



School health assessment tools: a systematic review of measurement in primary schools

Maryam Kazemitabar¹, Ali Moghadamzadeh¹, Mojtaba Habibi^{2,3}, Rezvan Hakimzadeh¹ and Danilo Garcia^{4,5,6,7,8}

¹ Department of Curriculum and Instruction, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran

² Health Promotion Research Center, Iran University of Medical Sciences, Tehran, Iran

³ Department of Health Psychology, School of Behavioral Sciences and Mental Health (Tehran Institute of Psychiatry), Iran University of Medical Sciences, Tehran, Iran

⁴ Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

⁵ Blekinge Center of Competence, Region Blekinge, Karlskrona, Sweden

⁶ Centre for Ethics, Law and Mental Health, University of Gothenburg, Gothenburg, Sweden

⁷ Department of Psychology, University of Gothenburg, Gothenburg, Sweden

⁸ Network for Well-Being, Sweden

ABSTRACT

Background. This systematic review aimed to investigate the psychometric properties of the school health's assessment tools in primary schools through COSMIN Risk of Bias checklist. We examined the studies that have addressed the measurement properties of school-health instruments to give a clear overview of the quality of all available tools measuring school health in primary schools. This systematic review was registered in PROPERO with the Registration ID: CRD42020158158.

Method. Databases of EBSCOhost, PubMed, ProQuest, Wiley, PROSPERO, and OpenGrey were systematically searched without any time limitation to find all full-text English journal articles studied at least one of the COSMIN checklist measurement properties of a school-health assessment tool in primary schools. The instruments should be constructed based on a school health model. The eligible studies were assessed by COSMIN Risk of Bias checklist to report their quality of methodology for each measurement property and for the whole study by rating high, moderate or low quality.

Results. At the final screening just seven studies remained for review. Four studies were tool development, three of them were rated as “adequate” and the other study as “very good”; five studies examined the content validity, three of them were appraised as “very good”, and the two remaining as “inadequate”. All seven studies measured structural validity, three of them were evaluated as “very good”, three other were scored as “adequate”, and the last study as “inadequate”. All the seven studies investigated the internal consistency, five of them were assessed as “very good”, one was rated as “doubtful”, and the last one as “inadequate”. Just one study examined the cross-cultural validity and was rated as “adequate”. Finally, all seven studies measured reliability, two of them were rated as “very good” and the rest five studies were appraised as “doubtful”. All rating was based on COSMIN checklist criteria for quality of measurement properties assessment.

Conclusion. The number of studies addressing school health assessment tools was very low and therefore not sufficient. Hence, there is a serious need to investigate

Submitted 5 February 2020

Accepted 9 June 2020

Published 21 July 2020

Corresponding authors

Maryam Kazemitabar,
maryam.kazemi64@ut.ac.ir,
ma.kazemi64@gmail.com
Danilo Garcia,
danilo.garcia@icloud.com

Academic editor

Yuriy Orlov

Additional Information and
Declarations can be found on
page 15

DOI 10.7717/peerj.9459

© Copyright
2020 Kazemitabar et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

the psychometric properties of the available instruments measuring school health at primary schools. Moreover, the studies included in the present systematic review did not fulfill all the criteria of the COSMIN checklist for assessing measurement properties. We suggest that future studies consider these criteria for measuring psychometric properties and developing school health assessment tools.

Subjects Health Policy, Pediatrics, Psychiatry and Psychology, Public Health, Science and Medical Education

Keywords School health, Psychometric properties, Reliability and validity, Health-promoting schools, School health assessment tools, COSMIN risk of bias checklist, Methodological quality assessment, Systematic review, School health in primary schools, Psychometric properties

BACKGROUND

Childhood consists of the golden years of everyone's life influencing many aspects of the rest of their lives among which *health* is of utmost importance. That is why the impact of childhood health status on adulthood health has been widely investigated (e.g., [Wang & Shen, 2016](#); [Ballard et al., 2015](#); [Péneau et al., 2017](#); [Kalmakis & Chandler, 2015](#); [Mehlhausen-Hassoen & Winstok, 2019](#)). In this context, the World Health Organization defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” ([World Health Organization, 1948](#)) and emphasizes the necessity of sufficient healthcare during childhood ([Bravo-Sanzana, Salvo & Mieres-Chacaltana, 2015](#)). Since children spend a tremendous time at school, the significance of providing a healthy environment at school in all aspects seems undeniable ([Pedersen, 2019](#); [Albert et al., 2019](#)). Moreover, research shows that healthy students learn better and they acquire better educational outcomes ([Neumann et al., 2017](#); [Lindegaard Nordin, Jourdan & Simovska, 2019](#); [Anderson et al., 2017](#); [Scott & Karberg, 2016](#)).

In this regard, [Allensworth & Kolbe \(1987\)](#) have presented eight components of the Coordinated School Health Program to promote health at schools, these components are: health education (1), physical education (2), health services (3), nutrition services (4), counseling, psychological and social services (5), healthy environment (6), school-site health promotion for staff (7), and family and community involvement (8) ([Lohrmann, 2008](#)). Furthermore, the World Health Organization has developed the Health-Promoting Schools framework in 1996 ([World Health Organization, 1996](#)), which introduced three main characteristics of School Curriculum, Ethos and/or Environment, and Families and/or Communities ([Langford et al., 2015](#)) which is divided into six key Health-Promoting School factors of healthy school policies, school's physical environment, school's social environment, community links, action competencies for healthy living, and school health care and health promotion ([World Health Organization, 2017](#)). Later the Whole School, Whole Community, Whole Child model released in 2014 included 11 components of health services, health education, employee wellness, counseling, psychological and social services, nutrition environment and services, physical education and physical activity, physical environment, social emotional climate, family engagement, and community involvement

(Lewallen et al., 2015). This model has endeavored to link health and educational outcomes and was a combination of the Coordinated School Health model and the Association for Supervision and Curriculum Development's Whole Child approach. Furthermore, the *National Association of School Nurses (2018)* has introduced the Framework for the 21st Century School Nursing Practice, a student-centered nursing care approach that connects students, family and school community. This framework is consisted of five components including standards of practice (1), care coordination (2), leadership (3), quality improvement (4), and community/public health (5) (*National Association of School Nurses, 2016*).

As briefly reviewed here, there have been various but similar models suggesting the components or factors of school health, thus, the question is if there are instruments to measure these factors. Measuring the factors of school health is important; it helps the health specials, school nurses, school principals, teachers, parents, politicians, and other stakeholders to realize what they need to improve in order to effectively promote student health, which in turn leads to better educational outcomes (*Best, Oppewal & Travers, 2018; Rahman et al., 2018*) promotes public health (*Kolbe, 2019; Birch & Auld, 2019*), improves mental health (*Deborah, 2019; Redfern et al., 2019; Holt, 2020*), increases health equity (*Peng et al., 2019; González, Etow & De La Vega, 2019*), promotes general health (*Boroumandfar et al., 2015; Krok-Schoen et al., 2018; Mishra et al., 2018*), prevents diseases (*Akihiro et al., 2017; Park et al., 2017; Jihene et al., 2015*), promotes physical activities (*Kelly et al., 2019; Dai, 2019*), healthier nutritional choices (*Anita, 2019; Shrestha et al., 2019*), and increases student's safety (*Voon & Ariff, 2019; Chalupka & Anderko, 2019; Mannathoko, 2019*), among many other benefits.

