Quality of health literacy instruments used in children and adolescents: a systematic review

Shuaijun Guo,1,2 Rebecca Armstrong,1 Elizabeth Waters,1 Thirunavukkarasu Sathish,1,3 Sheikh M Alif,1,4 Geoffrey R Browne,1 Xiaoming Yu5

To cite: Guo S, Armstrong R, Waters E, et al. Quality of health literacy instruments used in children and adolescents: a systematic review. BMJ Open 2018;8:e020080. doi:10.1136/bmjopen-2017-020080

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 (eg, 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 (eg, 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.

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 WHO,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 health and maintain good health’. The literature has shown that health literacy is an independent and more direct predictor of health outcomes than sociodemographics.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 al8 and Robinson et al,9 health literacy interventions for children and adolescents can bring about improvements in healthy behaviours and decreased use of emergency department services. Although health literacy in young people has gained

Strengths and limitations of this study

- The COSMIN (Consensus-based Standards for the selection of Health Measurement Instruments) 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.
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’s4 4D model,14 15 health literacy in children and adolescents is mediated by four additional factors compared with 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 multidimensional concept with varying definitions.16 This paper uses the definition by Nutbeam,17 who states that health literacy consists of three domains: functional, interactive and critical. 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. 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 definitions17–19 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.10 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 al10 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 6 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 assess health information and take control over health determinants. Although health literacy is sufficiently explained in terms of its definitions17–19 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.10 21

METHODS

Following the methods for conducting systematic reviews outlined in the Cochrane Handbook,22 we developed a review protocol (see online supplementary appendix 1, PROSPERO registration number: CRD42018013759) prior to commencing the study. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement23 (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 from 17 May 2014 to 31 January 2018 (period 2). The databases searched were Medline, PubMed, Embase, PsycINFO, Cumulative Index to Nursing and Allied Health Literature, Education Resources Information Center, and the Cochrane Library. The search strategy was designed on the basis of previous reviews4 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^w’ OR ‘health literacy instrument^w’ OR ‘health literacy tool^w’; and (3) age-related terms: ‘child^w’ OR ‘adolescent^w’ OR ‘student^w’ OR ‘youth’ OR ‘young people’ OR ‘teen^w’ OR ‘young adult’.

No language restriction was applied. The detailed search strategy for each database is available in online supplementary appendix 2. As per the PRISMA flow diagram,23 the references from included studies and from six previously published systematic reviews on health literacy5 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 6–24—this broad age range was used because the age range for ‘children’ (under the age of 18) and ‘adolescents’ (aged 10–24) overlap,26 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 (1) the full paper was not available (ie, only a conference abstract or protocol was available); (2) they were not peer-reviewed (eg, dissertations, government reports); or (3) they were qualitative studies.

Selection process

All references were imported into EndNote VX7 software (Thomson Reuters, New York, New York) and duplicate records were initially removed before screening. Next, one author (SG) screened all studies based on the 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 (SG and SMA). 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 the characteristics of the included studies (eg, first author, published year and country), general characteristics of instruments (eg, health topics, components and scoring systems), methodological quality of the study (eg, internal consistency, reliability and measurement error) and ratings of measurement properties of included instruments (eg, internal consistency, reliability and measurement error). Data extraction from full-text papers published during period 1 was performed by two independent authors (SG and TS), whereas data extraction from full-text papers published during period 2 was conducted by one author (SG) and then checked by a second author (TS).

Methodological quality assessment of included studies

The methodological quality of included studies was assessed using the COmponents of Measurement Instrument Study (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 5–18 evaluating items. For example, ‘internal consistency’ is evaluated against 11 items. Each item is scored using a 4-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 (ie, ‘worst score counts’). Two authors (SG 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 (SG) 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 the quality criteria proposed by Terwee et al,35 who are members of the group that developed the COSMIN checklist (see online supplementary 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,36 ‘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 online supplementary 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..."
Open access

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 (SG) 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, 361 full-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).

Figure 1 Flow chart of search and selection process according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

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 11 studies carried out in the USA. The target population (aged 7–25) could be roughly classified into three subgroups: children aged 7–12 (n=5), adolescents aged 13–17 (n=20) and young adults aged 18–25 (n=4). Schools (n=17) were the most common recruitment settings, compared with clinical settings (n=8) and communities (n=4).

