:exp: (/education) OR MeSH HEADING:exp: (Teaching) OR MeSH HEADING:exp: (Educational Status) OR MeSH HEADING:exp: (Education))) Refined by: MeSH HEADINGS: (ADOLESCENT OR YOUNG ADULT OR CHILD) Indexes=MEDLINE Timespan=1974-2014
# 2	3,880	TOPIC: ((((health) literacy assess* OR health literacy measur*) OR health literacy evaluat*) OR health literacy instrument*) OR health literacy tool*) Indexes=MEDLINE Timespan=1974-2014
# 3	352	#2 AND #1 Indexes=MEDLINE Timespan=1974-2014

2 PubMed search strategy

PubMed database was searched (Advanced search) on 16/05/2014 for the period 1974 to 16/05/2014.

Set	Results	
# 1	4910	Search (health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract]) Sort by: PublicationDate
# 2	3248385	Search (child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract]) Sort by: PublicationDate Because if we select age group including child, adolescent, and young adult, the newest papers such as published in 2014 will not be included, the reason maybe the database doesn't update properly. So we use these terms to identify.
# 3	1887	Search (health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*) Sort by: PublicationDate
# 4	581	Search ((((health literacy[MeSH Terms]) OR (health AND education AND literacy[Title/Abstract]))) AND ((health literacy assess* OR health literacy measur* OR health literacy evaluat* OR health literacy instrument* OR health literacy tool*))) AND ((child* OR adolescent* OR student* OR youth OR young people OR teen* OR young adult[Title/Abstract])) Filters: Publication date from 1974/01/01 to 2014/05/16 Sort by: PublicationDate

3 EMBASE (Ovid) search strategy

EMBASE database was searched using Ovid interface on 16/05/2014 for the period 1974 to current.

Using **.mp** as searching terms (Advanced Search):

Set	Results	
#1	6060	("health literacy" or (health and literacy and education)).mp.
#2	6043	limit 1 to yr="1974 -Current"
#3	671	limit 2 to (school child <7 to 12 years> or adolescent <13 to 17 years>)
#4	170	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	170	limit 4 to yr="1974 -Current"
#6	18	3 and 5

4 PsycINFO (EBSCO) search strategy

PsycINFO database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	786	health literacy OR (health AND literacy AND education)	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#2	133	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: School Age (6-12 yrs), Adolescence (13-17 yrs), Young Adulthood (18-29 yrs) Search modes - Boolean/Phrase
#3	133	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

5 CINAHL (EBSCO) search strategy

CINAHL database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	437	health literacy OR (health AND education AND literacy)	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#2	63	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 19740101-20140531; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase
#3	63	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*) AND (S1 AND S2)	Search modes - Boolean/Phrase

6 ERIC (EBSCO) search strategy

ERIC database was searched using EBSCO interface on 16/05/2014 for the period January 1974 to May 2014.

Advanced Search:

Set	Results		
#1	59	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#2	2,250	health literacy OR (health AND education AND literacy)	Limiters - Date Published: 19740101-20140531 Search modes - Boolean/Phrase
#3	59	S1 AND S2	Search modes - Boolean/Phrase

7 The Cochrane Library search strategy

The Cochrane Library database was searched on 30/05/2014 for the period January 1974 to May 2014.

Set	Results	Sub-database
#1	4	Cochrane Reviews: There are 4 results from 8483 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Cochrane Reviews'
#2	114	Trials: There are 114 results from 789657 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Trials'
#3	2	Methods Studies: There are 2 results from 15764 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Date from 1974 to 2014 in Methods Studies'
#4	120	

PART 2:

The above seven databases were searched using similar rationale as describe before for the timeframe of May 17 2014 to Jan 31 2018.

MEDLINE was searched using the Web of Science interface on 17/02/2018 for the period 2014 to 2018.