In this context, measuring the psychometric properties of instruments plays an important role in research, because it shows how valid and reliable a tool is and it helps researchers to choose the best tool available for their studies. In other words, the quality of tools or instruments is directly related to their psychometric properties (*Finch, 2002; Roach, 2006*). The most important psychometric properties are reliability and validity (*Terwee et al., 2007*). Low reliability of a tool means it lacks of generalizability power and low validity indicates the tool is not capable of measuring the intended construct (*Mauch, Rist & Kaelin, 2017*). In other words, in order for school health instruments to be of any use for health specials, school nurses, school principals, teachers, parents, politicians, and other stakeholders, the instrument needs to have sound psychometric properties.

In sum, due to the importance of good health during childhood and the necessity to measure school health using appropriate tools with good psychometric properties, this systematic review aimed to (1) identify existing instruments that measure the factors of school health at primary schools, and (2) evaluate these measures' psychometric properties. All studies have been screened for the risk of bias and quality of measurement using the modified COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. This review will help to clarify whether the current instruments, tools, scales, and indexes measure the school health's factors.

METHODS

Protocol and registration

The review protocol of the study used the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines. This review was registered with the International Prospective Registry of Systematic Reviews (PROSPERO) (Registration ID: CRD42020158158) before starting the research according to the PRISMA guideline recommendation (*Moher et al., 2009*).

Eligibility criteria

The criteria for the selection of the studies included in this systematic review were that the studies: (1) developed or evaluated an instrument to measure school health; (2) the instruments should measure the school health factors; (3) evaluated at least one of the psychometric properties described in COSMIN Risk of Bias checklist for a school health instrument; (4) the population of the study should be primary school children; (5) were original research journal articles (i.e., book chapters, thesis and case studies were excluded); and (6) were in English. All articles that were not available online in full text were also excluded from the study.

Search procedures

Two of the authors (Maryam Kazemitabar and Mojtaba Habibi) performed, independently from each other, a systematic search strategy to find articles with the specified criteria. The articles were divided into three groups of “completely related”, “somewhat related”, and “not related”. The authors selected the articles based on the relatedness of the title and abstract of the articles, then the full texts of those placed in the “completely related” and “somewhat related” groups were studied. In case of any disagreement, the second author (Ali Moghadamzadeh) intervened until consensus was reached.

The studies were searched through the EBSCO(host), ProQuest, PubMed, and Wiley databases, also a gray literature search was performed in PROSPERO, OpenGrey, and Google Scholar. No publication date was considered. Each database was queried from November 2019 using the Boolean operators (AND/OR). The terms searched were: “school health”, “health-promoting schools”, “coordinated school health”, “psychometric properties”, “reliability”, “validity”, “tool”, “scale”, “index”, “instrument”, “evaluation”, and “questionnaire”. We have also conducted hand search on the references lists of the selected studies. [Table 1](#) shows the databases and their search algorithms applied to the search strategy.

We have sent an Email to the corresponding author of one of the articles to access the full-text of it, but we have not received any reply so it has been excluded from this Systematic Review. The search for articles was finalized by December 2019. A total of 649 studies were identified for this review.

Study selection

From the 649 articles, we excluded all articles that did not report the psychometric properties of the instrument (cf. *Brener, Pejavarra & McManus, 2011*; *Chen & Lee, 2016*;

Table 1 Databases and their search algorithms.

Databases	Search algorithm	Articles founded
EBSCO(host)	TI (“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity) AND AB (“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)	118
ProQuest	ab(“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity) AND ti(“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)	38
PubMed	((“school health”)[Title] AND (scale[Title] OR index[Title] OR tool[Title] OR instrument[Title] OR questionnaire[Title] OR evaluation[Title] OR “psychometric properties”[Title] OR reliability[Title] OR validity)[Title])) AND (“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)[MeSH Terms])) AND (“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)	503
Wily	(“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)	55
PROSPERO	“school health“	65
OpenGrey	(“school health”) AND (scale OR index OR tool OR instrument OR questionnaire OR evaluation OR “psychometric properties” OR reliability OR validity)	173
Total		952

Burt et al., 1996; Weiler & Pigg Jr, 2004) or that did not include middle or high school students (*Kristiansen, Holmstrom & Olofsson, 2016; Yun et al., 2018; Sagatun et al., 2019*). After this procedure, seven articles were selected as clearly relevant for inclusion in the systematic review.

Data extraction and processing

The selected studies ($n = 7$) were assessed by the modified COSMIN checklist, which was developed to assess the methodological quality of studies on measurement properties. The original COSMIN checklist contains 10 parameters. In the modified version of the COSMIN checklist used in this systematic review, however, the parameters “*criterion validity*” and “*responsiveness*” were omitted because none of the studies reported these parameters. Thus, eight parameters were used in the present systematic review: PROM development, content validity, structural validity, internal consistency, cross-cultural

validity/measurement invariance, reliability, measurement errors, and hypotheses testing for construct validity.

Content validity refers to the degree to which the items in an instrument are an adequate reflection of the construct to be measured. Structural validity is the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured (Mokkink *et al.*, 2010). Cross-cultural validity refers to the degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument. Internal consistency refers to the degree of interrelatedness among the items and is often assessed by Cronbach's alpha (Prinsen *et al.*, 2018) and more recently by using coefficient Omega (Nima *et al.*, *in press*). Reliability refers to the overall consistency of a measure, a measure has a high reliability that results in similar scores under consistent conditions; in other words, a test is reliable if it gives the same repeated result under the same conditions. Measurement error is the difference between a measure quantity and its true value. There are two types: systematic error and random error. The appropriate statistics for calculating the measurement error are Standard Error of Measurement, Limits of Agreement, and Smallest Detectable Change (Prinsen *et al.*, 2018). Hypotheses testing for construct validity shows the extent of consistency between scores of the measure with hypotheses. It can be measured by comparing the instrument's score to other instruments or difference between relevant groups; in other words, studies that performed convergent and discriminative (divergent) validity or known-groups validity would be appraised as "very good".

The option "not applicable" applies for studies that have not investigated a specific psychometric property because it has not been applicable in that study and therefore were not included for scoring. Each parameter's check box has items to assess each specific psychometric property using a four-point scale ("very good", "adequate", "doubtful", and "inadequate"). The overall quality of each psychometric property is operationalized as the lowest rating of any standard in the box is taken (i.e., "the worst score counts" principle) (Mokkink *et al.*, 2018; Prinsen *et al.*, 2018; Terwee *et al.*, 2018). In this study each "very good" rating received score 1 and the options "adequate", "doubtful", and "inadequate" were received score 0. That is, the maximum score for a study was 8. The final quality of any individual study was rated based on high quality for any study with 50% of the scores or more, moderate quality for 30% to 50%, and low quality for below 30% of the scores.