General characteristics of included instruments
Compared with previous systematic reviews,10 11 13 this review identified 19 additional new health literacy instruments (eHealth Literacy Scale (eHEALS), short-form Test of Functional Health Literacy in Adults (s-TOFHLA), Diabetes Numeracy Test (DNT)-39, DNT-14, 51-item Health Literacy Assessment Tool (HLAT-51), 8-item Health Literacy Assessment Tool (HLAT-8), Child Health Literacy Test (CHLT), Visual Oral Health Literacy (VOHL), Health Literacy Assessment Scale for Adolescents (HAS-A), Questionnaire for Assessment of
| Study no | Author (year) | Country | Target population | Health literacy instrument | Sample size ( % male) | Sampling method | Recruitment setting |
|----------|---------------|---------|-------------------|-----------------------------|----------------------|----------------|-------------------|
| 1        | Davis et al (2006)<sup>41</sup> | 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 programmes |
| 2        | Norman and Skinner (2006)<sup>43</sup> | 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 (2007)<sup>48</sup> | USA     | Young people aged 13–17 years (mean age=14.7) | TOFHLA                  | 50 (48.0)      | na              | Children’s hospital |
| 4        | Steckelberg et al (2009)<sup>49</sup> | Germany | Students in grades 10–11 and university | CHC Test                  | Sample 1: 322 (36.6) Sample 2: 107 (32.7) | na              | Secondary schools, university |
| 5        | Schmidt et al (2010)<sup>46</sup> | Germany | Children aged 9–13 years (mean age=10.4) | HKCSS                    | 852 (52.9)     | na              | Primary school |
| 6        | Wu et al (2010)<sup>50</sup>     | Canada  | Students in grades 8–12 | HLAB                     | 275 (48.0)      | Convenience sampling | Secondary schools |
| 7        | Levin-Zamir et al (2011)<sup>51</sup> | Israel  | Adolescents in grades 7, 9 and 11 (approximately aged 13, 15 and 17) | MHL                       | 1316 (52.0)     | Probability sampling and random cluster sampling | Public schools |
| 8        | Chang et al (2012)<sup>52</sup> | Taiwan  | Students in high school (mean age=16.01±1.02) | c-sTOFHLAd (2012)<sup>53</sup> | 300 (52.6)      | Multiple-stage stratified random sampling | High schools |
| 9        | Hoffman et al (2013)<sup>54</sup> | 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 (2013)<sup>55</sup> | USA     | Adolescents aged 13–17 years (mean age=14.8) | MMAHL                    | 1208 (37.6)     | Sampling from a large health insurance network | Public health insurance network |
| 11       | Mulvaney et al (2013)<sup>56</sup> | 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 (2015)<sup>57</sup>  | Switzerland | Young adults aged 18–25 years (male mean age: 19.6; female mean age=18.8) | HLA-T-8                  | 7428 (95.5)     | Sampling from compulsory military service for men and two-stage random sampling for women | Compulsory military service, communities |
| 13       | Driessnack et al (2014)<sup>58</sup> | USA     | Children aged 7–12 years | NVS                      | 47 (53.0)       | Convenience sampling | The science centre |
| 14       | Harper (2014)<sup>59</sup>     | New Zealand | Students aged 18–24 years | HLA-T-8                  | 144 (41.0)      | Purposeful sampling | College |
| 15       | Warsh et al (2014)<sup>60</sup> | USA     | Children aged 7–17 years (median age=11) | NVS                       | 97 (46.0)       | Convenience sampling | Paediatric clinics |
| 16       | Liu et al (2014)<sup>61</sup>  | Taiwan  | Children in grade 6 | CHLT                     | 162,609 (51.1)  | National sampling | Primary schools |
| 17       | Ueno et al (2014)<sup>62</sup> | 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 (2015)<sup>63</sup> | USA     | Youth aged 12–19 years (mean age=15.6) | HAS-A                    | 272 (37.0)      | Convenience sampling | A paediatric clinic and the community |
| 19       | Guttersturd et al (2015)<sup>64</sup> | Uganda | Pregnant adolescents aged 15–19 years | MaHeLi                   | 384 (0)         | Random sampling | Health centres |
| 20       | de Jesus Loureiro (2015)<sup>65</sup> | Portugal | Adolescents and young people aged 14–24 years (mean age=16.75±1.62) | QuALiSMental             | 4938 (43.3)     | Multistage cluster random sampling | Schools |

Continued
Mental Health Literacy (QuALiSMental), Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer (FCCHL-AYAC), Interactive and Critical Health Literacy (ICHHL), Health Literacy Measure for Adolescents (HELMA), Health Literacy for School-aged Children (HLSAC), Rapid Estimate of Adolescent Literacy in Medicine Short Form (REALM-TeenS), Functional Health Literacy Scale for Young Adults (funHLS-YA), Health Literacy Scale for Thai Childhood Overweight; HLSAC, Health Literacy for School-aged Children; ICHHL, Interactive and Critical Health Literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMaHL, Multidimensional Measure of Adolescent Health Literacy; na, no information available; 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-TOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; CHC Test, Critical Health Competence Test; CHLT, Child Health Literacy Test; 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; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; HLSAC, Health Literacy Scale for Adolescents; HLS-TCO, Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Booklet; HELMA, Health Literacy Assessment Tool; 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-TOFHLAd, short-form Test of Functional Health Literacy in Adolescents; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

