REVIEW

Application of Rasch Analysis for Development and Psychometric Properties of Adolescents' Quality of Life Instruments: A Systematic Review

Sahar Dabaghi [iD][1]
Fatemeh Esmaielzadeh [iD][1]
Camelia Rohani [iD][2,3]

[1]Student Research Committee, School of Nursing & Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [2]Community Health Nursing Department, School of Nursing & Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [3]Ersta Sköndal Bräcke University College, Campus Ersta, Stockholm SE-100 61, Sweden

Background: Due to the importance of assessing quality of life (QoL) in healthy and ill adolescents, the evaluation of psychometric properties of these questionnaires is important.
Objective: To investigate the application of Rasch analysis in psychometric assessment studies on adolescents' QoL instruments, and to evaluate the quality of reporting Rasch parameters in these studies.
Methods: This systematic review was conducted by searching for papers in electronic databases PubMed, Web of Science, EMBASE, Cochrane Library and Scopus until December 2018.
Results: After screening 122 papers, 31 remained in the study. Around 68% of the studies used the Rasch analysis for instrument testing and 32% for the development of new instruments. In 77.4% of studies, both classical and Rasch methods were used parallel to data analysis. In 32.2% of studies, healthy adolescents were the main target group. The most commonly used instrument in Rasch studies was, KIDSCREEN, administered in different countries. Six Rasch parameters were reported with a higher percentage in the studies. Major reported parameters of Rasch analysis were application of the software program (96.7%), test of item fit to the Rasch model (93.5%), unidimensionality (80.6%), type of the identified mathematical Rasch model (74.1%), threshold (58%) and differential item functioning (54.8%). Based on the psychometric evaluation of the QoL instruments, 71% of studies showed acceptable results.
Conclusion: The application of the Rasch model for psychometric assessment of adolescents' QoL questionnaires has increased in recent decades. But, there is still no strong and commonly used critical appraisal tool or guideline for the evaluation of these papers.
Keywords: adolescence, instrument, psychometric, quality of life, Rasch analysis, systematic review

Introduction

One of the concerns of healthcare providers is adolescents and children's quality of life (QoL).[1] Adolescence, is a transitional phase between the childhood and adulthood period. The World Health Organization (WHO) defines it as a life period between ages 10 to 19.[2] This range is overlapping with the definition of the WHO for young people between 10 to 24.[2] Substance abuse, injuries, obesity, chronic diseases, mental health problems, sexual and reproductive health disorders are important issues during the adolescence period.[3,4] To assess physical, social and mental health as well as evaluating population-based intervention programs,

Correspondence: Camelia Rohani
Ersta Sköndal Bräcke University College,
Campus Ersta, Stigbergsgatan 30, Box
11189, Stockholm SE-100 61, Sweden
Tel +46-766289637
Email camelia.rohani@esh.se

outcome variables like QoL and health-related quality of life (HRQoL) are measured among adolescents and children.[5] Thus, the assessment of QoL among adolescents and children is important to identify individuals with poor QoL condition and to plan intervention strategies.[6] One important part of an individual's general well-being is defined as QoL.[7] As described by the WHO, QoL is an individuals' understanding of their situations in life within the context of the culture and value systems in which they live, in relation to their goals, expectations, standards, and attitudes.[8] However, HRQoL contains at least physical, psychological and social functioning aspects in the context of a disease.[9] Despite various definitions for QoL and HRQoL, the researchers agree on three important features, consisting of multidimensionality, subjectivity, and dynamicity of these concepts.[10]

Since, the 1970s, after efforts to improve QoL, the number of instruments has increased exponentially, some of which are made for adolescents.[11] In clinical and non-clinical settings, the use of QoL measurement tools can provide valuable information for the best possible individual care and evaluation of interventions.[12] Considering the importance of measuring the QoL of adolescents, it is valuable to use valid and reliable questionnaires. The necessity and importance of providing sufficient information about validity and reliability of the questionnaires have been mentioned in previous studies.[13,14] Thus, evaluation of the psychometric properties of these questionnaires is essential.[15] To this end, various tools have been designed and are available as generic and specific-disease tools to measure QoL.[16]

Although, evaluation of psychometric properties of instruments is mostly done by Classical Test Theory (CTT) methods, researchers' attention has increased towards application of more complicated models. Also, numerous researchers have recommended, that new theories should replace the classical theories in various disciplines. The Rasch model is one of the most practical and common approaches used for this purpose.[17] The Rasch analysis, based on the Item Response Theory (IRT), is used for evaluation of the psychometric properties of measurement tools, and is more complex than the classical methods. Relevant findings show that this complexity is superior to the classical methods because more accurate results are obtained. There are some problems regarding the use of classical methods in which raw scores, linear combinations of these scores and responses that have ordinal scale, essentially are considered as interval-scale data. Modern analysis offers solutions to these problems. Rasch analysis is a statistical technique that can be used to convert ordinal data into interval data in questions with two or more options to answer.[18]

QoL Questionnaires with several subscales can be evaluated by Rasch models.[19] In these models, it is assumed that, all the questions and materials of a test measuring a construct constitute a sort of ordered relationship. The second assumption in these models is that a correct answer to one question does not influence the other questions.[17]

Evidence shows that the number of QoL instruments is increasing, some of them specific for adolescents. Due to the importance of assessing QoL in healthy and ill adolescents, the evaluation of psychometric properties of these instruments is important.[17] Thus, a systematic review was conducted on studies that investigated psychometric properties of adolescents' QoL Instruments by Rasch analysis. The following research questions were also put forth in this review:

How is the application of Rasch analysis in psychometric assessment studies on adolescents' QoL Instruments?

How is the quality of reporting Rasch parameters in selected studies?