A customized data extraction form was developed by the first reviewer (Maryam Kazemitabar), and then characteristics data of the studies and participants were extracted; the second reviewer (Mojtaba Habibi) also extracted the data independently, the results of the two extractions were compared and in case of any differences the third reviewer (Ali Moghadamzadeh) intervened to reach consensus. The data regarding characteristics of the selected studies included: first author name, year of publication, country, sample size, time to answer the items, number of items, response scale, participant age/grade, participant sex, and instrument factors.

The instruments studied in this systematic review were: Scale for Health-Promoting Schools (SHPS), Quality of Life in School, School Health Policies and Program Study

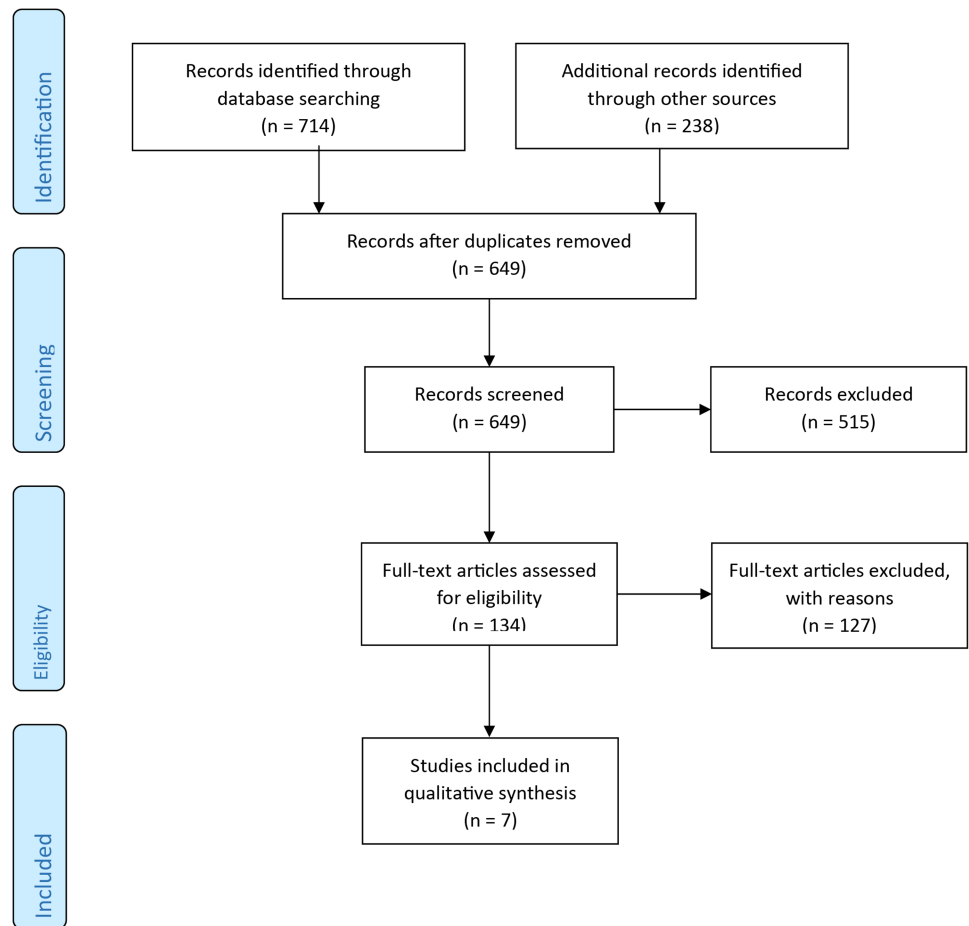


Figure 1 PRISMA flowchart of search outcomes.

[Full-size](#) [DOI: 10.7717/peerj.9459/fig-1](https://doi.org/10.7717/peerj.9459/fig-1)

2000 Questionnaire, Healthy School Indicator Tool (HSIT), and School Health Evaluation Instrument.

RESULTS

Study selection

The search strategy was conducted throughout 6 data bases: EBSCOhost, PubMed, ProQuest, Wiley, PROSPERO, and OpenGrey. This resulted in 952 articles; and then in 649 after removing the duplicates. After the screening of titles and abstracts 515 articles were selected; then the full-text of the articles were studied for investigating the eligibility criteria including original journal articles, articles that studies at least one of the COSMIN Risk of Bias checklist measure properties of a school health assessment tool, study's population should be primary schools, articles of English language and articles with full-text availability. After this procedure, seven articles remained for inclusion in this systematic review (Fig. 1).

Study characteristics

The following descriptive characteristics of the articles included in this systematic review was extracted: first author name, publication year, country, sample size, time to answer the items/number of items, response scale (multiple/forced choice, Likert scale, fill-in-the blank, true or false, yes or no, and text response), participant age/grade, participant sex, instrument's factors (Table 2).

Risk of bias of school health measurement properties

Table 3 shows the modified COSMIN risk of bias checklist for the studies included in this systematic review. In the modified version of the checklist “criterion validity” and “responsiveness” were omitted because none of the studies reported these parameters. In this modified version, the term “tool development” substituted the term “Patient Reported Outcome Measure (PROM) development” from the original COSMIN checklist, in order to reflect the real target population of the studies which are students, teachers, and principals instead of “patient”.

The articles were assessed by three raters who agreed that none of the seven articles scored “high quality”, three of the articles were assessed as “moderate quality”, and the remaining four articles as “low quality”.

Tool development

There are some requirements for developing an instrument. First of all the theory, framework or model used for developing the instrument should be clearly described, the target population should be specified. Context of use and qualitative data collection method should be clarified. The interviewer should be skilled and the data collected from qualitative study should be recorded and transcribed and coding section should be performed by at least two researchers independently. In addition a pilot study should be conducted to test the comprehensibility of the items, and the appropriate number of examinees for qualitative studies should be ≥ 7 and for quantitative studies ≥ 50 to score as “very good”.

Four articles of the seven articles included in this review (Weintraub & Erez, 2009; Lee et al., 2014; Andrews & Conte, 2005; Pinto et al., 2016) were development studies; three of the studies' quality were assessed as “adequate” (Weintraub & Erez, 2009; Lee et al., 2014; Pinto et al., 2016) because there was not enough information whether the coding was performed independently, if the data collection continued until saturation, and if the group meetings have been recorded and transcribed. The other article (Andrews & Conte, 2005) was scored as “very good” because it fulfilled all criteria in the “tool development” box of the COSMIN checklist, including a clear description of steps for developing the instrument such as five focus groups indulging in determining indicators, table of specifications, creating items and response format, and conducting a pilot study. They have also mentioned that members of the advisory group have independently studied the indicators as well as reviewing relevant literature.

Table 2 Characteristics of studies investigating psychometric properties of school health instruments.