Health literacy domains and components

Next, Nutbeam’s three-domain health literacy model 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, and one instrument measured only critical health literacy. There was one instrument measuring functional and interactive health literacy, one measuring functional and critical health literacy, and one measuring interactive and critical health literacy. Fifteen instruments measured health literacy by all three domains (functional, interactive and critical).
| 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<sup>36,52</sup> | 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 3min. | Interviewer-administered and performance-based. |
| 2  | TOFHLA<sup>20</sup> | 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 and 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 min (8.9–17.3 min). | Interviewer-administered and performance-based. |
| 3  | s-TOFHLA<sup>25</sup> | Functional HL 1. Reading comprehension (36). | Demographic patterns. | Instruction for preparation for an upper gastrointestinal series (fourth 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 and performance-based. |
| 4  | c-sTOFHLA<sup>41</sup> | Functional HL 1. Reading comprehension (36). | Developmental change. Demographic patterns. | Instruction for preparation for an upper gastrointestinal series (fourth 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. | 20min class period. | Self-administered and performance-based. |
| 5  | REALM-Teen<sup>41,90</sup> | Functional HL 1. Reading recognition (66). | Developmental change. Demographic patterns. | 66 health-related words such as weight, prescription and tetanus (sixth grade). | Open-ended. | Score range: 0–66; ordinal category: 0–37: ≤3rd; 38–47: 4th–5th; 48–58: 6th–7th; 59–62: 8th–9th; 63–66: ≥10th | 2–3min. | Interviewer-administered and performance-based. |
| 6  | HLAB<sup>40</sup> | 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 and performance-based. |
| 7  | MMAHL<sup>34</sup> | Functional, interactive and critical HL 1. Patient-provider encounter (4). 2. Interaction with the healthcare 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. | Experiences of how to access, navigate and manage one's healthcare and preventive health needs (sixth grade). | 5-point Likert scale. | Score range: na; continuous score. | na. | Self-administered and self-reported. |
| 8  | MHL<sup>39</sup> | 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 and multiple choice. | Score range: 0–24; continuous score. | na. | Video-assisted interviewer-administered and 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         | 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 and performance-based. |
| 10 | DNT-14         | 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 and performance-based. |
| 11 | eHEALS         | Functional, interactive and critical HL 1. Accessing health information (4). 2. Evaluating health information (2). 3. Applying health information (2). | Developmental change. Dependency. | General health topics about online health information (pilot-tested). | 5-point Likert scale. | Score range: na; continuous score. | na. | Self-administered and self-reported. |
| 12 | CHC Test       | 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. | Echinacea and common cold, MRI in knee injuries, treatment of acne, breast cancer screening (pilot-tested). | Open-ended and multiple choice. | na. | Less than 90 min. | Interviewer-administered and performance-based. |
| 13 | HKACSS         | Functional and interactive HL 1. Health knowledge (3). 2. Health attitudes (4). 3. Health communication (3). 4. Self-efficacy (3). | Developmental change. | Physical activities, nutrition, smoking, vaccination, oral health and general health (na). | 2 response options; 5-point Likert scale; 4-point Likert scale. | Score range: na; continuous score. | na. | Self-administered and performance-based and self-reported. |
| 14 | HLAT-51        | Functional, interactive and critical HL 1. Comprehension skill (20). 2. Health knowledge (11). 3. Media literacy (8). 4. Digital literacy (12). | Developmental change. | 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 min. | Self-administered and performance-based and self-reported. |
| 15 | HLAT-8         | 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. | 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 and self-reported. |
| 16 | CHLT           | Functional, interactive and critical HL 1. Health knowledge (11). 2. Health attitude (16). 3. Health skills (5). | Developmental change. | Personal hygiene, growth and ageing, 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 and performance-based. |
| 17 | VOHL           | Functional HL 1. Health knowledge (2). | Developmental change. | Oral health for tooth and gingiva (na). | Visual drawing. | Score range: 0–6; continuous score. | na. | Self-administered and 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 |
|----|---------------|--------------------------------------|------------------------------------------|---------------------------------------------|------------------|----------------|--------|--------------------|
| 18 | HAS-A**<sup>35</sup> | Functional, interactive and critical HL 1. Understanding health information (6). 2. Communication health information (5). 3. Confusion about health information (4).  | Developmental change. Demographic patterns. Dependency. | General health topics in daily life (pilot-tested). | 5-point Likert scale. | Score range: 0–24 (understanding), 0–20 (communication), 0–16 (confusion); continuous score. | na. | Self-administered and self-reported. |
| 19 | MaHeLi**<sup>37</sup> | 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 and self-reported. |
| 20 | QuALiSMental**<sup>41</sup> | 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. | na. | 40–50 min. |
| 21 | FCCHL-AYAC**<sup>36</sup> | 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 (second grade). | 4-point Likert scale. | Score range: 13–52; continuous score. | na. | Self-administered and self-reported. |
| 22 | IOHL**<sup>38</sup> | 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 min (together with other measures). | Self-administered and self-reported. |
| 23 | HELMA**<sup>39</sup> | 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 min. | Self-administered and self-reported. |
| 24 | HLSAC**<sup>40</sup> | 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 (seventh grade). | 4-point Likert scale. | na. | 45 min (together with the HBSC survey). | Self-administered and 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 |
|----|---------------|-------------------------------------|------------------------------------------|---------------------------------------------|------------------|----------------|--------|---------------------|
| 25 | REALM-TeenS⁴⁶ | Functional HL 1. Reading recognition (10). | Developmental change. Demographic patterns. | 10 health-related words such as diabetes (sixth grade). | Open-ended. | Score range: 0–10; ordinal category: 0–2: ≤3rd; 3–4: 4th–5th; 5–6: 6th–7th; 7–8: 8th–9th; 9–10: ≥10th. | 13.6s (range: 7.8–23.0). | Interviewer-administered and performance-based. |
| 26 | funHLS-YA⁴¹ | Functional HL 1. Word recognition and comprehension (19). | Developmental change. | Diseases and symptoms, nutrition and diet, biology of the human body (na). | Multiple choice. | Score range: na; continuous score. | 5min. | Self-administered and performance-based. |
| 27 | HLS-TCO⁶² | 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; fair: 21–27.99 for FHL, 33–43.99 for IHL; high: 28–35 for FHL, 44–54.9 for IHL, 36–45 for CHL. | na. | Self-administered and performance-based and self-reported. |
| 28 | HLRS-Y⁶³ | 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. | 15–20min. | Self-administered and self-reported. |
| 29 | p_HLAT-8⁶⁷ | Functional, interactive and critical HL 1. Understanding health information (2). 2. Searching health information (2). 3. Communicating health information (2). 4. 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 and self-reported. |