Basic search:

Set	Results	
# 5	**35**	#4 AND #3 Indexes=MEDLINE Timespan=2014-2018
# 4	**14,198**	MeSH MAJOR TOPIC:exp: ((((((child*) OR adolescent*) OR student*) OR youth) OR young people) OR teen*) OR young adult) Indexes=MEDLINE Timespan=2014-2018
# 3	**1,779**	#2 AND #1 Indexes=MEDLINE Timespan=2014-2018
# 2	**3,482**	((((TOPIC: (health literacy assess*) OR TOPIC: (health literacy measur*)) OR TOPIC: (health literacy instrument*)) OR TOPIC: (health literacy tool*)) ORTOPIC: (health literacy evaluat*)) Indexes=MEDLINE Timespan=2014-2018
# 1	**2,654**	((MeSH HEADING:exp: (health literacy) OR MeSH MAJOR TOPIC:exp: (health literacy)) OR TITLE: (health literacy)) OR MeSH MAJOR TOPIC: ((health) AND education) AND literacy) Indexes=MEDLINE Timespan=2014-2018

Pubmed was searched (Advanced search) on 17/02/2018 for the period 2014 to 31/01/2018.

Set	Results	
#6	**26**	Search (((((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract]))))) AND ((child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract])) Filters:Publication date from 2014/05/16 to 2018/01/31
#5	48	Search (((((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract]))))) AND ((child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract]))
#4	288	Search (((((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract]))) AND ((health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract]))
#3	288	Search (health literacy assess*[Title/Abstract] OR health literacy measur*[Title/Abstract] OR health literacy evaluat*[Title/Abstract] OR health literacy instrument*[Title/Abstract] OR health literacy tool*[Title/Abstract])
#2	1636528	Search (child*[Title/Abstract] OR adolescent*[Title/Abstract] OR student*[Title/Abstract] OR youth[Title/Abstract] OR young people[Title/Abstract] OR teen*[Title/Abstract] OR young adult[Title/Abstract])
#1	8495	Search (((health literacy[MeSH Terms]) OR health literacy[Title/Abstract]) OR (health[Title/Abstract] AND education[Title/Abstract] AND literacy[Title/Abstract])) OR (health[Title/Abstract] AND education[Title/Abstract] AND numeracy[Title/Abstract])

EMBASE was searched using Ovid interface on 17/02/2018 for the period 2014 to current.

Using **.mp** as searching terms (Basic Search):

Set	Results	
#1	11966	("health literacy" or (health and literacy and education)).mp.
#2	5862	limit 1 to yr="2014 -Current"
#3	639	limit 2 to (school child <7 to 12 years> or adolescent <13 to 17 years>)
#4	372	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	**26**	3 and 4

PsycINFO was searched using Ovid interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results	
#1	4331	("health literacy" or (health and literacy and education)).mp.
#2	2077	limit 1 to yr="2014 -Current"
#3	754	limit 2 to (100 childhood <birth to age 12 yrs> or 180 school age <age 6 to 12 yrs> or 200 adolescence <age 13 to 17 yrs> or 320 young adulthood <age 18 to 29 yrs>)
#4	216	(health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*).mp.
#5	**40**	3 and 4

CINAHL was searched using EBSCO interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results		
S1	health literacy OR ((health AND education AND literacy))	Limiters - Published Date: 20140501-20180131; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase	View Results (467)
S2	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Published Date: 20140501-20180131; Age Groups: Child: 6-12 years, Adolescent: 13-18 years Search modes - Boolean/Phrase	View Results (118)
S3	S1 AND S2	Search modes - Boolean/Phrase	View Results (118)

ERIC was searched using EBSCO interface on 17/02/2018 for the period May 2014 to Jan 2018.

Basic Search:

Set	Results		
S1	health literacy OR ((health AND education AND literacy))	Limiters - Date Published: 20140501-20180131 Search modes - Boolean/Phrase	View Results (292)
S2	health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool*	Limiters - Date Published: 20140501-20180131 Search modes - Boolean/Phrase	View Results (13)
S3	(S1 AND S2)	Search modes - Boolean/Phrase	View Results (13)

Cochrane Library was searched on 17/02/2018 for the period May 2014 to Jan 2018.