Author(s)/Publication year	Country	Sample size	Time/No. of items	Response scale	Participant age/-grade	Participant sex	Instrument factors
<i>Öncel & Sümen (2018)</i>	Turkey	1236	15-20 min/37	Likert	Teachers	Female & male	nutrition services, healthy school policies, physical environment, social environment, community links, individual health skills and action competencies, health services
<i>Pinto et al. (2016)</i>	Brazil	53	NR/28	yes/no	School principals	Female & male	pedagogical dimension (drug and sexuality, violence and prejudice, self-care, peace and quality of life), structural dimension (access, structure conservation and equipment, sanitary conditions), relational dimension (community relations and relationship at school)
<i>Ghotra et al. (2016)</i>	Canada	629	NR/36	Likert	Grades 4-6	Female & male	teacher student relationship and social activities, physical environment, negative feelings towards school, positive feelings towards school
<i>Lee et al. (2014)</i>	Korea	728	NR/37	Likert	Principals, vice principal, health teachers, physical education teachers, classroom teachers, nutritionists and school counselors	Female & male	healthy school policies, physical environment, social environment, community links, individual health skills and action competencies, health services
<i>Weintraub & Erez (2009)</i>	Israel	526	NR/36	Likert	Grades 3, 4, 5, 6	Female & male	teacher-student relationship and school activities, physical environment, negative feelings toward school, positive feelings toward school
<i>Andrews & Conte (2005)</i>	Canada	570	NR/87	Likert	Teachers & Administrators	Female & male	physical health, active living, safety, nutrition, interpersonal relationships, mental health, academic health, sexual health, partnerships, substance abuse, and physical environment
<i>Brener, Kann & Smith (2003)</i>	USA	1498	NR/23 questionnaires	Categorical, ordinal, interval, continuous range	Elementary, middle & high school	Female & male	health services, mental health and social services, food service, school policy and environment, and faculty and staff health promotion

Notes.

*NR, not reported

Table 3 Risk of bias for school health assessment tools in primary schools.

Authors	Items of modified COSMIN risk of bias checklist*								Percent	Rating
	1	2	3	4	5	6	7	8		
Oncel & Sumen (2018)	NA	IA	VG	VG	NR	DF	NR	NR	28%	Low quality
Pinto et al. (2016)	AD	VG	AD	VG	NA	VG	NR	NR	42%	Moderate quality
Ghotra et al. (2016)	NA	NA	AD	VG	AD	DF	NR	VG	33%	Moderate quality
Lee et al. (2014)	AD	VG	VG	VG	NA	DF	NR	NR	42%	Moderate quality
Weintraub & Erez (2009)	AD	IA	VG	VG	NA	DF	NR	NR	28%	Low quality
Andrews & Conte (2005)	VG	VG	AD	DF	NA	DF	NR	NR	28%	Low quality
Brener, Kann & Smith (2003)	NA	NA	IA	IA	NA	VG	NR	NR	16%	Low quality

Notes.

NR, not reported; NA, not applicable; VG, very good; AD, adequate; DF, doubtful; IA, inadequate.

*Items of the modified COSMIN risk of bias checklist. Items: 1: tool development; 2: content validity; 3: structural validity; 4: internal consistency; 5: cross-cultural validity/measurement invariance; 6: reliability; 7: measurement errors; 8: hypotheses testing for construct validity.

Content validity

According to the checklist the study should examine items both in an expert group and the target population to determine relevance and comprehensiveness by skilled interviewers. Other criteria are similar to the “tool development” parameter.

Five of the seven studies examined content validity of the instrument at hand. Three of them have met these criteria and received “very good” (*Lee et al., 2014; Andrews & Conte, 2005; Pinto et al., 2016*). One of the studies used the Delphi method and pilot study for evaluating content validity (*Pinto et al., 2016*); another has used focus group and a pilot study as the method of validation (*Andrews & Conte, 2005*); one study has examined only the comprehensibility of the items from students’ points of view and did not evaluate content validity by asking professionals, so it was scored as “inadequate” (*Weintraub & Erez, 2009*), and the last study used only experts and did not perform a pilot study and therefore was scored “inadequate” (*Öncel & Sümen, 2018*).

Structural validity

Measuring structural validity is only applicable for instruments that are based on a reflective model, something that all instruments included in this systematic review were. The studies that have conducted confirmatory factor analysis and had a sample size that was seven times the number of items and ≥ 100 get scored as “very good”, studies with exploratory factor analysis and a sample size at least five times the number of items and ≥ 100 or the sample size at least six times the number of items but < 100 get scored as “adequate”.

All of the seven articles have investigated the structural validity of the instrument at hand. Three articles have performed confirmatory factor analysis with a sufficient sample size (seven times the number of items) and were scored as “very good” (*Öncel & Sümen, 2018; Weintraub & Erez, 2009; Lee et al., 2014*), three studies have performed exploratory factor analysis (*Ghotra et al., 2016; Pinto et al., 2016; Andrews & Conte, 2005*) and were assessed as “adequate”; the last one was assessed as “inadequate” because it has not performed confirmatory or exploratory factor analysis (*Brener, Kann & Smith, 2003*).

Internal consistency

In COSMIN Risk of Bias checklist any study that measures internal consistency for each of the subscales independently is considered to have a “very good” quality and any study presenting only the total internal consistency or not calculating the internal consistency at all, receives the score “inadequate”. Five studies have calculated Cronbach’s alpha for each subscale or unidimensional scale and were therefore evaluated as “very good” (*Öncel & Sümen, 2018; Weintraub & Erez, 2009; Lee et al., 2014; Pinto et al., 2016; Ghotra et al., 2016*), one of the studies was scored as “doubtful” because internal consistency was unclear and was not fully reported (*Andrews & Conte, 2005*), the last study was scored as “inadequate” because the internal consistency has not been calculated (*Brener, Kann & Smith, 2003*).

Cross-cultural validity/measurement invariance

Cross-cultural validity or measurement invariance is needed for a translated or culturally adapted instrument to compare performance of the items to the original version of the instrument. The samples should be similar for relevant characteristics except for the group variance; the sample size should be 200 subjects per group to be scored as “very good” and 150 subjects per group to be scored as “adequate”. Cross-cultural validity assesses if the instrument is Measurement Invariance and not Differential Item Functioning occurs; this indicates that examinees from different groups with the same latent trait level act invariant to an item. Appropriate methods for Classical Test Theory (CCT) studies are regression analyses or confirmatory factor analysis, while for Item-Response Theory (IRT), the appropriate method is Differential Item Functioning (*Teresi et al., 2009*).

Only two of the studies could be investigated for cross-cultural validity. One study was scored as “adequate”, because it did not report comparison between different groups clearly (*Ghotra et al., 2016*), and the other study has not reported cross-cultural validity at all (*Öncel & Sümen, 2018*).

Reliability

As the reliability value should be stable on the construct to be measured between repeated test administrations, the time interval should be appropriate, the test condition should be similar to original instrument (e.g., administration, environment, and instructions).

Moreover, the intraclass correlation coefficient should be calculated for continuous scores. For dichotomous/nominal/ordinal scores Cohen’s kappa should be calculated, for ordinal scored weighted kappa should also be calculated and it is recommended to describe weighting scheme for ordinal scores (*Prinsen et al., 2018*).

Two of the studies received score “very good” because one of them has performed split-half reliability (*Pinto et al., 2016*) and the other has performed a test-retest reliability for continuous data and Kappa has also been calculated for dichotomous scores (*Brener, Kann & Smith, 2003*). The rest five studies were scored as “doubtful” because they have not presented a clear description about stability across respondents, time interval, similar conditions or calculating the intraclass correlation coefficient.