c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; CHC Test, Critical Health Competence Test; CHL, critical health literacy; CHLT, Child Health Literacy Test; DNT, Diabetes Numeracy Test; eHEALS, e-Health 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; 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; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; HLSAC, Health Literacy for School-aged Children; ICHL, Interactive and Critical Health Literacy; IHL, interactive health literacy; MaHeLi, Material Health Literacy; MHL, Media Health Literacy; MMHIL, Multidimensional Measure of Adolescent Health Literacy; na, no information available; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QUALSMental, Questionnaire for Assessment of Mental Health Literacy; 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.
Consideration of participants’ characteristics
As per Forrest et al’s 4D model, the 29 included instruments were examined for whether participant characteristics were considered when developing a new instrument or validating an existing instrument. The results showed most of the health literacy instruments considered developmental change, dependency and demographic patterns. In contrast, only seven instruments considered differential epidemiology.

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 (eg, general health topics, oral health or mental health), while only three instruments measured health literacy in the specific context of eHealth or media health. In relation to the readability of tested materials, only eight health literacy instruments reported their readability levels, ranging from 2nd to 19.5th grade.

Burden and forms of administration
The time to administer was reported in seven instruments, ranging from 3 to 90min. 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, 15 instruments were performance-based, 11 instruments were self-report and 3 included both performance-based and self-report items.

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, 8 studies assessed test–retest/inter-rater reliability, 2 studies assessed cross-cultural validity and only 1 study assessed responsiveness.

Evaluation of instruments’ measurement properties
After the methodological quality assessment of included studies, the measurement properties of each health literacy instrument were examined according to Terwee’s quality criteria (see online supplementary appendix 5). The rating results of the 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. 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.