Set	Results	Sub-database
#1	2	Cochrane Reviews: There are 2 results from 10210 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Year from 2014 to 2018 in Cochrane Reviews'
#2	199	Trials: There are 199 results from 1121096 records for your search on 'health literacy in Title, Abstract, Keywords and child* OR adolescent* OR student* OR teen* OR youth OR young adult OR young people in Title, Abstract, Keywords and health literacy assess* or health literacy measur* or health literacy evaluat* or health literacy instrument* or health literacy tool* in Title, Abstract, Keywords , Publication Year from 2014 to 2018 in Trials'
#3	201	

Appendix 3. Quality criteria for measurement properties of health literacy instruments

Property	Rating	Quality criteria
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	-	MIC ≤ SDC OR MIC equals or inside LOA
Reliability	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80
	?	Neither ICC/weighted Kappa nor Pearson's r determined
	-	ICC/weighted Kappa < 0.70 OR Pearson's r < 0.80
Validity		
Content validity	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement
	-	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity		
Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	-	Factors explain < 50% of the variance
Hypotheses testing	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Responsiveness		
Responsiveness	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs

Note: AUC, Area Under the Curve; ICC, Intra-class Correlation Coefficient; LOA, Limits of Agreement; MIC, Minimal Important Change; SDC, Smallest Detectable Change. + positive rating; ? indeterminate rating; - negative rating.

Appendix 4. Levels of evidence for the overall rating of measurement properties

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodological quality

Note: + positive result; - negative result; ±conflicting result; ? unknown result.

Appendix 5. Reliability and validity results for included instruments

Appendix Table 1. The methodological quality of each study based on reliability for each health literacy instrument

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
NVS (Warsh *et al.*, 2014)	na	na	na	na	na	na
NVS (Driessnack *et al.*, 2014)	α=0.71 (n=47)	Poor	na	na	na	na
NVS (Hoffman *et al.*, 2013)	α=0.67 (n=229)	Poor	na	na	na	na
c-sTOFHLAd (Chang *et al.*, 2012)	α=0.85 (n=300) Item-total correlation=0.44-0.86	Fair	Correlation of test and retest was 0.95 (*P*<0.001)	Test-retest	1 week	Fair
TOFHLA (Chisolm and Buchanan, 2007)	na	na	na	na	na	na
s-TOFHLA (Hoffman *et al.*, 2013)	α=0.89 (n=229)	Poor	na	na	na	na
REALM-Teen (Davis *et al.*, 2006)	α=0.94 (n=388)	Poor	γ=0.98	Test-retest	1 week	Fair
REALM-Teen (Hoffman *et al.*, 2013)	α=0.92 (n=229)	Poor	na	na	na	
HLAB (Wu *et al.*, 2010)	α=0.92 (n=275) Understanding α=0.88 (n=275) Evaluating α=0.82 (n=275)	Fair	Concordance rate=95%	Inter-rater	na	Poor
MMAHL(Massey *et al.*, 2013)	α=0.83 (n=1208) Item-total correlation=0.39-0.74	Good	na	na	na	na
MHL (Levin-Zamir *et al.*, 2011)	α=0.74 (n=1316) Coefficient of reproducibility=0.84 Coefficients of scalability=0.54-0.80	Poor	na	na	na	na
DNT-39 (Mulvaney *et al.*, 2013)	α=0.93 (n=61)	Fair	na	na	na	na
DNT-14 (Mulvaney *et al.*, 2013)	α=0.82 (n=133) α=0.80 (n=61)	Fair	na	na	na	na

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
	α=0.83 (n=72)					
eHEALS (Norman and Skinner, 2006)	α=0.88 (n=664) Item-scale correlation coefficient=0.51-0.76	Fair	The correlations between administrations ranged 0.68-0.40.	Test-retest	Immediately after the intervention; 3-month; 6-month	Fair
CHC Test (Steckelberg *et al.*, 2009)	na	na	Cohen's Kappa was excellent for 277 ratings (κ=0.9-1.0), moderate or good for 31 ratings (κ=0.7-0.89) and poor for 5 ratings (κ=<0.7)	Inter-rater	na	Poor
HKACSS (Schmidt *et al.*, 2010)	Health knowledge χ^2=6.45, *P*=0.17 (n=852) Health communication α=0.73 (n=852) Health attitudes α=0.57 (n=852)	Excellent	na	na	na	na
HLAT-51 (Harper, 2014)	Goodness of fit statistic was calculated by each domain (CFI=0.33-0.88; TLI=0.66-0.84; RMSEA=0.09-0.17). The internal consistency statistic was not calculated.	Poor	na	na	na	na
HLAT-8 (Abel *et al.*, 2014)	α=0.64 (n=7097 for male) α=0.65 (n=331 for female)	Excellent	na	na	na	na
CHLT (Liu *et al.*, 2014)	α=0.87 (the entire scale); subscales α ranged 0.59 to 0.81	Fair	na	na	na	na
VOHL (Ueno *et al.*, 2014)	na	na	The kappa value of scoring among the dentists ranged from 0.60 tooth score to 0.70 for gingiva score.	Inter-rater	na	Fair
HAS-A (Manganello *et al.*, 2015)	α=0.77 (communication) α=0.73 (confusion) α=0.76 (understanding)	Fair	na	na	na	na
MaHeLi (Naigaga *et al.* 2015)	The person separation index for the original 20-item scale was 0.91 and α=0.92. After	Fair	na	na	na	na