Measurement error

Conditions for measurement error is similar to reliability, the same examinees, appropriate time interval, and similar conditions for measurements. Standard Error of Measurement, Small Detectable Change or Limits of Agreements should be calculated. Percentage (positive and negative) agreement should be calculated for dichotomous/nominal/ordinal scores. Calculation of Standard Error Measurement through Cronbach's alpha is not appropriate, because it does not take the variance between time points into account and it should be calculated by a test-retest design (*De Vet et al., 2006*). None of the included studies calculated Standard Error of Measurement, Limits of Agreement or Smallest Detectable Change in the instruments being used in each study.

Hypotheses testing for construct validity

Hypothesis testing for construct validity requires comparisons of the instrument under development with other outcome measurement instruments (convergent validity) through Pearson correlation and describing distribution of scores or mean scores. Performing confirmatory factor analysis or structural equation modeling over scales are also proper methods to measure differences between similar constructs. Discrepancy in scores among 'known groups' (discriminative or known group validity study) should be also measured. In discriminative or known group validity study significant characteristics of the groups under study such as age, gender, language etc. should be reported.

Only one of the studies (*Ghotra et al., 2016*) has calculated construct validity and was assessed as "very good" because it reported both convergent validity and discriminative validity for the instrument under investigation. The rest of the studies have not reported construct validity by comparing the school health instruments with another instrument or throughout known group validity.

DISCUSSION

This systematic review assessed the measurement properties of school health assessment tools used in primary schools through the modified COSMIN methodology checklist. The selected instruments measure school health's different factors. Despite the importance of existing valid and reliable tools to measure school health, there are only a few number of school health assessment tools. Moreover, for instruments like CDC's School Health Index and CDC's School Health Profile, our search strategy was not able to find any study measuring their psychometric properties. Nevertheless, we systematically found seven studies that were recognized as eligible for our criteria. As far as we know, this is the first systematic review that has methodologically studied the quality of school health assessment tools.

The results showed that all of the seven selected studies have performed structural validity, internal consistency, and reliability. Most of the studies had satisfactory results in internal consistency, that is, they have conducted Cronbach's alpha for each subscale or have tested the instruments' unidimensionality.

Concerning structural validity and content validity, more numbers of the studies were scored "very good" among the other measurement properties after the internal consistency,

indicating a relatively strong level of evidence on these measurement properties compared to others. Conversely, none of the studies have calculated measurement error and just one study has considered the hypotheses testing for construct validity. If not specified otherwise, the random or biased errors of the scores are not obvious; which implies that we do not know the difference between the observed scores and the true scores (*Schmidt & Hunter, 1999*). That is, we do not have enough information to understand how reliable is the school health score obtained from the instrument.

Two instruments have been translated to other languages (*Öncel & Sümen, 2018; Ghotra et al., 2016*) and therefore should have calculated cross-cultural validity; but only one of them has assessed this property (*Ghotra et al., 2016*). Assessment of cross-cultural validity is important because the instrument may vary across different samples (e.g., healthy and patients), different gender (male and female), different languages (e.g., English and Persian), different cultures etc. (*Hjemdal et al., 2015*). Since the perception of health differs between cultures, certain school health factors might be less or more valued depending on cultural aspects (*Eves & Cheng, 2007; Grunert, Hieke & Juhl, 2018; Chiu et al., 2016*). *Öncel & Sümen (2018)* have assessed the psychometric properties of the Scale for Health Promoting Schools (*Lee et al., 2014*) in Turkish society without calculating cross-cultural validity, and *Ghotra et al. (2016)* have investigated the psychometric properties of the instrument developed by (*Weintraub & Erez, 2009*) named Quality of Life in School and estimated cross-cultural validity, although not sufficiently.

Reliability measurement requires a double test for the instrument with stable samples and proper time interval and calculating intraclass correlation coefficient or Kappa. Just two studies could be scored as “very good” (*Ghotra et al., 2016; Brener, Kann & Smith, 2003*) that shows the poor quality for this measurement property in the literature. Reliability is essential as it leads to trust in the instrument being applied as well as its results; for instance, if the stakeholders of school health assessment do not know how reliable the results obtained from the instrument are and whether it reproduces similar scores in different samples, then the validity of the instrument decreases as well and the results are not trustworthy (*Melchers & Beck, 2018*).

Eventually, none of the studies scored as “high quality” because they have not measured all the required measurement properties using the correct statistical methods; three studies were scored as “moderate quality” (*Pinto et al., 2016; Ghotra et al., 2016; Lee et al., 2014*); and the other four studies were scored as “low quality” (*Öncel & Sümen, 2018; Weintraub & Erez, 2009; Andrews & Conte, 2005; Brener, Kann & Smith, 2003*).

The Scale for Health-Promoting School’ items (*Lee et al., 2014*) were developed based on the regional guideline provided by the WHO (1996) on development of Health-Promoting Schools in South Korea ($\alpha = 0.97$, RMSEA = 0.07, CFI = 0.98). The internal consistency between factors has ranged from 0.86 to 0.91. The Quality of Life at School Questionnaire (*Weintraub & Erez, 2009*) has been developed by two sources of theoretical literature related to children’s quality of life in general, school quality of life and interview with students, parents, and teachers by a semi-structured questionnaire in Israel. Internal consistency among its factors has ranged from 0.68 to 0.91; correlation between total score and factors has been $0.51 < r < 0.69$. The School Health Evaluation Instrument (*Pinto et al., 2016*) has

been developed based on WHO guideline and interview by school managers of Brazil. The Kaiser-Meyer-Olkin (KMO), split-half reliability and Cronbach's alpha has been measured for each factor separately. For pedagogical dimension KMO has been 0.705, split-half test of 0.856 and Cronbach's alpha above 0.7; for structural dimension KMO has been equal to 0.639, split half reliability of 0.805 and Cronbach's alpha above 0.7; and finally for relational dimension KMO has been 0.681, split half test of 0.646 and Cronbach's alpha below 0.7.

Brener, Kann & Smith (2003) have measured the psychometric properties of School Health Policies and Programs Scale 2000 (SHPPS), this instrument has been developed by the Center for Disease Control and Prevention (CDC) containing 23 questionnaires. In this study Kappa has been calculated as a measure of reliability for dichotomous answers ranged from 23.3% to 88.5%, and Pearson correlation ranged from 0.38 to 0.80. This study has reported reliability for categorical, ordinal and interval items by percent without describing the exact values, but Kappa values have been reported completely. Healthy School Indicator Tool (HSIT) has been developed by *Andrews & Conte (2005)* based on Comprehensive School Health initiative (1995) and partnership of health and education professionals in Calgary, Canada. Cronbach's alpha for the nine factors have been reported above 0.7, and also a split-half reliability for all subscales and total HSIT above 0.7.

Ghotra et al. (2016) have measured psychometric properties of Quality of Life at School Scale in Canada. Cronbach's alpha has ranged from 0.75 to 0.93, correlation between subscales have been above 0.60. This study includes a split-half reliability but the values have not been reported. The last study performed by *Öncel & Sümen (2018)* has measured psychometric properties of the Scale for Health-Promoting Schools. They have reported Cronbach's alpha for total scale equal to 0.95 and for subscales ranged from 0.55 to 0.93, goodness of fit equal to 0.63, RMSEA of 0.12, KMO of 0.9, and a correlation between factors from 0,1 to 0.88. They have also reported a content validity index ranged from 0.91 to 1.