Methods

The present study was carried out based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009. PRISMA Statement is a critical appraisal tool and helps authors improve the reporting of systematic review and meta-analysis studies' results.[20] Selected studies included all papers reporting psychometric properties of QoL instruments in adolescents by the Rasch method. A systematic search was done by two authors independently using five electronic databases; PubMed, Web of Science, EMBASE, Cochrane Library and Scopus, with no time limit until December 2018 (Table 1). A number of key terms with the assistance of Boolean expressions (AND, OR, NOT) were assessed in the title and abstract of the identified papers. Key terms included "Quality of Life" OR "Life Quality" OR "Health-Related" OR "Health-Related Quality Of Life" OR "HRQOL" AND "Rasch analysis" OR "Rasch analyses" OR "Rasch measurement" OR "Rasch measurement model" OR "Rasch model" OR "Rasch modeling" OR "Rasch rating scale analysis" OR "Rasch rating scale model" AND "Adolescent" OR "Teenager" OR "Teen"

Table 1 Search Syntax in Different Databases

Database	Search Syntax	Number of Outputs
PubMed	("Quality of life"[tiab] OR "Life Quality"[tiab] OR "Health-Related Quality Of Life"[tiab] OR "Health Related Quality Of Life"[tiab] OR HRQOL[tiab]) AND ("Rasch analysis"[tiab] OR "Rasch analyses"[tiab] OR "Rasch measurement"[tiab] OR "Rasch measurement model"[tiab] OR "Rasch model"[tiab] OR "Rasch modeling"[tiab] OR "Rasch modelling"[tiab] OR "Rasch rating scale analysis"[tiab] OR "Rasch rating scale model"[tiab] AND Adolescent[tiab] OR Teenager[tiab] OR Teen[tiab] OR Stripling[tiab])	33
Embase	("Quality of life":ab,ti OR "Life Quality":ab,ti OR "Health-Related Quality Of Life":ab,ti OR "Health Related Quality Of Life":ab,ti OR HRQOL:ab,ti) AND ("Rasch analysis":ab,ti OR "Rasch analyses":ab,ti OR "Rasch measurement":ab,ti OR "Rasch measurement model":ab,ti OR "Rasch model":ab,ti OR "Rasch modeling":ab,ti OR "Rasch modelling":ab,ti OR "Rasch rating scale analysis":ab,ti OR "Rasch rating scale model":ab,ti) AND (Adolescent:ab,ti OR Teenager:ab,ti OR Teen:ab,ti OR Stripling:ab,ti)	14
Web of Science	(TS=(Quality of life) OR TS=(Life Quality) OR TS=(Health-Related Quality Of Life) OR TS=(Health Related Quality Of Life) OR TS=(HRQOL)) AND (TS=(Rasch analysis) OR TS=(Rasch analyses) OR TS=(Rasch measurement) OR TS=(Rasch measurement model) OR TS=(Rasch model) OR TS=(Rasch modeling) OR TS=(Rasch modelling) OR TS=(Rasch rating scale analysis) OR TS=(Rasch rating scale model)) AND (TS=(Adolescent) OR TS=(Teenager) OR TS=(Teen) OR TS=(Stripling))	50
Cochrane Library	(("Quality of life"):ti,ab,kw OR ("Life Quality"):ti,ab,kw OR ("Health-Related Quality Of Life"):ti,ab,kw OR ("Health Related Quality Of Life"):ti,ab,kw OR ("HRQOL"):ti,ab,kw) AND (("Rasch analysis"):ti,ab,kw OR ("Rasch analyses"):ti,ab,kw OR ("Rasch measurement"):ti,ab,kw OR ("Rasch measurement model"):ti,ab,kw OR ("Rasch model"):ti,ab,kw OR ("Rasch modeling"):ti,ab,kw OR ("Rasch modelling"):ti,ab,kw OR ("Rasch rating scale analysis"):ti,ab,kw OR ("Rasch rating scale model"):ti,ab,kw) AND (("Adolescent"):ti,ab,kw OR ("Teenager"):ti,ab,kw OR ("Teen"):ti,ab,kw OR ("Stripling"):ti,ab,kw)	3
SCOPUS	TITLE-ABS("Quality of life") OR TITLE-ABS("Life Quality") OR TITLE-ABS("Health-Related Quality Of Life") OR TITLE-ABS("Health Related Quality Of Life") OR TITLE-ABS (HRQOL) AND TITLE-ABS ("Rasch analysis") OR TITLE-ABS ("Rasch analyses") OR TITLE-ABS ("Rasch measurement") OR TITLE-ABS("Rasch measurement model") OR TITLE-ABS("Rasch model") OR TITLE-ABS("Rasch modeling") OR TITLE-ABS("Rasch modelling") OR TITLE-ABS("Rasch rating scale analysis") OR TITLE-ABS("Rasch rating scale model") AND TITLE-ABS (Adolescent) OR TITLE-ABS(Teenager) OR TITLE-ABS(Teen) OR TITLE-ABS(Stripling)	22

OR "Stripling". The inclusion criteria for study selection were all papers published in a scientific journal in the English language, they used Rasch analysis to evaluate psychometric properties of adolescents QoL instruments, and there was access to the full text of the paper. Exclusion criteria included non-English language papers, adult target population, and review papers. After removing duplicates in the first screening phase, two authors independently reviewed titles and abstracts of the studies. Papers containing the full text, were included in the review and were screened separately by two authors. In cases of disagreement between the two authors, the papers were reviewed by the third author. Two tables were designed in the research team to collect the data from different papers. The following data were extracted from all studies: author's name, year of publication, country, aim of the study, study type, measurement instrument, item numbers, rating scales, measurement construct, sample and age, data collection procedure, study results, type of software, type of Rasch parameters, validity (Classical and Rasch methods), reliability (Classical and Rasch methods), and mixed results of both Rasch and Classical methods. Three authors individually evaluated the Rasch parameters using a Quality Identifiers (QI) checklist. All processes of studies' selection and data extraction were discussed within the research team. The QI checklist which was introduced by Tennant and Conaghan with 10 items, was used to evaluate quality of the reporting of Rasch parameters.[21,22] The criteria for this 10-item checklist are: (1) Stating the name of the software for Rasch analysis, (2) Reporting mathematical derivation of the Rasch model, (3) Evaluation of the threshold ordering, (4) Test of item fit to the Rasch model, (5) Test of person fit, (6) Testing for differential item functioning (DIF), (7) Reliability, (8) Response dependency, (9) Unidimensionality, and (10) Transformation table.[21] Here there is a short explanation for each item.