Instrument	Internal consistency		Reliability			
	Result	COSMIN score	Result	Design	Time interval	COSMIN score
	item reduction, the person separation index for 12-item scale was 0.90.					
QuALiSMental (de Jesus Loureiro *et al.*, 2015)	α=0.55-0.72 (component 2 and 3) α=0.44-0.59 (component 4) α=0.60-0.82 (component 5)	Fair	na	na	na	na
FCCHL-AYAC (McDonald *et al.*, 2016)	α=0.73 (FHL) α=0.63 (IHL) α=0.85 (CHL)	Fair	na	na	na	na
ICHL (Smith *et al.*, 2016)	na	na	na	na	na	na
HELMA (Ghanbari *et al.*, 2016)	α=0.93 (the entire scale); subscales α ranged 0.61 to 0.89	Good	The intraclass correlation coefficient was 0.93.	Test-retest	Two weeks	Good
HLSAC (Paakkari *et al.*, 2016)	α=0.93 (the entire scale); subscales α ranged 0.69 to 0.77	Fair	The standardised stability estimate was 0.83.	Test-retest	Two weeks	Fair
REALM-TeenS (Manganello *et al.*, 2017)	α=0.82	Good	na	na	na	na
funHLS-YA (Tsubakita *et al.*, 2017)	α=0.75	Fair	na	na	na	na
HLS-TCO (Intarakamhang *et al.*, 2017)	α=0.70-0.82 for five subscales; KR-20=0.76 for health knowledge scale	Fair	na	na	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017)	α=0.88 (Knowledge) α=0.94 (Self-advocacy/ support) α=0.93 (Resiliency)	Fair	na	na	na	na
p_HLAT-8 (Quemelo *et al.*, 2017)	α=0.74 (the entire scale), subscales α ranged 0.41 to 0.71	Fair	na	na	na	na

Note: na, no information available. CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; FCCHL-AYAC, the Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, Functional Health Literacy; funHLS-YA, Functional Health

Literacy Scale for Young Adults; HAS-A, the Health Literacy Assessment Scale for Adolescents; HELMA, the Health Literacy Measure for Adolescents; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, The Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; IHL, Interactive Health Literacy; MaHeLi, the Maternal Health Literacy; MHL, the Media Health Literacy; MMAHL, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, the Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, the Rapid Estimate of Adolescent Literacy in Medicine Short Form; RMSEA, Root Mean Square Error of Approximation; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults; VOHL, the Visual Oral Health Literacy.

Appendix Table 2. The methodological quality of each study based on validity for each health literacy instrument