Implications for school health measurement

Our findings suggest that measuring all aspects or factors of school health is not widespread. Most of the studies have measured just one, two or some factors related to school health. Measuring health at school is necessary, since it affects the policies and decisions of stakeholders of schools. Further investigations are required into measuring comprehensive and precise psychometric testing on existing tools, as well as advanced development of new tools considering appropriate statistical methods for measuring cross-cultural validity, reliability, measurement error, and construct validity. Nevertheless, among the existing tools, the Scale for Health-Promoting Schools (*Lee et al., 2014*) and the School Health Evaluation Instrument (*Pinto et al., 2016*) are the best available options for primary schools in terms of methodological evaluation. In fact, *Lee et al. (2014)* instrument has a good validity and *Pinto et al. (2016)* instrument has a good reliability. Therefore, although they require more evaluations regarding afore mentioned measurement properties, we recommend researchers, scholars and stakeholders to apply them in primary schools and to continue the psychometric testing and development of these instruments.

Limitations

In sum, we found only a few instruments assessing school health. What is more, there are only a few studies measuring the psychometric properties of these instruments. Importantly, this small number of studies have not considered some of the measurement properties mentioned in the COSMIN checklist. Thus, there is not enough evidence on the quality of these instruments. This systematic review suggests that researchers need to conduct more studies on school health assessment tools in order to acquire better conclusions on these instrument's measurement properties.

CONCLUSIONS

The studies included in this systematic review show that their instruments had a relatively strong internal consistency, a moderate quality for both structural validity and content validity, and a poor quality for tool development, reliability, cross-cultural validity, measurement error, and hypotheses testing for construct validity. Due to the inconsistency of school health assessment tools, we recommend that future studies on measurement properties of school health assessment tools adhere to the measurement properties of the COSMIN Risk of Bias checklist as it can be used as a helpful detailed manual or guideline to report the psychometric properties of school health instruments.

“Let no one ignorant of geometry enter”
—Plato

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by a grant from The Swedish Research Council (Dnr. 2015-01229). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The Swedish Research Council: Dnr. 2015-01229.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Maryam Kazemitabar conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ali Moghadamzadeh and Mojtaba Habibi performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Rezvan Hakimzadeh analyzed the data, prepared figures and/or tables, and approved the final draft.

- Danilo Garcia conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplementary Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.9459#supplemental-information>.

REFERENCES

- Akihiro K, Hiromi K, Satoko Y, Mika N, Md Moshir R. 2017.** Factors affecting children's cholesterol levels with a view to prevent cardiovascular disease—the unique benefits of health screening at the elementary school. *Health* **9**:172–188 DOI [10.4236/health.2017.91012](https://doi.org/10.4236/health.2017.91012).
- Albert L, Amelia Siu Chee L, Mei Wan K, Chi Ming Amy K, Kwok Keung W. 2019.** Effective health promoting school for better health of children and adolescents: indicators for success. *BMC Public Health* **19**:1088 DOI [10.1186/s12889-019-7425-6](https://doi.org/10.1186/s12889-019-7425-6).
- Allensworth DD, Kolbe LJ. 1987.** The comprehensive school health program: exploring an expanded concept. *Journal of school health* **57(10)**:409–412.
- Anderson I, Lyons JG, Luke JN, Reich HS. 2017.** *Health determinants and educational outcomes for indigenous children*. Palgrave: Macmillan DOI [10.1057/978-1-137-53435-4_12](https://doi.org/10.1057/978-1-137-53435-4_12).
- Andrews JJ, Conte R. 2005.** The development and use of the healthy school indicator tool (HSIT). *Canadian Journal of School Psychology* **20(1–2)**:98–110 DOI [10.1177/0829573506295456](https://doi.org/10.1177/0829573506295456).
- Anita EP. 2019.** The association between under-nutrition, school performance and perceptual motor functioning in first-grade South African learners: the North-West Child Health Integrated with Learning and Development study. *Health SA Gesondheid: Journal of Interdisciplinary Health Sciences* **0**:e1 DOI [10.4102/hsag.v24i0.1046](https://doi.org/10.4102/hsag.v24i0.1046).
- Ballard ED, Van Eck K, Musci RJ, Hart SR, Storr CL, Breslau N, Wilcox HC. 2015.** Latent classes of childhood trauma exposure predict the development of behavioral health outcomes in adolescence and young adulthood. *Psychological Medicine* **45(15)**:3305–3316 DOI [10.1017/S0033291715001300](https://doi.org/10.1017/S0033291715001300).
- Best NC, Oppewal S, Travers D. 2018.** Exploring school nurse interventions and health and education outcomes: an integrative review. *Journal of School Nursing* **34(1)**:14–27 DOI [10.1177/1059840517745359](https://doi.org/10.1177/1059840517745359).
- Birch DA, Auld ME. 2019.** Public health and school health education: aligning forces for change. *Health Promotion Practice* **20(6)**:818–823 DOI [10.1177/1524839919870184](https://doi.org/10.1177/1524839919870184).
- Boroumandfar Z, Kohan S, Mohaghegh N, Mohamadirizi S, Hassanzadeh A. 2015.** The prevalence and types of child abuse among teachers and its relationship with their

- demographic characteristics and general health in primary schools. *International Journal of Pediatrics* 3(4.2):853–859 DOI 10.22038/ijp.2015.4585.
- Bravo-Sanzana M, Salvo S, Mieres-Chacaltana M. 2015.** The importance of emotional health in schools, a factor to consider for the improvement of mental health and learning achievements. *Salud Publica de Mexico* 57(2):111–112.
- Brener ND, Kann L, Smith TK. 2003.** Reliability and validity of the School Health Policies and Programs Study 2000 questionnaires. *Journal of School Health* 73(1):29–37 DOI 10.1111/j.1746-1561.2003.tb06556.x.
- Brener ND, Pejavara A, McManus T. 2011.** Applying the School Health Index to a nationally representative sample of schools: update for 2006. *Journal of School Health* 81(2):81–90 DOI 10.1111/j.1746-1561.2010.00564.x.
- Burt CJ, Beetem N, Iverson C, Hertel V, Peters DA. 1996.** Preliminary development of the school health intensity rating scale©. *Journal of School Health* 66(8):286–290 DOI 10.1111/j.1746-1561.1996.tb03400.x.
- Chalupka S, Anderko L. 2019.** Climate change and schools: implications for children’s health and safety. *Creative Nursing* 25(3):249–257 DOI 10.1891/1078-4535.25.3.249.
- Chen FL, Lee A. 2016.** Health-promoting educational settings in Taiwan: development and evaluation of the Health-Promoting School Accreditation System. *Global Health Promotion* 23(1_suppl):18–25 DOI 10.1177/1757975916638286.
- Chiu MM, Chow BWY, McBride C, Mol ST. 2016.** Students’ sense of belonging at school in 41 countries: cross-cultural variability. *Journal of Cross-Cultural Psychology* 47(2):175–196 DOI 10.1177/0022022115617031.
- Dai CL. 2019.** School health program: impacting physical activity behaviors among disadvantaged students. *Journal of School Health* 89(6):468–475 DOI 10.1111/josh.12758.
- De Vet HC, Terwee CB, Knol DL, Bouter LM. 2006.** When to use agreement versus reliability measures. *Journal of Clinical Epidemiology* 59(10):1033–1039 DOI 10.1016/j.jclinepi.2005.10.015.
- Deborah H. 2019.** Promoting positive mental health in the primary school. In: *Theory into practice*. London: Routledge DOI 10.4324/9780429504051.
- Eves A, Cheng L. 2007.** Cross-cultural evaluation of factors driving intention to purchase new food products—Beijing, China and south-east England. *International Journal of Consumer Studies* 31(4):410–417 DOI 10.1111/j.1470-6431.2007.00587.x.
- Finch E. 2002.** Physical rehabilitation outcome measures. In: *A guide to enhanced clinical decision making*. Ottawa: Canadian Physiotherapy Association.
- Ghotra S, McIsaac JLD, Kirk SF, Kuhle S. 2016.** Validation of the Quality of Life in School instrument in Canadian elementary school students. *PeerJ* 4:e1567 DOI 10.7717/peerj.1567.
- González T, Etow A, De La Vega C. 2019.** Health equity, school discipline reform, and restorative justice. *Journal of Law, Medicine and Ethics* 47((2_suppl)):47–50 DOI 10.1177/1073110519857316.
- Grunert KG, Hieke S, Juhl HJ. 2018.** Consumer wants and use of ingredient and nutrition information for alcoholic drinks: a cross-cultural study in six EU countries. *Food Quality and Preference* 63:107–118 DOI 10.1016/j.foodqual.2017.08.005.