Stating the Name of the Software for Rasch Analysis

There are several software programs to analyze the data using the Rasch method, but the main software programs which are used in the studies are WINSTEPS, RUMM, ConQuest and WINMIRA. Each of these software programs report the findings in a slightly different process. But, the main hypothesis in all of them is testing and matching observed response pattern in the data with theoretical pattern expected by the model.[21–24]

Reporting Mathematical Derivation of the Rasch Model

The first step in Rasch analysis is determining and selecting the mathematical derivation of the Rasch model. If the Likert scale items have two response options the chosen model is a dichotomous model. If the items have three or more levels of response options (Polytomous model), the Rasch model will be slightly different and is commonly referred to as the Andrich Rating Scale Model (RSM) or Masters Partial Credit Model (PCM). The two models differs in the way of behaving with thresholds (probabilistic midpoint between two adjacent categories). In the RSM, any category on the response scale is modeled to be "equal" throughout all items. In other words, the gap between thresholds is considered equal across items. In fact, the threshold is a metric distance between items and should be the same across the items of a scale. Studies should explain the type of the model which was used and the reason for this selection.[21–24]

Evaluation of the Threshold Ordering

This evaluation applies to polytomous items using the PCM and the RSM models. It is important to identify the classification structure of response levels to the items and to ensure that answers to the items are consistent with the metric estimate of the underlying construct. This means that the transition from one response level to the next should increase the hidden property of the measure. When this does not happen, disordered thresholds have occurred. Disorder at threshold means collapse of response options (response categories) over one another, followed by disorder at response levels due to relevance, redundancy, or misallocation of the items in the instrument. In the "Rasch-Andrich thresholds" or "step calibrations", disordering appears when some categories do not match to intervals on the latent variable. Andrich disputes that there

is a requirement at the level of the item for the categories to be ordered. The latent response processes must therefore be dependent so that an outcome of being successful on the second latent process and failing the first latent process cannot occur. Therefore, a threshold cannot be passed unless all prior thresholds have been passed. Andrich imposes this order requirement through what he calls a Guttman structure.[25,26] Studies should identify threshold as well as measures to eliminate the disorder.[21–24]

Test of Item Fit to the Rasch Model

There may be a difference between observed and expected responses in each study, so model testing can give the researcher more information. There are different fit statistics in different software programs. Although they cannot be compared to each other, there are standardized values for all individuals and they can subsequently be comparable. For example, WINSTEPS software outputs include infit, outfit, and ZSTD (standardized fit statistics). In the RUMM 2020 software, outputs are Chi-square statistic and residual statistic (standardized sum of all differences between observed and expected values aggregated across all persons) which are very similar to outfit and ZSTD in WINSTEPS software. Infit is the difference between observed and expected responses for those items with difficulty levels close to the individual's ability. The outfit deals with the differences between all items (regardless of difficulty of the item and based on one's ability). Thus, infit gives more weight to the responses for items close to one's ability. Chi-square test shows the difference between expected and observed values of measured properties between groups, indicating different levels of ability. If its value is less than 0.05, then it is considered as a misfit. Another output of the RUMM 2020 is item-trait interaction Chi-square, representing the variance of measured property. If Chi-square becomes significant, hierarchical ordering of response levels varies across attributes.[21–24]

Test of Person Fit

Answers with unusual patterns can influence fit item levels. These incorrect responses can be triggered by unregistered diseases such as cognitive deficits in the respondent. Thus, deleting such responders can lead to improved internal construct validity of the tool. Such response patterns must be identified by high positive residuals. So, person fit is important and should be reported.[21–24]

Testing for DIF (Differential Item Functioning)

DIF or item bias can influence the fit model. This happens when different groups with different demographic characteristics in a sample (eg, youth, the elderly, women, and men) respond differently to the same item while equal levels of response are measured for a latent variable. The goal is not to disregard age group differences, but to have access to the same levels of a trait regardless of the age variable. There may be a difference in different demographic settings, but an important point is that, DIF should be evaluated at least in different age and sex groups in each study.[21–24]

Reliability

There are different outputs for reliability in different software programs. For example, the output of the WINSTEPS software is reported as "person reliability" and "person separation". "Person reliability" indicates to how well the instrument constantly reproduces a participant's score. A "person reliability" coefficient of 0.70 shows an acceptable level, 0.80 a good level, and 0.90 an excellent level. While a "person separation" indicates how well the instrument differentiates persons into separate ability groups. It means "reproducibility of relative measure location, therefore, the greater the person separation statistic, the more distinct strata are disclosed (not less than 2). In RUMM 2030, this is provided by the "Person Separation Index" (PSI). The PSI is equivalent to Cronbach's alpha, though it uses person estimates in logit instead of raw scores. Similar to Cronbach's alpha, a value close to 1 indicates high internal consistency and a value less than 0.70 indicates model misfit.[21–24]

Response Dependency

Response dependency arises when the responses to some items are correlated, that is, answering one item determines the answer to another. Dependent items can be determined by calculating a residual correlation matrix, which should not actually show a significant correlation.[21–24]

Unidimensionality

According to the Rasch model, it is assumed that, all items collectively represent a unidimensional tool and measure one trait. There are several methods used for the measurement of unidimensionality that vary depending on the software which is used in data analysis, but Rasch

Principal Component Analysis (PCA) for residuals or PCAR is the common indicator for the analysis of residuals. In WINSTEPS software, the first remaining contrast is an important indicator and usually should not be higher than two (eigenvalue of <2). However, if the first contrast has an eigenvalue value >2, it indicates that another dimension may be exist. It specifies that the dimension has at least three items.[27] Another way to measure it using RUMM software is reporting a series of independent *t*-tests guided by pairs of person measures fitted from two subsets of positive and negative loading items on the first component of the residuals. Person estimates derived from the positive set of items are compared against those derived from the negative set. A series of unique *t*-tests are performed to compare the probabilities for each individual. Percentage of these tests outside the range of −1.96 to +1.96 should not exceed 5%. The differences in estimates derived from the two sets of the items are normally distributed.[21–24]

Transformation Table

When the Rasch model is used for psychometric study of an instrument, the options of instrument items can be converted to an interval scale. In this case, a transformation table for this change must be reported in the study. Equal interval scaling is important in clinical trials for accurately reporting the change in scores and reporting responsiveness, such as minimal important differences and effect size.[21–24]

Results

From a total of 122 papers which were found until December 2018, 31 papers were finally included in this study. The process of the study selection is shown in Figure 1 (the PRISMA 2009 diagram). The general characteristics of the selected studies are shown in Table 2.

In response to the first research question "How is the application of Rasch analysis in psychometric assessment studies on adolescents' QoL Instruments?" the following results were obtained.