DISCUSSION

Summary of the 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 with previous three systematic reviews, this review identified 19 additional new health literacy instruments and critically appraised the 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 (eg, 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 (ie, functional, interactive and critical), especially in the past 5 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, rather than a healthcare 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. 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 and Okan et al, 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...
literacy or media health literacy. Second, researchers had different research purposes for their studies. Some researchers used what were originally adult instruments to measure adolescent health literacy, whereas others developed new or adapted instruments. Third, the research settings affected the measurement process. As clinical settings were busy, short surveys were inconsistent. Some researchers measured general health literacy, while others measured eHealth literacy or media health literacy. Second, researchers had different research purposes for their studies. Some researchers used what were originally adult instruments to measure adolescent health literacy, whereas others developed new or adapted instruments.

| Health literacy instrument (author, year) | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cultural validity | Responsive ness |
|------------------------------------------|----------------------|-------------|-------------------|-----------------|---------------------|---------------------|-----------------------|-----------------|
| NVS (Hoffman et al, 2013)                 | Poor                 | na          | na                | Poor            | na                  | Fair                | na                    | na              |
| NVS (Driesnack et al, 2014)              | Poor                 | na          | na                | Poor            | Poor                | na                  | na                    | na              |
| NVS (Warsh et al, 2014)                  | na                   | na          | na                | Poor            | na                  | Fair                | na                    | na              |
| TOFLHA (Chisom and Buchanan, 2007)       | na                   | na          | Poor              | na              | na                  | Fair                | na                    | na              |
| s-TOFLHA (Hoffman et al, 2013)           | Poor                 | na          | Poor              | na              | na                  | Fair                | na                    | na              |
| c-sTOFLAld (Chang et al, 2012)           | Fair                 | Fair        | Good              | Good            | na                  | na                  | na                    | na              |
| REALM-Teen (Davis et al, 2006)           | Poor                 | Fair        | Good              | na              | na                  | na                  | na                    | na              |
| REALM-Teen (Hoffman et al, 2013)         | Poor                 | na          | Poor              | na              | na                  | na                  | na                    | na              |
| HLAB (Wu et al, 2010)                    | Fair                 | Poor        | na                | Good            | na                  | Fair                | na                    | na              |
| MMAHL (Massey et al, 2013)               | Excellent            | Good        | Good              | na              | na                  | na                  | na                    | na              |
| MHL (Levin-Zamir et al, 2011)            | Poor                 | na          | Good              | na              | Good                | na                  | na                    | na              |
| DNT-39 (Mulvane et al, 2013)             | Fair                 | na          | Poor              | na              | na                  | Fair                | na                    | na              |
| DNT-14 (Mulvane et al, 2013)             | Fair                 | na          | Poor              | na              | na                  | Fair                | na                    | na              |
| eHEALS (Norman and Skinner, 2006)       | Fair                 | Fair        | Good              | na              | na                  | Good                | na                    | na              |
| CHC Test (Steckelberg et al, 2009)       | na                   | Poor        | Good              | Poor            | na                  | na                  | na                    | na              |
| HKACSS (Schmidt et al, 2010)             | Excellent            | Good        | Good              | na              | na                  | na                  | na                    | na              |
| HLAT-51 (Harper, 2014)                   | Poor                 | na          | Good              | Poor            | na                  | na                  | na                    | na              |
| HLAT-8 (Abel et al, 2015)                | Excellent            | Poor        | Good              | Excellent       | Good                | na                  | na                    | na              |
| CHLT (Liu et al, 2014)                   | Fair                 | na          | Good              | Fair            | na                  | na                  | na                    | na              |
| VOHL (Ueno et al, 2014)                  | na                   | Fair        | na                | na              | na                  | Fair                | na                    | na              |
| HAS-A (Manganello et al, 2015)           | Fair                 | na          | Good              | Fair            | na                  | na                  | na                    | na              |
| MaHeLi (Guttersrud et al, 2015)          | Fair                 | na          | Poor              | na              | na                  | na                  | na                    | na              |
| QuALiSMental (de Jesus Loureiro, 2015)   | Fair                 | na          | Excellent         | na              | na                  | na                  | na                    | na              |
| FCCHL-AYAC (McDonald et al, 2016)        | Fair                 | na          | Good              | na              | na                  | na                  | na                    | na              |
| ICHL (Smith and Samar, 2016)             | na                   | na          | na                | Na              | na                  | na                  | na                    | na              |
| HELMA (Ghanbari et al, 2016)             | Good                 | Good        | Good              | Good            | na                  | na                  | na                    | na              |
| HLSAC (Paakkari et al, 2016)             | Fair                 | Fair        | Good              | Good            | na                  | na                  | na                    | na              |
| REALM-TeenS (Manganello et al, 2017)     | Good                 | na          | Good              | na              | na                  | na                  | na                    | na              |
| funHLS-YA (Taubakita et al, 2017)        | Poor                 | na          | Poor              | na              | na                  | na                  | na                    | na              |
| HLS-TCO (Intarakamhang and Intarakamhang, 2017) | Fair | na | Poor | na | na | Good | Fair | na | na |
| HLRS-Y (Bradley-Klug et al, 2017)        | Fair                 | na          | Excellent         | na              | na                  | na                  | na                    | na              |
| p_HLAT-8 (Quemel et al, 2017)            | Fair                 | na          | Excellent         | na              | na                  | na                  | na                    | na |
were more appropriate than long surveys. On the other hand, health literacy in school settings was often measured using long and comprehensive surveys. Fifth, researchers considered different participant characteristics when measuring health literacy in children and adolescents. For example, some