- Hjemdal O, Roazzi A, Maria da Graça BB, Friberg O. 2015.** The cross-cultural validity of the Resilience Scale for Adults: a comparison between Norway and Brazil. *BMC Psychology* **3**(1):18 DOI [10.1186/s40359-015-0076-1](https://doi.org/10.1186/s40359-015-0076-1).
- Holt D. 2020.** *Promoting Positive Mental Health in the Primary School*. London: Routledge DOI [10.4324/9780429504051](https://doi.org/10.4324/9780429504051).
- Jihene MM, Imed H, Rim G, Sonia H, Mylene B, Slim S, Hassen G. 2015.** Baseline characteristics of a school based intervention to prevent non communicable diseases risk factors: project together in Health. *La Tunisie medicale* **93**(01).
- Kalmakis KA, Chandler GE. 2015.** Health consequences of adverse childhood experiences: a systematic review. *Journal of the American Association of Nurse Practitioners* **27**(8):457–465 DOI [10.1002/2327-6924.12215](https://doi.org/10.1002/2327-6924.12215).
- Kelly C, Luna C, Carpenter D, Field J, Tucker E, Behrens TK, Holeva-Eklund WM. 2019.** Increasing physical activity in schools: strategies for school health practitioners. *Health Promotion Practice* **20**(5):697–702 DOI [10.1177/1524839919857983](https://doi.org/10.1177/1524839919857983).
- Kolbe LJ. 2019.** School health as a strategy to improve both public health and education. *Annual review of public health* **40**(1):443–463 DOI [10.1146/annurev-publhealth-040218-043727](https://doi.org/10.1146/annurev-publhealth-040218-043727).
- Kristiansen L, Holmstrom MR, Olofsson N. 2016.** Assessing the construct validity and reliability of school health records using the ‘health dialogue questionnaire’ in the eleventh grade. *AIMS Public Health* **3**(3):470–486 DOI [10.3934/publichealth.2016.3.470](https://doi.org/10.3934/publichealth.2016.3.470).
- Krok-Schoen JL, Bernardo BM, Weier RC, Peng J, Katz ML, Reiter PL, Paskett ED. 2018.** Belief about mandatory school vaccinations and vaccination refusal among ohio appalachian parents: do demographic and religious factors, general health, and political affiliation play a role? *The Journal of Rural Health* **34**(3):283–292 DOI [10.1111/jrh.12285](https://doi.org/10.1111/jrh.12285).
- Langford R, Bonell C, Jones H, Poulidou T, Murphy S, Waters E, Campbell R. 2015.** The World Health Organization’s Health Promoting Schools framework: a Cochrane systematic review and meta-analysis. *BMC Public Health* **15**(1):130 DOI [10.1186/s12889-015-1360-y](https://doi.org/10.1186/s12889-015-1360-y).
- Lee EY, Shin YJ, Choi BY, Cho HSM. 2014.** Reliability and validity of a scale for health-promoting schools. *Health Promotion International* **29**(4):759–767 DOI [10.1093/heapro/dat021](https://doi.org/10.1093/heapro/dat021).
- Lewallen TC, Hunt H, Potts-Datema W, Zaza S, Giles W. 2015.** The whole school, whole community, whole child model: a new approach for improving educational attainment and healthy development for students. *Journal of School Health* **85**(11):729–739 DOI [10.1111/josh.12310](https://doi.org/10.1111/josh.12310).
- Lindgaard Nordin L, Jourdan D, Simovska V. 2019.** (Re)framing school as a setting for promoting health and well-being: a double translation process. *Critical Public Health* **29**(3):325–336 DOI [10.1080/09581596.2018.1449944](https://doi.org/10.1080/09581596.2018.1449944).
- Lohrmann DK. 2008.** A complementary ecological model of the coordinated school health program. *Public health reports* **123**(6):695–703 DOI [10.1177/003335490812300605](https://doi.org/10.1177/003335490812300605).

- Mannathoko MC. 2019.** Health and safety concerns in art and design primary school classrooms in Botswana: reflections on the art and design module. *Art Education* 72(6):20–24 DOI 10.1080/00043125.2019.1648141.
- Mauch M, Rist HJ, Kaelin X. 2017.** Reliability and Validity of Two Measurement Systems in the Quantification of Jump Performance. *Swiss Journal of Sports Medicine & Sports Traumatology* 65(1):57–63.
- Mehlhausen-Hassoen D, Winstok Z. 2019.** The association between family violence in childhood and mental health in adulthood as mediated by the experience of childhood. *Child Indicators Research* 12(5):1697–1716 DOI 10.1007/s12187-018-9605-9.
- Melchers RE, Beck AT. 2018.** Structural reliability analysis and prediction. 3rd edition. Hoboken: John Wiley & Sons.
- Mishra P, Fareed N, Battur H, Khanagar S. 2018.** Effectiveness of an integrated hand-wash and mouthrinse programme on general health of young adolescent school children: a randomized controlled trial. *Journal of Indian Association of Public Health Dentistry* 16(2):116–121 DOI 10.4103/jiaphd.jiaphd_181_17.
- Moher D, Liberati A, Tetzlaff J, Altman DG. 2009.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339:b2535 DOI 10.1186/2046-4053-4-1.
- Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, Terwee CB. 2018.** COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research* 27(5):1171–1179 DOI 10.1007/s11136-017-1765-4.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. 2010.** The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology* 63(7):737–745 DOI 10.1016/j.jclinepi.2010.02.006.
- National Association of School Nurses. 2016.** Framework for 21st century school nursing practice: National Association of School Nurses. *NASN School Nurse* 31(1):45–53 DOI 10.1177/1942602X15618644.
- National Association of School Nurses. 2018.** Shaping Your Practice and Changing Mindsets: framework for 21st Century School Nursing Practice™. *NASN School Nurse* 33(4):236–238 DOI 10.1177/1942602X18778292.
- Neumann L, Combe L, Lambert P, Bartholomew K, Morgan S, Bobo N. 2017.** Whole school, whole community, whole child: implications for 21st century school nurses. Position statement. National Association of School Nurses.
- Nima AA, Cloninger KM, Luchesse F, Sikström S, Garcia D. 2020.** Validation of a general subjective well-being factor using classic test theory. *PeerJ* In Press.
- Öncel S, Sümen A. 2018.** Adaptation of the scale for health promoting schools to Turkish society: a validity and reliability study. *Türkiye Klinikleri. Tıp Bilimleri Dergisi* 38(1):46–56 DOI 10.5336/medsci.2017-581321.
- Park BZ, Cantrell L, Hunt H, Farris RP, Schumacher P, Bauer UE. 2017.** State public health actions to prevent and control diabetes, heart disease, obesity and associated