Characteristics of Studies

67.7% of the studies (n = 21) used the Rasch analysis for instrument testing and 32.3% (n = 10) for the development of new instruments. In 77.4% of studies (n = 24), both Classical and Rasch methods were used parallel to data analysis. Mean age of adolescents in selected studies was equal to 12 years old (range: 6–22 years). The most
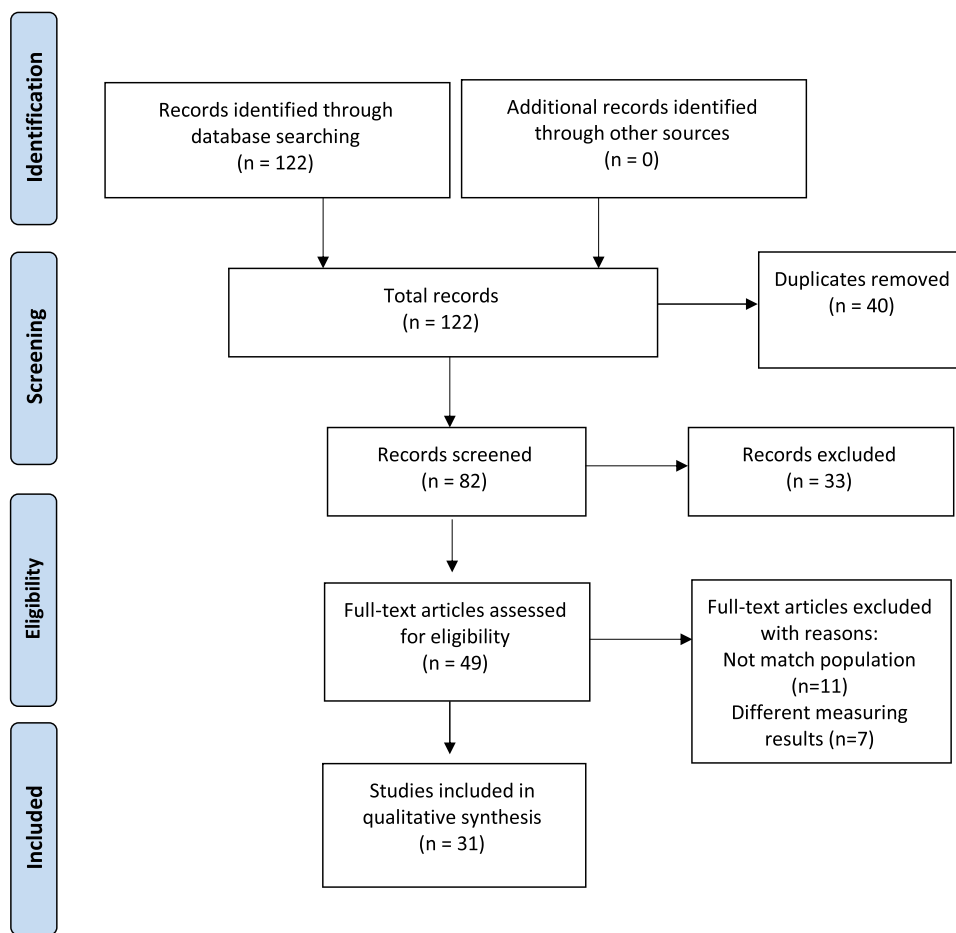
Figure 1 PRISMA diagram for the selection of studies.
Notes: Adapted from Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA GROUP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6 (7): 1000097. doi: 10. 1371/journal. pmed 1000097.[20]

commonly used instrument in Rasch studies was KIDSCREEN, administered in different countries. An important point is that, the majority of QoL of instruments, such as KIDSCREEN (KIDSCREEN Health-Related Quality of Life Questionnaire), DCGM-37 (DISABKIDS Chronic Generic Module), PedsQL (Pediatric Quality of Life Inventory), CFQ (Cystic Fibrosis Questionnaire) cover a wide range of ages, including childhood and adolescence, making it impossible to separate study results on adolescents from children. Various tools have been used, and the main target group for evaluating validity and reliability of the QoL tools was healthy adolescents (n = 78,149) in 32.2% of studies.[24,28–36] Regarding target groups of patients, 9.4% of studies were related to adolescents with Cerebral Palsy (CP), accounting for the highest number of patients.[37–39] Other disorders studied, including kyphosis, skin diseases, diabetes, cancer, glaucoma, bladder dysfunction, visual impairment, Life-Threatening Conditions (LTC), Duchenne muscular dystrophy, cystic fibrosis, mental health, idiopathic scoliosis, spine deformity, mental health, underweight, obesity and overweight have also been the focus of interest. Also, 22.6% of the studies (n = 7) were cross-cultural surveys, with the highest share among European countries (Table 2).

Adolescents' QoL Instruments

Twenty types of QoL instruments were evaluated by the Rasch method. Among various QoL instruments, there were 18 specific tools for adolescents and children; ISYQOL (Italian Spine Youth Quality of Life), WAItE (Weight-specific Adolescent Instrument for Economic evaluation), Skindex-Teen (skin disease scale), VSP-A (Vécu et Santé Percue de l'Adolescent), KIDSCREEN, T-QoL (Teenagers' Quality of Life), DISABKIDS (Disability Kids), PedsQL (Pediatric Quality of Life Inventory), DCGM-37 DISABKIDS (Chronic Generic Module), CHAQ (The Childhood Health Assessment Questionnaire), Sizing Me Up (a self-reported weight-related QoL

Table 2 General Characteristics of the Selected Studies with Application of Rasch Analysis in Adolescents' QoL Instruments (n = 31)

No.	Author (Year); Country	Aim of the Study	Study Type	Instrument; Item Numbers	Rating Scales	Measurement Area	Sample; Age	Data Collection	Results
1	Caronni et al(2017); Italy[45]	Develop a new questionnaire for measuring HRQoL in young people with a spine deformity	Cross-sectional study; Rasch analysis was run on version 1 and 2 of the ISYQOL	ISYQOL; 20	5-point	Spine deformity	Adolescents with idiopathic scoliosis or Scheuermann juvenile kyphosis (n=402); under 18 in two phases: ISYQOL-II; 14.6 ±1.99; and with ISYQOL-II; (15±2.1)	Self-report questionnaire	Two valid and reliable versions of the questionnaires are provided for patients with and without brace.
2	Caronni et al (2014); Italy[51]	Present Rasch analysis of the SRS-22 questionnaire and to develop a short form and Rasch approved of the SRS-7	Psychometric study	SRS-22; 22 SRS-7; 7	SRS-7: four items= 4-point; one item= 3-point; one item=5-point	Idiopathic scoliosis	AIS patients who never were treated before; (n=300); 10–16 years	Self-report questionnaire	The SRS-22 showed a poor clinimetric properties; the SRS-7 showed better targeting of the participants' population.
3	Oluboyede et al (2017); UK[54]	Development and refinement of the WAItE, a new obesity-specific QoL measure for adolescents	Two phases: qualitative phase (one-to-one and focus group interviews); quantitative phase (Psychometric assessments and Rasch analysis)	WAItE; 7	5-point	Overweight and obesity	treatment-seeking above normal weight status and non-treatment-seeking (school sample-adolescents) (n=315); 11–18 years	Self-report questionnaire	The WAItE focuses on aspects of life affected by weight that are important to adolescents. It showed the potential for adding key information to the assessment of weight management interventions aimed in adolescents.