| Health literacy instrument (author, year) | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cultural validity | Responsiveness |
|------------------------------------------|----------------------|-------------|-------------------|-----------------|---------------------|-------------------|------------------------|-----------------|
| NVS (Hoffman et al, 2013)               | –                    | na          | na                | ?               | na                  | –                 | na                     | na              |
| NVS (Driessnack et al, 2014)            | +                    | na          | na                | ?               | na                  | +                 | na                     | na              |
| NVS (Warsh et al, 2014)                 | na                   | na          | na                | ?               | na                  | +                 | na                     | na              |
| TOFHLA (Chisolm and Buchanan, 2007)     | na                   | na          | na                | ?               | na                  | –                 | na                     | na              |
| s-TOFHLA (Hoffman et al, 2013)          | +                    | na          | na                | ?               | na                  | –                 | na                     | na              |
| c-sTOFHLAd (Chang et al, 2012)          | +                    | +           | na                | +               | ?                   | +                 | ?                      | na              |
| REALM-Teen (Davis et al, 2006)          | +                    | +           | na                | +               | na                  | +                 | na                     | na              |
| REALM-Teen (Hoffman et al, 2013)        | +                    | na          | na                | ?               | na                  | –                 | na                     | na              |
| HLAB (Wu et al, 2010)                   | +                    | +           | na                | +               | na                  | –                 | na                     | na              |
| MAAHL (Massey et al, 2013)              | +                    | na          | +                 | –               | na                  | na                | na                     | na              |
| MHL (Levin-Zamir et al, 2011)           | +                    | na          | na                | +               | na                  | +                 | na                     | na              |
| DNT-39 (Mulvaney et al, 2013)           | +                    | na          | na                | ?               | na                  | –                 | na                     | na              |
| DNT-14 (Mulvaney et al, 2013)           | +                    | na          | na                | ?               | na                  | –                 | na                     | na              |
| eHEALS (Norman and Skinner, 2006)       | +                    | –           | na                | +               | –                   | +                 | na                     | na              |
| CHC Test (Steckelberg et al, 2009)      | na                   | +           | na                | +               | na                  | na                | na                     | na              |
| HKACSS (Schmidt et al, 2010)            | + (HC) – (HA)        | na          | +                 | na              | +                   | na                | na                     | na              |
| HLAT-51 (Harper, 2014)                  | ?                    | na          | na                | +               | ?                   | na                | na                     | na              |
| HLAT-8 (Abel et al, 2015)               | –                    | na          | na                | ?               | +                   | +                 | na                     | na              |
| CHLT (Li et al, 2014)                   | +                    | +           | na                | +               | +                   | na                | na                     | na              |
| VOHL (Ueno et al, 2014)                 | na                   | – (TS) + (GS)| na                | na              | na                  | –                 | na                     | +               |
| HAS-A (Manganello et al, 2015)          | +                    | na          | na                | +               | +                   | –                 | na                     | na              |
| MaHeLi (Guttersrud et al, 2015)         | +                    | na          | na                | ?               | +                   | na                | na                     | na              |
| QuALiSMental (de Jesus Loureiro, 2015)  | –                    | na          | na                | +               | +                   | na                | na                     | na              |
| FCCHL-AAYAC (McDonald et al, 2016)      | + (FHL) – (IHL) + (CHL)| na      | na                | +               | –                   | +                 | na                     | na              |
| ICHL (Smith and Samar, 2016)            | na                   | na          | na                | +               | na                  | +                 | na                     | na              |
| HELMA (Ghanbari et al, 2016)            | +                    | +           | na                | +               | na                  | +                 | na                     | na              |
| HLSAC (Paakkari et al, 2016)            | +                    | +           | na                | +               | –                   | +                 | na                     | na              |
| REALM-TeenS (Manganello et al, 2017)    | +                    | na          | na                | +               | na                  | +                 | na                     | na              |
| funHLS-Y (Tsunabaki et al, 2017)         | +                    | na          | na                | +               | –                   | +                 | na                     | na              |
| HLS-TCO (Intarakamhang and Intarakamhang, 2017) | + | na          | na                | +               | +                   | +                 | na                     | na              |
| HLRS-Y (Bradley-Klug et al, 2017)       | +                    | na          | na                | +               | +                   | +                 | na                     | na              |
| p_HLAT-8 (Quemelo et al, 2017)          | +                    | na          | na                | +               | –                   | +                 | na                     | na              |