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
NVS (Warsh *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in item generation.	Poor	na	na	Hypotheses regarding correlation between scores of a comparator instrument of Gray Silent Reading Test (GSRT) and NVS were formulated before data collection. The NVS and GSRT scores were highly correlated (ρ=0.71, p<0.0001). The NVS score increased with child age (ρ=0.53, p<0.0001).	Fair	na	na
NVS (Driessnack *et al.*, 2014)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in item generation.	Poor	na	na	A moderate positive correlation was found between children's NVS scores and their age, and between children's NVS scores and their reports of books numbers (γ_s=0.43, p=0.003; γ_s=0.36, p=0.012, respectively), but not found with their parents' report of the number of children's books at home (γ_s=0.06, p=0.671).	Poor	na	na
NVS (Hoffman *et al.*, 2013)	A panel of heath literacy experts developed the NVS according to previous experience. The NVS was then refined after feedback from patients, interviewers, and data analysts. No target population is involved in	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.49 (p<0.01).	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	item generation.							
c-sTOFHLAd (Chang *et al.*, 2012)	The c-sTOFHLAd was translated from the short-version of TOFHLA according to translation procedures and was tested among 30 adolescents to ensure appropriateness.	Good	Confirmatory factor analysis was conducted to determine structural validity. One-factor model indicated an acceptable fit to the data according structural equation modelling analysis.	Fair	Convergent validity was measured between c-sTOFHLAd and the rapid estimate of adult literacy in medicine (REALM), with a correlation coefficient of 0.74 (p<0.001).	Fair	Semantic equivalence was measured by the content validity index (CVI). All items were rated by the experts as having a CVI>0.85. Thirty adolescents were chosen to determine and ensure the cultural congruence of the instrument.	Fair
TOFHLA (Chisolm and Buchanan, 2007)	The TOFHLA was developed from a literacy expert after reviewing commonly used hospital texts and a pilot test. No target population is involved in item generation.	Poor	na	na	The reading comprehension component was significantly correlated with the WRAT3 and the REALM (ρ=0.59, p<0.001; ρ=0.60, p<0.001 respectively), however, no correlation were found with the numeracy component (ρ=0.11, p=0.45; ρ=0.18, p=0.22 respectively).	Fair	na	na
s-TOFHLA (Hoffman *et al.*, 2013)	The s-TOFHLA was developed based on previous data analysis, perceived importance and frequency of the task in the healthcare settings.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.28 (p<0.01).	Fair	na	na
REALM-Teen (Davis *et al.*, 2006)	The REALM-Teen was developed based on a preliminary test and a structured interview among adolescents. And a	Good	na	na	Convergent validity was measured between REALM-Teen and the WRAT-3 (r=0.83) and SORT-R (r=0.93).	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	panel of experts reviewed the word list.							
REALM-Teen (Hoffman *et al.*, 2013)	The REALM-Teen was developed based on a preliminary test and structured interview among adolescents. And a panel of experts reviewed the word list.	Poor	na	na	Convergent validity was measured between NVS and the TerraNova academic achievement test, with a correlation coefficient of 0.40 (p<0.01).	Poor	na	na
HLAB (Wu *et al.*, 2010)	Previous experience and literature review were used to develop items; 10 students were pilot-tested for appropriateness of wording, content and format of the final instrument.	Good	na	na	Correlations were assumed between socio-demographic variables and the overall scores. Socio-demographics of gender, age when came to Canada to live, speaking a language other than English were correlated with the scores of HLAB (β=-0.18, p=0.004; β=-0.22, p=0.014; β=-0.20, p=0.008 respectively). No convergent validity is assessed.	Fair	na	na
MMAHL (Massey *et al.*, 2013)	Domains were established from literature review and focus group. Items were developed either using adaptation of existing relevant items or created by the research team.	Good	Explorative principal components factor analysis was conducted and 49.8% of the variance was accounted by 6 factors.	Good	na	na	na	na
MHL (Levin-Zamir *et al.*, 2011)	The face validity was discussed in the focus group during pilot test. The content validity was	Good	na	na	As hypothesised, MHL was associated with socio-economic determinants, particularly with gender (β=1.25, p<0.001) and	Good	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	analysed using theory and operational definitions of health literacy and media literacy, and adolescents were invited to write detailed, anonymous responses.				mother's education (β=0.16, *p*=0.04). In addition, MHL was also associated with health behaviours (β=0.03, *p*=0.05) and health empowerment (β=0.36, *p*<0.001).			
DNT-39 (Mulvaney *et al.*, 2013)	The DNT-39 was developed from the original 43-item version DNT-43 by eliminating questions specific to type 2 diabetes. An expert team developed the DNT-43 and refined it.	Poor	na	na	The DNT-39 was associated with WRAT-3 and parent education (ρ=0.40, *p*=0.001; ρ=0.29, *p*=0.028 respectively)	Fair	na	na