(*Continued*)

Table 2 (Continued).

No.	Author (Year); Country	Aim of the Study	Study Type	Instrument; Item Numbers	Rating Scales	Measurement Area	Sample; Age	Data Collection	Results
4	Smidt et al (2010); US[55]	Development and validation a QoL index for adolescents with skin diseases	Prospective, longitudinal cohort study. Two phases: qualitative phase (expert opinions, and review of the literature); quantitative phase (psychometric assessments and Rasch analysis)	Skindex-Teen; 21	5-point	Skin diseases (Acne, Atopic dermatitis, Nevi or halo nevi, Psoriasis, Verrucae, Alopecia or alopecia areata, Morphea or localized scleroderma, Vitiligo)	Patients with a skin condition (n=205); 12–17 years	Self-report questionnaire	Valid and reliable questionnaire is provided for patients with skin diseases.
5	Sapin et al (2005); France[28]	Evaluation of psychometric properties of the VSP-A questionnaire on two populations, ill and healthy	Psychometric study	VSP-A; 37	5-point	Global HRQoL in ill medical, surgical or psychiatric conditions and healthy adolescents	Two sub-samples: the first with adolescents attending school, and the second with inpatient youths with various conditions (n= 1938); 10–17 years	Self-report questionnaire	Results support the reliability and validity of the VSP-A as a generic report of HRQoL for adolescents.
6	Basra et al (2018); UK[56]	Development and validation of the T-QoL instrument for adolescents with skin diseases	Two phases: qualitative phase (semi-structured interviews with adolescents); quantitative phase (psychometric assessments and Rasch analysis)	T-QoL; 18	5-point	Skin diseases (Acne, Eczema, Psoriasis, Moles, Nonspecific dermatitis, Warts)	Adolescents with skin diseases (n= 426); 12–19 years	Self-report questionnaire	Valid and reliable questionnaire is provided for patients with skin diseases.

#	Author	Objective	Study design	Instrument	Scale	Condition	Sample	Method	Results
7	Ravens-Sieberer et al (2008); 13 European countries[29]	Reliability and validity of the European KIDSCREEN-52 generic HRQoL questionnaire	Cross-sectional study	KIDSCREEN; 52	5-point	HRQoL in healthy adolescents	Children and adolescents (n= 22,827); 8–11 years	Self-report questionnaire	The KIDSCREEN –52 showed a good clinimetric properties.
8	Gothwal et al (2018); India[40]	Psychometric properties of KIDSCREEN-27 and child-parent agreement regarding child's HRQoL in children operated for CG	Prospective cohort with a psychometric evaluation	KIDSCREEN; 27 (child and adolescent version, and parent version)	5-point	Congenital Glaucoma	Children operated for CG (n= 121); 8–18 years	Self-report questionnaire	Items of the KIDSCREEN –27 were fitted with the Rasch model, but a lack of unidimensionality was observed. There was a discordance between CG child's self-report of HRQoL and parent's report.
9	Nielsen et al (2017); Denmark[57]	Translate the DCGM-37 and DSM-10 questionnaires into Danish and use the Rasch model to determine their internal validity and reliability in children and adolescents with T1D	Two phases: translated into the Danish language and psychometric assessments and Rasch analysis	DCGM; 37 and DSM; 10	5-point	T1D	Children and adolescents with T1D (n=413); 8–18 years	Self-report questionnaire	Danish translations of the DISABKIDSDCGM- 37 demonstrated good validity and adequate reliability.
10	Ng et al (2015); China[30]	Reliability and validity of the KIDSCREEN-52 and KIDSCREEN-27	Two phases: translated into the Chinese language and psychometric assessments and Rasch analysis	KIDSCREEN; 27, 52	5-point	HRQoL of Students	Children and adolescents (n= 1379 + 555); 8–14 years	Self-report questionnaire	Chinese version of the KIDSCREEN demonstrated good validity and reliability.

(*Continued*)

Table 2 (Continued).

No.	Author (Year); Country	Aim of the Study	Study Type	Instrument; Item Numbers	Rating Scales	Measurement Area	Sample; Age	Data Collection	Results
11	Jervaeus et al (2013); Sweden[41]	Present Rasch analysis results on the KIDCSREEN-27	Cohort study	KIDCSREEN; 27	5-point	Childhood cancer	Participants diagnosed with cancer (n= 63); 12–22 years	Telephone-interview	The KIDSCREEN-27, showed a good psychometric properties, with the exception of Autonomy and Parent Relations, due to non-satisfactoryunidimensionality, for use among adolescents and young adults who have survived childhood cancer.
12	Jafari et al (2012); Iran[31]	Present Rasch analysis results of the Persian version of the KIDCSREEN-27	Psychometric study	KIDCSREEN; 27	5-point	HRQoL of Students	Children and adolescents (n= 1083); 8–18 years	Self-report questionnaire	Persian version of the KIDSCREEN-27 showed good validity and reliability.
13	Aires et al (2011); Brazil[32]	Psychometric properties of the Brazilian-Portuguese version of the VSP-A	Two phases: translated into the Brazilian language and psychometric assessments and Rasch analysis	VSP-A; 36	5-point	HRQoL in high school students	Adolescents (n= 446); 14–18 years	Self-report questionnaire	Brazilian version of the VSP-A demonstrated good validity and reliability.
14	Kook & Varni (2008); Korea[24]	Psychometric properties of the Korean version of the PedsQL4.0	Two phases: translated into the Korean language and psychometric assessments and Rasch analysis	PedsQL; 23	5-point	QoL of students	School children and adolescents (n= 1425); 8–18 years	Self-report questionnaire	Korean version of the PedsQL showed moderate validity and reliability.