- risk factors, and promote school health. *Preventing Chronic Disease* **14**(12):E127 DOI 10.5888/pcd14.160437.
- Pedersen E. 2019.** Cook county health stresses the importance of back-to-school physicals and vaccines for children. Paddock Publications.
- Peng Y, Finnie RK, Hahn RA, Truman BI, Johnson RL, Fielding JE, Muntaner C, Fullilove MT, Zhang X, Community Preventive Services Task Force. 2019.** Expanded in-school instructional time and the advancement of health equity: a community guide systematic review. *Journal of Public Health Management and Practice* **25**(6):584–589 DOI 10.1097/PHH.0000000000000834.
- Pinto RO, Pattussi MP, Fontoura LDP, Poletto S, Grapiglia VL, Balbinot AD, Teixeira VA, Horta RL. 2016.** Validation of an instrument to evaluate health promotion at schools. *Revista de Saude Publica* **50**:2 DOI 10.1590/S01518-8787.2016050005855.
- Péneau S, Giudici KV, Gusto G, Goxe D, Lantieri O, Herberg S, Rolland-Cachera MF. 2017.** Growth trajectories of body mass index during childhood: associated factors and health outcome at adulthood. *The Journal of Pediatrics* **186**:64–71 DOI 10.1016/j.jpeds.2017.02.010.
- Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, Terwee CB. 2018.** COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research* **27**(5):1147–1157 DOI 10.1007/s11136-018-1798-3.
- Rahman MA, Todd C, John A, Tan J, Kerr M, Potter R, Kennedy J, Rice F, Brophy S. 2018.** School achievement as a predictor of depression and self-harm in adolescence: linked education and health record study. *The British Journal of Psychiatry* **212**(4):215–221 DOI 10.1192/bjp.2017.69.
- Redfern A, Jolley S, Bracegirdle K, Browning S, Plant D. 2019.** Innovations in Practice: CUES-Ed: an in-service evaluation of a new universal cognitive behavioural early mental health intervention programme for primary school children. *Child and Adolescent Mental Health* **24**(2):187–191 DOI 10.1111/camh.12309.
- Roach KE. 2006.** Measurement of health outcomes: reliability, validity and responsiveness. *Journal of Prosthetics and Orthotics* **18**(6):P8–P12.
- Sagatun Å, Kvarme LG, Misvær N, Myhre M, Valla L, Holen S. 2019.** Evaluating a web-based health-promoting dialogue tool in school health services: feasibility and user experiences. *The Journal of School Nursing* Epub ahead of print Oct 9 2019 DOI 10.1177/1059840519879489.
- Schmidt FL, Hunter JE. 1999.** Theory testing and measurement error. *Intelligence* **27**(3):183–198 DOI 10.1016/S0160-2896(99)00024-0.
- Scott ME, Karberg E. 2016.** Measuring children’s care arrangements and their educational and health outcomes internationally. *Global Social Welfare* **3**:75–89 DOI 10.1007/s40609-016-0059-z.
- Shrestha RM, Ayer R, Dhital R, Jimba M, Ghimire M, Shakya P. 2019.** School health and nutrition program implementation, impact, and challenges in schools of Nepal: Stakeholders’ perceptions. *Tropical Medicine and Health* **47**(1): Article 32 DOI 10.1186/s41182-019-0159-4.

- Souza ACD, Alexandre NMC, Guirardello EDB. 2017.** Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde* 26:649–659 DOI [10.5123/S1679-49742017000300022](https://doi.org/10.5123/S1679-49742017000300022).
- Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, Lai JS, Choi SW, Hays RD, Reeve BB, Reise SP, Pilkonis PA, Cella D. 2009.** Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): an item response theory approach. *Psychology Science* 51(2):148–180.
- Terwee CB, Bot SD, De Boer MR, Can der Windt DA, Knol DL, Dekker J, Se Vet HC. 2007.** Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60(1):34–42 DOI [10.1016/j.jclinepi.2006.03.012](https://doi.org/10.1016/j.jclinepi.2006.03.012).
- Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Mokkink LB. 2011.** COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research* 27(5):1159–1170.
- Voon H, Ariff TM. 2019.** Safety and health in schools. *International Journal of Recent Technology and Engineering* 2(2 Special Issue 3):251–258 DOI [10.35940/ijrte.B1043.0782S319](https://doi.org/10.35940/ijrte.B1043.0782S319).
- Wang Q, Shen J. 2016.** Childhood health status and adulthood cardiovascular disease morbidity in rural china: are they related? *International Journal of Environmental Research and Public Health* 13(6):565 DOI [10.3390/ijerph13060565](https://doi.org/10.3390/ijerph13060565).
- Weiler RM, Pigg Jr RM. 2004.** The school health portfolio system: a new tool for planning and evaluating coordinated school health programs. *Journal of School Health* 74(9):359–364 DOI [10.1111/j.1746-1561.2004.tb06629.x](https://doi.org/10.1111/j.1746-1561.2004.tb06629.x).
- Weintraub N, Erez ABH. 2009.** Quality of life in school (QoLS) questionnaire: development and validity. *The American Journal of Occupational Therapy* 63(6):724–731 DOI [10.5014/ajot.63.6.724](https://doi.org/10.5014/ajot.63.6.724).
- World Health Organization. 1948.** Preamble to the constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948. Available at http://www.who.int/governance/eb/who_constitution_en.pdf.
- World Health Organization. 1996.** *Regional guidelines: development of health-promoting schools—a framework for action*. Manila: WHO Regional Office for the Western Pacific.
- World Health Organization. 2017.** *Health promoting schools: experiences from the Western Pacific Region*. Manila: WHO Regional Office for the Western Pacific. Available at <https://iris.wpro.who.int/handle/10665.1/13551>.
- Yun YH, Kim Y, Sim JA, Choi SH, Lim C, Kang JH. 2018.** Development and validity testing of the school health score card. *Journal of School Health* 88(8):569–575 DOI [10.1016/j.jclinepi.2010.02.006](https://doi.org/10.1016/j.jclinepi.2010.02.006).