DNT-14 (Mulvaney *et al.*, 2013)	The DNT-14 was developed from the original 15-item version DNT-15 by eliminating 1 question specific to type 2 diabetes. An expert team developed the DNT-15 by data analysis from DNT-43.	Poor	na	na	The DNT-14 was associated with the Wide-Ranging Achievement Test (WRAT3), parent education, diabetes problem solving and HbA1c (ρ=0.36, *p*=0.005; ρ=0.31, *p*=0.019; ρ=0.27, *p*=0.023; ρ=-0.34, *p*=0.004 respectively)	Fair	na	na
eHEALS (Norman and Skinner, 2006)	The eHEALS was developed by the expert team and pilot-tested and refined by feedback from participants.	Good	Explorative principal components factor analysis was conducted and 56% of the variance was accounted by a single factor. The factor loadings ranged from	Fair	Correlations were assumed between eHEALS and other measured variables (gender, age, use of information technology overall, self-evaluations of health). However, only gender difference was found at baseline level of eHealth literacy	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
			0.60-0.84 among the 8 items.		(t=2.236, p=0.026). No convergent validity is assessed.			
CHC Test (Steckelberg *et al.*, 2009)	The CHC Test was developed by the research team and pre-tested by collecting qualitative data and quantitative field test.	Good	IRT test for determining dimensionality was performed.	Poor	na	na	na	na
HKACSS (Schmidt *et al.*, 2010)	The HKACSS items were taken from a previous health survey and selected basing on consideration of item content.	Good	na	na	As hypothesised, health communication, attitudes and self-efficacy were significantly related to each other (ρ=0.15-0.38, P<0.05). And children from higher educational background showed a better knowledge and communicated more about health topics (β=0.16, p<0.05).	Good	na	na
HLAT-51 (Harper, 2014)	The expert team evaluated the initial items using a 5-point Likert scale according to their research experience. And 144 college students were invited to complete a pilot test.	Good	Comprehension (CFI=0.80; TLI=0.78; RMSEA=0.09); health numeracy (CFI=0.57; TLI=0.48; RMSEA=0.09); media literacy (CFI=0.88; TLI=0.84; RMSEA=0.07); digital literacy (CFI=0.33; TLI=0.06; RMSEA=0.16); health information seeking (CFI=0.80; TLI=0.66; RMSEA=0.17)	Poor	na	na	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
HLAT-8 (Abel *et al.*, 2014)	The research team developed the HALT-8 drawing on literature review and their own experience. No target population is involved in item generation.	Poor	Explorative principal components factor analysis was conducted and 72.96% of the variance was accounted by four factors among male. In addition, the factor structure was validated using confirmatory factor analysis (CFI=0.99; TLI=0.97; RMSEA=0.03; SRMR=0.03).	Excellent	Hypotheses were formulated a priori regarding correlations between health literacy and gender, socio-cultural characteristics and health values. Results showed that female, higher educational status, and a stronger health valuation were associated with higher HL scores (p<0.05, respectively).	Good	na	na
CHLT (Liu *et al.*, 2014)	The research team developed the CHLT drawing on literature review, expert consultation and pilot test. 12 six graders were piloted about the instrument's readability.	Good	Confirmatory factor analysis was conducted to test the uni-dimensionality of each subscale. The factor loadings ranged from 0.20-0.58.	Fair	Hypotheses were formulated a priori regarding correlations between health literacy and gender, self-reported health and health behaviours. Results showed that female, better health status, normal BMI and healthy behaviours were positively associated with HL scores (p<0.05, respectively). Health-risky behaviours were negatively associated with health literacy scores (p<0.05).	Fair	na	na
VOHL (Ueno *et al.*, 2014)	na	na	na	na	Correlations were conducted between health literacy and gender. Results showed female students had higher gingiva scores than male students	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
					(p<0.05). However, no gender differences were found regarding tooth scores.			
HAS-A (Manganello *et al.*, 2015)	The research team developed the HAS-A drawing on literature review, expert consultation and pilot test. Scale items were piloted with undergraduates.	Good	Exploratory factor analysis was conducted and 41% of the variance was accounted by three factors.	Fair	Communication scale, confusion scale, and understanding scale were all correlated with the AURA scale (r=0.69, p<0.001; r=-0.50, p<0.001; r=-0.42, p<0.001). The correlation between communication scale, confusion scale and understanding scale and REALM-Teen and NVS were small, ranging from -0.26 to 0.08. Also health literacy scores were compared by demographics. There was no difference in scores by sex or age, but a significant difference by race/ethnicity (p<0.001).	Fair	na	na
MaHeLi (Naigaga *et al.* 2015)	The research team developed the MaHeLi based on the health belief model and integrated model of health literacy. No target population is involved in item generation.	Poor	The health-seeking behaviour (HSB) subscale brought substantial multidimensionality into the MeHeLi scale. After removing most items of the HSB subscale, the MeHeLi scale showed a uni-dimensionality	Fair	na	na	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