15	Simeoni et al (2007); 7 European countries[46]	Present psychometric properties of the HRQoL DISABKIDS Chronic Generic Measure and to develop a short form by Rasch analysis	Cross-sectional study	DCGM; 37	5-point	HRQoL of children and adolescents	Students with chronic health conditions (asthma, arthritis, epilepsy, cerebral palsy, diabetes, atopic dermatitis, cystic fibrosis (n= 1153); 8–16 years	Self-report questionnaire	The DCGM-37 showed a good clinimetric properties.
16	Robitail et al (2006); 7 European countries[42]	Psychometric properties of the European proxy KIDSCREEN-52	Cross-sectional study	KIDSCREEN; 52	5-point	HRQoL of children and adolescents with Special Health Care Needs	Children and adolescents with special health care needs and their proxies (parents) (n= 2505); 8–18 years	Self-report questionnaire	The KIDSCREEN-52 showed a good psychometric properties.
17	Henry et al (2003); France[52]	Development of the CFQ for assessing QoL in pediatric and adult patients	Two phases: qualitative phase (expert panel and review of the literatures and interview with CF teenagers); quantitative phase (psychometric assessments and Rasch analysis)	CFQ; 25	4-point	Cystic Fibrosis	Patients (101 teenagers and 151 adults) and 141 parents of children with CF; 8–13 years	Self-report questionnaire	Both the CFQ 14 + and the CFQ child parent-proxy French questionnaires showed a good psychometric properties.
18	Chae et al (2018); Korea[37]	Psychometric properties of the Korean version of the CHAQ by applying the Rasch model	Psychometric study	CHAQ; 30	4-point	CP	Children with CP (n= 65); 6–15 years	Proxy-report Questionnaire;	The CHAQ-30, adapted to the Korean population in assessing HRQL in children with CP showed a poor psychometric properties; item development and modification were determined to be necessary.

(Continued)

Table 2 (Continued).

No.	Author (Year); Country	Aim of the Study	Study Type	Instrument; Item Numbers	Rating Scales	Measurement Area	Sample; Age	Data Collection	Results
19	Landfeldt et al (2018); UK & US[47]	Psychometric properties of the English (UK and US) version of the PedsQL NMM	Psychometric study	PedsQL NMM; 25	5-point	Duchenne muscular dystrophy	Patients with Duchenne muscular dystrophy (n= 278); 9–23 years	Self-report questionnaire	The PedsQL NMM-25 showed a poor clinimetric properties.
20	Bray et al (2017); Australia[48]	Two parts: Part one- determine whether a single score as a valid reflection of the HRQoL could be derived from ESM data. Part two- external validity by comparing the measure score derived from the ESM with that of the retrospective measure	Psychometric study	ESM a diary method; 19 PedsQL Generic Core scales; 23	2-, 3-, or 4-point	Duchenne muscular dystrophy	Boys with Duchenne muscular dystrophy (n= 31); 9–18 years	Self-report questionnaire	The ESM −19 as a reported day-to-day experience of QoL showed a good psychometric properties in children.
21	Strong et al (2017); Australia[49]	Psychometric properties of the Taiwan version of the Sizing Me Up for adolescents with normal weight and underweight children in East Asia using Rasch analysis and CTT methods	Two phases: translated into the Taiwanese language and psychometric assessments and Rasch analysis	Sizing Me Up; 22	4-point	Obese, overweight, normal weight, and underweight children	Students in third to sixth grades (n= 497); under 11 years	Self-report questionnaire	The Sizing Me Up showed a good psychometric properties for underweight and normal weight children.

22	Vélez et al (2016); Columbia[33]	Psychometric properties of the Colombian version of the KIDSCREEN-27 children and parent-proxy versions through Rasch analysis	Psychometric study	KIDSCREEN; 27	5-point	HRQoL of ill and healthy children	Children and adolescents: n=321; and parents-proxies: n=1150; 8–18 years	Self-report questionnaire and interview for children; self-report parent-proxy questionnaire	The KIDSCREEN –27 showed a good psychometric properties.
23	Ravens-Sieberer et al (2014); 13 European countries[34]	Development and validation of a new KIDS-CAT based on KIDSCREEN experiences	Psychometric study	KIDS-CAT; KIDSCREEN 52, 27.10	5-point	QoL and well-being in ill and healthy students	Children and adolescents (n= 22,827); 8–18 years	Self-report and proxy versions of the questionnaires	The KIDS-CAT showed a good psychometric properties.
24	Huang et al (2012); US[44]	Develop a HRQoL tool for YASCC based on three legacy instruments (QOL-CS, SF-36, and QLACS)	Psychometric study	HRQoL tool for YASCC;123	7-point	Childhood cancer	Young adult survivors of childhood cancer (N=151); 23–29 years	Interview	The HRQOL tool for YASCC showed a poor psychometric properties.
25	Park et al (2012); Korea[38]	Evaluation of the PODCI in patients with cerebral palsy using Rasch analysis	Cohort study	PODCI; 39	Not reported	Cerebral Palsy	Adolescents with CP, (n= 720); 8–14 years	Self-report and proxy-report versions of the questionnaires	The PODCI showed a poor psychometric properties.

(*Continued*)

Table 2 (Continued).

No.	Author (Year); Country	Aim of the Study	Study Type	Instrument; Item Numbers	Rating Scales	Measurement Area	Sample; Age	Data Collection	Results
26	Huang et al (2011); US[43]	Psychometric properties of a generic HRQoL instrument, the PedsQL 4.0, for children with LTC	Psychometric study	PedsQL; 21	5-point	Life threatening conditions	Parents of children with LTC (n= 257); 2–18 years	Parent proxy-report by telephone interview	The PedsQL showed a poor psychometric properties.
27	Erhart et al (2010); 7 European countries[35]	Compare item reduction analysis based on the CTT with analysis based on the Rasch Partial Credit Model item-fit for children and adolescents' HRQoL items	Cross-sectional study	A dimension from a preliminary KIDSCREEN; 19	5-point	Physical well-being dimension	Children and adolescents; (n= 3019); 8–18 years	Self-report questionnaire	Both types of item reduction analysis should be accompanied by additional analyses. However, the results support the usability of the Rasch method for developing new HRQoL measures for children and adolescents.
28	Khadka, et al (2010); UK[53]	Development and validation of the CVAQC for adolescents with a visual impairment	Two phases: qualitative phase (focus groups and interview); quantitative phase (psychometric assessments and Rasch analysis)	CVAQC; 25	4-point	With and without a visual impairment	Children and young people (N=109); 5–18 years	Self-report questionnaire	The CVAQC showed a good psychometric properties.