+, positive rating; –, negative rating; ?, indeterminate rating; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AAYAC, 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; ICHL, Critical Health Literacy Tool; HELAT-51, 51-item Health Literacy Assessment Tool; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; HLSAC, Health Literacy for School-aged Children; 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; na, no information available; 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 Adolescents 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.
researchers took considerations of students' cognitive development, some focused on adolescents' resources and environments (eg, friends and family contexts, eHealth contexts, media contexts), and others looked at the effect of different cultural backgrounds and socioeconomic status. Based on Forrest et al's 4D model, this review showed that most health literacy instruments considered participants' development, dependency and demographic patterns, with only seven instruments considering differential epidemiology.

| Health literacy instrument | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cultural validity | Responsiveness |
|----------------------------|----------------------|-------------|-------------------|------------------|---------------------|--------------------|------------------------|----------------|
| NVS39 50 52 | ? | na | na | ? | na | ± | na | na |
| TOFHLA48 | na | na | na | ? | na | - | na | na |
| s-TOFHLA50 | ? | na | na | ? | na | - | na | na |
| c-sTOFHLAd51 | + | + | na | ++ | ? | + | ? | na |
| REALM-Teen41 50 | ? | + | na | ++ | na | + | na | na |
| HLAB40 | + | ? | na | ++ | na | - | na | na |
| MMAHL48 | ++ | na | na | ++ | - | - | na | na |
| MHL49 | ? | na | na | ++ | na | + | na | na |
| DNT-39 51 | + | na | na | ? | na | - | na | na |
| DNT-14 53 | + | na | na | ? | na | - | na | na |
| eHEALS43 | + | na | na | ? | na | - | na | na |
| CHC Test47 | na | ? | na | ++ | ? | na | na | na |
| HKACSS50 | +++ (HC) – – – (HA) | na | na | ++ | na | ++ | na | na |
| HLAT-5142 | ? | na | na | ++ | ? | na | na | na |
| HLAT-845 | – – – | na | na | ? | +++ | ++ | na | na |
| CHLT54 | + | na | na | ++ | + | + | na | na |
| VOHL56 | na | – (TS) + (GS) | na | na | na | – | na | + |
| HAS-A56 | + | na | na | ++ | + | - | na | na |
| MaHeLi57 | + | na | na | ? | + | na | na | na |
| QuaLIS Mental54 | – | na | na | +++ | + | + | na | na |
| FCCHL-AYAC66 | + (FHL) – (HL) + (CHL) | na | na | ++ | + | – | na | na |
| ICHL58 | na | na | na | ++ | na | + | na | na |
| HELMA59 | ++ | na | na | ++ | ++ | na | na | na |
| HLSAC60 | + | na | na | ++ | - | + | na | na |
| REALM-TeenS66 | ++ | na | na | ++ | na | + | na | na |
| funHLS-YA61 | + | na | na | ? | + | - | na | na |
| HLS-TCO62 | + | na | na | ++ | + | + | na | na |
| HLSRS-Y63 | + | na | na | +++ | + | + | na | na |
| p_HLAT-867 | + | na | na | ++ | + | - | + | na |

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

+ or –, limited evidence and positive/negative result; ++ or – – –, strong evidence and positive/negative result; ±, conflicting evidence; ±, unknown due to poor methodological quality or indeterminate rating of a measurement property; c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; CHC Test, Critical Health Competence Test; CHL, Critical Health Literacy; CHLT, Child Health Literacy Test; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Youth and Young Adults; 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; HLS-TCO, Health Literacy Scale for Thai Children; HLSAC, Health Literacy Scale for School-aged Children; 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; na, no information available; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuaLIS Mental, 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.
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 Newest Vital Sign (NVS), the Test of Functional Health Literacy in Adults (TOFHLA), the s-TOFHLA, the DNT-99 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 unidimensionality analysis for internal consistency. As explained by the COSMIN group,76 a set of items can be inter-related and multidimensional, whereas unidimensionality 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 unidimensionality analysis (eg, factor analysis).