			construct with some but not too noticeable multi-dimensionality.					
QuALiSMental (de Jesus Loureiro *et al.*, 2015)	The questionnaire was developed based on mental health literacy framework and adapted among Portuguese adolescents and young people.	Excellent	Exploratory factor analysis was conducted for each component of the questionnaire. A five-factor solution explained 46.84% of the total variance for component 1 and 40.00% for components 2 and 3. A three-factor solution explained 47.24% of the variance for component 4 and a two-factor solution explained 55.63% for component 5.	Fair	The relationship between mental health components and mental health help-seeking intension was examined using a binary logistic regression analysis. Results showed higher levels of mental health literacy tended to associate with mental health-seeking intentions.	Fair	na	na
FCCHL-AYAC (McDonald *et al.*, 2016)	The instrument was adapted from the functional, communicative, and critical health literacy scale to be suitable for adolescents and young adults diagnosed with cancer.	Good	Exploratory factor analysis was conducted for the entire scale. The screen plot suggested the extraction of three factors (53.1% variance explained)	Fair	Health literacy scores were examined by gender, whether the measure was completed online or on paper, whether the participant was on or off treatment. Results showed no significant difference was found.	Fair	na	na
ICHL (Smith *et al.*, 2016)	The instrument was developed from formative interviews with 20 deaf/hard-of hearing high	Good	na	na	The relationship between ICHL and standard health literacy measures were examined. Result showed most ICHL items	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
	school students. Also the instrument was piloted with 18 individuals including content-expert and content-naïve deaf and hearing colleagues, teachers interpreters and students.				were related to health literacy skills instrument-short form, s-TOFHLA, and comprehensive heart disease knowledge questionnaire (p<0.05).			
HELMA (Ghanbari *et al.*, 2016)	All items were initially generated by in-depth interviews with 67 adolescents. Then items were assessed by an expert panel review and 16 adolescents.	Good	Exploratory factor analysis was conducted and 53.37% of the variance was accounted by eight factors.	Good	na	na	na	na
HLSAC (Paakkari *et al.*, 2016)	The research team developed the HLSAC drawing on literature review, expert review and pilot test. Scale items were piloted with 401 pupils (7th graders and 9th graders).	Good	The five-factor structure was tested using confirmatory factor analysis (RMSEA=0.08; CFI=0.96; TLI=0.92; SRMR=0.03). However, due to high correlations between factors, one-factor structure was finally determined (RMSEA=0.08; CFI=0.94; TLI=0.92; SRMR=0.04).	Fair	Correlations were assumed between the final 10-item scale and the original 15-item scale. Results showed the 10-item HLSAC predicted approximately 97% of the variance of the 15-item instrument.	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
REALM-TeenS (Manganello *et al.*, 2017)	This instrument was derived from the original 66-item REALM-Teen using the item response theory. Also, ten teenage patients were piloted.	Good	na	na	The REALM-TeenS scores were correlated with the REALM-Teen (r=0.92, p<0.001). Item fit analysis using the differential item functioning showed the REALM-TeenS functioned well for different groups of sex, race/ethnicity, and language spoken at home.	Good	na	na
funHLS-YA (Tsubakita *et al.*, 2017)	Items were generated from health materials that were frequently used in young adults and reviewed by the research team. No target population was involved in pilot test.	Poor	1-factor model was supported by the exploratory factor analysis.	Fair	The correlation between funHLS-YA and the comparator instrument of functional health literacy was 0.191 (p<0.001).	Fair	na	na
HLS-TCO (Intarakamhang *et al.*, 2017)	Items were developed from theories, documents and related research. Also, focus group and expert review were used to develop the instrument. 100 samples of overweight children were piloted.	Good	Confirmatory factor analysis was conducted for each subscale and results showed the model was acceptable, with factor loading ranging 0.39-0.73.	Fair	The path model of health literacy for obesity prevention behaviours was conducted using structural equation modelling. Results showed the hypothetical causal model was consistent with empirical data (chi-square=60.10, p=0.00, df=12, RMSEA=0.05, CFI=0.99; AGFI=0.99).	Fair	na	na
HLRS-Y (Bradley-Klug *et al.*, 2017)	Items were generated by focus group, expert review and a pilot test with 25 participants.	Excellent	Exploratory factor analysis was conducted, and results showed a three-factor structure of the instrument.	Fair	The relationships between health literacy scores and demographics were examined and results showed insurance type and knowledge, time since diagnosis and knowledge and	Fair	na	na