29	Erhart, et al (2009); 7 European counties[39]	Rasch Measurement Properties of the KIDSCREEN-52 in Children with CP and DIF between Children with and without CP	Psychometric study	KIDSCREEN; 52	5-point	Cerebral palsy	Children (n=3219) and parents (n=2126) in the general population (KIDSCREEN project); and children with CP (n=501) and their parents (n=823) (SPARCLE project); 8–12 years	Self-report and parent version of the questionnaire	The KIDSCREEN-52 items were understood in the same way by children with and without CP. All items of the tool were fitted with the Rasch model. Some small deviation from unidimensionality was observed in two dimensions of parent-reported version, but not influenced on the measurement.
30	Robitail et al (2007); 13 European countries[36]	Present Rasch analysis of the KIDSCREEN and to develop a short form and Rasch approved of the KIDSCREEN-52	Cross-sectional study	KIDSCREEN; 27	5-point	HRQoL of young people	Children and adolescent (n= 22,827); 8–18 years	Self-report questionnaire	The KIDSCREEN-27 showed a good psychometric properties.
31	Bower et al (2006); China[50]	Development of a validated cross-cultural QoL instrument, specific for children with bladder dysfunction; PinQ	Methodological study in three phases for instrument development	PinQ; 21	2, 4-point	Bladder Dysfunction	Children with Bladder Dysfunction from 10 countries (n= 156); 6–17 years	Self-report questionnaire	The PinQ with two factors showed a good psychometric properties.

Abbreviations: HRQoL, health-related quality of life; ISYQOL, Italian Spine Youth Quality of Life; SRS, Scoliosis Research Society; AIS, Adolescent Idiopathic Scoliosis; WAItE, Weight-Specific Adolescent Instrument for Economic-Evaluation; QoL, quality of life; VSP-A, Vécu et Santé Perçue de l'Adolescent; T-QoL, Teenagers' Quality of Life; CG, congenital glaucoma; DCGM, DISABKIDS Chronic-Generic Module, it is a joint Chronic-Generic Module from the European project of DISABKIDS for developing HRQoL instruments in children and adolescents with chronic conditions; DSM, Diabetes-Specific Module; T1D, type 1 diabetes; PedsQL, Pediatric Quality of Life Inventory; CFQ, Cystic Fibrosis Questionnaire; CF, cystic fibrosis; CHAQ, Childhood Health Assessment Questionnaire; CP, cerebral palsy; PedsQL NMM, Pediatric Quality of Life Inventory 3.0 Neuromuscular Module; ESM, experience sampling method; CAT, computerized adaptive test; YASCC, young adult survivors of childhood cancer; PODCI, Pediatric Outcomes Data Collection Instrument; LTC, life-threatening conditions; CVAQC, Cardiff Visual Ability Questionnaire for Children; PinQ, pediatric QoL tool; CTT, classical test theory; DIF, differential item functioning.

instrument), KIDS-CAT (computerized adaptive test), HRQOL tool for YASCC (Young Adult Survivors of Childhood Cancer), PODCI (Pediatric Outcomes Data Collection Instrument), CVAQC (Cardiff Visual Ability Questionnaire for Children), PinQ (Pediatric QoL), ESM (Experience Sampling) a diary method, SRS-7 (Scoliosis Research Society), and CFQ (Cystic Fibrosis Questionnaire). The most commonly used tool was KIDSCREEN, administered in 11 studies (35.5%) in different countries, examining its psychometric properties.[29–31,33,36,39–41] This questionnaire was designed in the KIDSCREEN project in 10 dimensions with 52 questions and is suitable for measuring QoL of children and adolescents at 8–18 years of age. It has 10 dimensions, including physical well-being, psychological well-being, emotions and moods, social support and peer, parent's relationships, self-perception, autonomy, school environment, social acceptance, and financial resources, measuring QoL across a 5-point Likert scale. The highest score for each item is 4 and the lowest score is 0.[29,30,41] In the present review, studies used the short (10, 19, 27 items) and long versions (52 items) of the questionnaire.[29,31,33–36,39–42] It was used as a self-report questionnaire in most studies, while telephone-interview,[41,43] proxy interview,[33,34,37–39,43] and face-to face interview versions[33,44] have also been reported.

Several types of Likert-scales were used in various studies, but 5-point Likert-scales were mostly used. Based on the psychometric assessment of the instruments, 9 studies (29%) reported unsatisfactory psychometrics results. The rest of studies reported satisfactory psychometrics results (n = 22, 71%) (Table 2). Furthermore, some short forms of the instruments, such as SRS-7, KIDSCREEN-23, and CHAQ-28 showed good psychometric results.

Rasch Parameters

In response to the second research question "How is the quality of reporting Rasch parameters in selected studies?" the following results were obtained:

Tables 2 and 3 represent the detailed results for psychometric assessment of the studies. Rasch parameters using the QI checklist are shown in Table 3. The results show that most studies (n = 30, 96.7%) reported the name of the software program for Rasch analysis; WINSTEPS (n = 23, 74.2%), RUMM (n = 3, 9.7%), and WINMIRA (n = 3, 9.7%), respectively. One study (3.2%) used two software programs; WINMIRA and MULTIRA.[29] Two studies (6.4%) used other software programs; DIGRAM and FACETS.[48,57] Also, 15

studies (48.4%) applied PCM[29,31,33–36,38,39,41,45–50] and 6 studies (19.4%) applied RSM as a mathematical Rasch model.[24,28,40,51–53] One study (3.2%) used both models.[42] In 8 studies (25.8%), no mathematical Rasch model was mentioned.[30,32,37,43,44,54–57] As threshold is important for instruments with polytomous item responses, it was observed in 58.1% of the studies (n = 18). It was not mentioned in several studies (n=13, 41.9%).[28,30,31,38,42–44,52,55] In studies with disordered threshold, poor items were treated by item deletion[45,46,54] or category rescoring.[33,40,41,47,51,53,56]

Among current studies, most of them (n = 29, 93.5%) reported infit and or outfit mean square for item fit, and a few reported another statistic test (Q-index) (n = 2, 6.4%). Out of these, 2 studies (6.4%) did not address item fit. Only 8 studies (25.8%) reported person fit. 17 studies (54.8%) reported DIF, but 14 studies (45.1%) did not mention it. Based on the reliability results, PSI (n = 10, 32.2%), item separation ratio (n = 6, 19.3%), and Cronbach's alpha by RUMM (n = 1, 3.2%) were reported in studies. In 3 studies (9.6%), indicators of reliability were not reported. Except for 6 studies (19.3%), the rest of the studies reported unidimensionality (n = 25, 80.6%). In none of the studies was a transformation table drawn (Table 3). According to the QI checklist, around 71% of the studies (n = 22) reported between 4 to 7 Rasch parameters (Table 4). Rasch parameters, including application of the software program (n = 30, 96.7%), test of item fit to the Rasch model (n = 29, 93.5%), unidimensionality (n = 25, 80.6%), type of the identified mathematical Rasch model (n = 23, 74.1%) and threshold (n = 18, 58%) were reported with a higher percentage, respectively. Reliability was reported in most studies using two different approaches; Rasch (n = 12, 38.7%) and classical methods (n = 16, 51.6%). Reliability was not reported in only three studies (9.6%).