Critical appraisal of measurement properties for included instruments

This review demonstrated that of all instruments reviewed, three instruments (the Chinese version of short-form Test of Functional Health Literacy in Adolescents (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. 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 with the overall reliability rating of the s-TOFHLA,50 the c-sTOFHLAd showed better results.51 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.71 Second, the overall rating results of content validity for all included instruments were similar (ie, 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.72 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 al26 this review demonstrated that only one included study contained evidence of responsiveness. Ueno et al55 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 Multidimensional Measure of Adolescent Health Literacy44) discussed the potential to measure health literacy over time. Studies that measure health literacy over time in populations are needed, because this is a prerequisite for longitudinal studies and 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,74 75 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.76 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 overestimated results.77 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.69 The difference between self-report and performance-based instruments of health literacy has been discussed in the literature,28 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 Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen), the NVS, the s-TOFHLA, the c-TOFHLAd, the eHEALS, the Critical Health Competence Test (CHC Test), the Health Knowledge, Attitudes, Communication and Self-efficacy Scale (HKACSS), the Health Literacy Assessment Booklet (HLAB), the Media Health Literacy (MHL), the HLAT-51, the CHLT, the VOHL, the QuaLiSMental, the HELMA, the HLSAC, the funHLSYA, 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, 6 tested functional health literacy (the REALM-Teen, the NVS, the s-TOFHLA, the c-TOFHLAd, the CHC Test), 1 measured critical health literacy (the CHC Test), 1 measured functional and interactive health literacy (the HKACSS), 1 examined critical health literacy (the CHC Test), 1 examined critical 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 construct. Meanwhile, although the p_HLAT-8 (Portuguese version) is not available in English, the original HLAT-8 is. After comparing measurement domains and measurement properties, 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 peers15; (2) it is a short but comprehensive tool that captures Nutbeam’s three-domain nature of health literacy17; (3) it showed satisfactory structural validity (RMSEA=0.03; CFI=0.99; TLI=0.97; SRMR=0.03)16; and (4) it has good feasibility (eg, 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 used yet in school-based studies, 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. 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 ‘hypothetical 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 (cf. Ormshaw et al, Perry and 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.

Acknowledgements
The authors 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 SMA screened the literature. SG and TS extracted the data. SG drafted the manuscript. SG, GRB, RA, EW, XY, SMA and TS reviewed and revised the manuscript. All authors contributed to the final manuscript.

Funding
This paper is part of SG’s PhD research project, which is supported by the Melbourne International Engagement Award. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests
None declared.

Patient consent
Not required.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data sharing statement
There are no additional data available.

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

REFERENCES
1. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008;67:2072–8.
2. Nutbeam D. Health Promotion Glossary. Health Promot Int 1998;13:349–64.
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31(Suppl 1):19–26.
4. Squiers L, Peinado S, Berkman N, et al. The health literacy skills framework. J Health Commun 2012;17(Suppl 3):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;155:97–107.
6. Donovan S, Simpson, 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 JA. Health literacy and adolescents: a framework and agenda for future research. Health Educ Res 2008;23:840–7.
8. Diamond C, Saintonge S, August P, et al. The development of building wellness, a youth health literacy program. J Health Commun 2011;16(Suppl 3):103–18.
9. Robinson LD, Calma R, Bazmarg M. The impact of literacy enhancement on asthma-related outcomes among underserved children. J Natl Med Assoc 2008;100:892–6.
10. Ormshaw MJ, Paakkari LT, Kannas LK. Measuring child and adolescent health literacy: a systematic review of literature. Health Educ 2013;113:433–55.
11. Perry EL. Health literacy in adolescents: an integrative review. J Spec Pediatr Nurs 2014;19:210–8.
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;18:166.
14. Forrest CB, Simpson L, Clancy C. Child health services research. Challenges and opportunities. JAMA 1997;277:1789–93.
15. Rothman RL, Yin HS, Mulvany S, et al. Health literacy and quality: focus on chronic illness care and patient safety. Pediatrics 2009;124(Suppl 3):S315–26.
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;12:80.
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;15:259–67.
18. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15(Suppl 2):9–19.
19. Nutbeam D. Defining and measuring health literacy: what can we learn from literacy studies? Int J Public Health 2006;54:303–5.
20. Abel T. Measuring health literacy: moving towards a health-promotion perspective. Int J Public Health 2008;53:169–70.
21. Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. J Health Commun 2014;19:1481–96.
22. Higgins JP, Green S. In: Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.1. 0 [updated Mar 2011], 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;151:264–9.
24. Simonds SK. Health education as social policy. Health Educ Monogr 1974;21(suppl):1–10.
25. Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. Nurs Health Sci 2009;11:77–89.
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;64:368–79.
27. Sanders LM, Federico S, Klass P, et al. Literacy and child health: a systematic review. Arch Pediatr Adolesc Med 2009;163:131–40.