Instrument	Content validity		Structural validity		Hypotheses-testing		Cross-cultural validity	
	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score	Results	COSMIN score
					self-advocacy.			
p_HLAT-8 (Quemelo *et al.*, 2017)	The p_HLAT-8 was translated from the HLAT-8 according to translation procedures and was tested among 10 university students to ensure appropriateness.	Good	Confirmatory factor analysis was conducted, and results showed the 4-factor model fit was fair (CFI=0.97, GFI=0.98, TLI=0.95, RMSEA=0.03).	Fair	Convergent validity was examined for each sub-scale, but the results showed that only the factor 'search for information' was adequate. Discriminant validity was only adequate for two factors ('search for information' and 'understanding information').	Fair	Three experts in the field of health forward and backward translated the scale independently. Ten university students were piloted to test and ensure the cultural congruence of the scale. Confirmatory factor analysis showed a 4-factor structure fit the model.	Fair

Note: na, no information available. AGFI, Adjusted Goodness of Fit Index; AURA, Ask, Understand, Remember and Assessment; CFI, Comparative Fit Index; CHC Test, the Critical Health Competence Test; CHLT, Child Health Literacy Test; c-sTOFHLAd, the Chinese version of short-form Test of Functional Health Literacy in Adolescents; DNT, the Diabetes Numeracy Test; eHEALS, the eHealth Literacy Scale; FCCHL-AYAC, the Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, the Health Literacy Assessment Scale for Adolescents; HELMA, the Health Literacy Measure for Adolescents; HKACSS, the Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HLAB, Health Literacy Assessment Booklet; HLAT-8, the 8-item Health Literacy Assessment Tool; HLAT-51, the 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLSAC, The Health Literacy for School-aged Children; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; ICHL, Interactive and Critical Health Literacy; MaHeLi, the Maternal Health Literacy; MHL, the Media Health Literacy; MMAHLA, the Multidimensional Measure of Adolescent Health Literacy; NVS, the Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMenty, the Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, the Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, the Rapid Estimate of Adolescent Literacy in Medicine Short Form; RMSEA, Root Mean Square Error of Approximation; SRMR, Standardized Root Mean Square Residual; SORT-R, Slosson Oral Reading Test-Revised; s-TOFHLA, the short-form Test of Functional Health Literacy in Adults; TLI, Tucker-Lewis Index; TOFHLA, the Test of Functional Health Literacy in Adults; VOHL, the Visual Oral Health Literacy; WRAT-3, Wide-Range Achievement Test-Revised.

Appendix Table 3. The methodological quality of each study based on responsiveness for each health literacy instrument

Instrument	Responsiveness	
	Results	**COSMIN score**
VOHL (Ueno *et al.*, 2014)	Comparison of health literacy scores before and after health education showed both tooth and gingiva scores significantly increased after health education.	Fair

Note: As there was only one study examining the instrument's responsiveness, we only presented the instrument of VOHL. VOHL, the Visual Oral Health Literacy.

Research Checklist. PRISMA checklist for reporting systematic review

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Appendix 1 (CRD42018013759)
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8-9
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	10

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	10
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	11; Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11; 15-16; Table 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	25; Table 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	25; Table 4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	25; Table 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	25; Table 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	32-38
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38-39
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	39
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.