Discussion

The present study was conducted aimed at providing a systematic review of the application of Rasch analysis in psychometric assessment studies on adolescents' QoL instruments. Overall results showed that, most of the selected studies used the Rasch method for instrument testing and fewer for instrument development. Results of this study showed that in the last decade, the number of these studies has increased. The diversity of QoL instruments using Rasch analysis is high, and KIDSCREEN has been mostly used in those studies, as it is applicable to a wide range of healthy and ill adolescents and children

Table 3 Psychometric Properties of the Selected QoL Studies in Adolescents (n = 31)

No.	Author (Year)	Rasch Software	Mathematical Derivation of Rasch Model	Rasch Parameters (Type)	Validity (Rasch and Classical Methods)*	Reliability (Rasch and Classical Methods)*	Mixed Results (Rasch and Classical Methods)
1	Caronni et al (2017)[45]	WINSTEPS	PCM	Infit MnSq, Outfit MnSq, ZSTD, DIF, Unidimensionality, Threshold, person reliability, PSI	PCAR,	Person reliability, PSI	−
2	Caronni et al (2014)[51]	WINSTEPS	RSM	Infit, Outfit, DIF, Unidimensionality, Threshold, Person reliability, PSR	PCAR, Floor-ceiling effect	Person reliability, PSR	-
3	Oluboyede et al (2017)[54]	RUMM	Not reported	Fit residuals (item fit residuals, person fit residuals), Chi-square (difference between observed and expected responses), DIF, Item-threshold probability curve, Unidimensionality, Local dependency, PSI	EFA, CFA, PAF, Floor effect	PSI, Internal Consistency (Cronbach's α), corrected item-total correlation	+
4	Smidt et al (2010)[55]	WINSTEPS	Not reported	Fit statistics (MnSq),Separation index, Item scale reliability	EFA, Construct, Content, Face validity	Separation index (in scale level), Internal consistency (Cronbach's alpha), ICC, Item-scale reliability, Responsiveness	+
5	Sapin et al (2005)[28]	BIGSTEPS/ WINSTEPS	RSM	Unidimensionality	Content validity, Construct validity (inter-item correlations, item- dimension correlations, PCA), Item discriminant validity, Item-convergent validity, External validity, Clinical validity (Known-group), Floor-ceiling effect	Internal consistency (Cronbach's alpha), Sensitivity	+

(*Continued*)

Table 3 (Continued).

No.	Author (Year)	Rasch Software	Mathematical Derivation of Rasch Model	Rasch Parameters (Type)	Validity (Rasch and Classical Methods)*	Reliability (Rasch and Classical Methods)*	Mixed Results (Rasch and Classical Methods)
6	Basra et al (2018)[56]	RUMM	Not reported	DIF, Item responses, Individual item fit, Unidimensionality, Threshold, PSI	Convergent validity, Construct validity, Face and Content validity, EFA, CFA	Cronbach's alpha, ICC, PSI, Internal consistency, Sensitivity (SRM, ES)	+
7	Ravens-Sieberer et al (2008)[29]	WINMIRA, MULTIRA	PCM	DIF, Infit, Item threshold parameter, Person threshold parameter	Convergent and Discriminant validity, Construct validity (between groups by ES), CFA	Internal consistency (Cronbach's alpha), ICC	+
8	Gothwal et al (2018)[40]	WINSTEPS	RSM	Infit, DIF, Unidimensionality, Threshold, Mean item and person location	PCA	Reliability (not reported the type of reliability)	+
9	Nielsen et al (2017)[57]	DIGRAM	Not reported	DIF, Unidimensionality, Threshold, local dependence	Construct validity, GLLRM, overall test-of-fit (using CLR tests)	Cronbach's α, Monte Carlo method	+
10	Ng et al (2015)[30]	WINSTEPS	Not reported	Infit, Dimensional map	CFA, Convergent and Divergent validity, Floor-ceiling effect	Internal consistency (Cronbach's alpha), ICC	+
11	Jervaeus et al (2013)[41]	WINSTEPS	PCM	Item and person goodness of fit statistics (MnSq residuals and standardized z-values), DIF, Unidimensionality, Threshold	PCAR, Internal scale validity, Person response validity	Not reported separately	-
12	Jafari et al (2012)[31]	WINSTEPS	PCM	Infit, Outfit, Item difficulty, CPC	Convergent validity, Discriminant validity	Internal consistency (Cronbach's alpha)	+

#	Author (year)	Software	Model		Validity	Reliability	
13	Aires et al (2011)[32]	WINSTEPS	Not reported	DIF, Infit, Unidimensionality,	CFA (Internal construct validity), External construct validity (by comparison between groups), IIC, IDV, Floor-ceiling effect	Cronbach's alpha, ICC	+
14	Kook & Varni (2008)[24]	WINSTEPS	RSM	Infit, Outfit, Item hierarchy, Threshold, Person reliability, Item reliability, PSI, Item reliability index	Construct validity (Known-groups), Structural validity (CFA), Floor-ceiling effect	Person reliability, Item reliability, PSI	+
15	Simeoni et al (2007)[46]	WINSTEPS	PCM	Infit, Unidimensionality	CFA (Construct validity), Convergent and Discriminant validity, IIC, IDV	Internal consistency (Cronbach's alpha), ICC	+
16	Robitail et al (2006)[42]	WINSTEPS	PCM, RSM	Iinfit, Unidimensionality	CFA, Construct validity, External validity, Convergent and Divergent validity, Floor-ceiling effect	Cronbach's alpha, ICC	+
17	Henry et al (2003)[52]	BIGSTEPS	RSM	Infit, Outfit, Unidimensionality, Item calibration	Convergent and Discriminant validity, External construct validity, Clinical validity	Internal consistency (Cronbach's alpha), ICC, Responsiveness	+
18	Chae et al (2018)[37]	WINSTEPS	Not reported	Infit, Outfit, Item difficulty, Item hierarchy, Threshold, Separation reliability, Separation index	Construct validity, Concurrent validity	Separation reliability (for subject, item), separation index (for subject, item)	-
19	Landfeldt et al (2018)[47]	RUMM	PCM	Person fit (mean fit residual), item fit (mean fit residual), Local item dependency, DIF, Unidimensionality, Item hierarchy, Threshold, PSI, Cronbach's alpha	PCAR	PSI, Cronbach's alpha	-
20	Bray et al (2017)[48]	FACETS	PCM	Infit, Outfit, Unidimensionality, Item hierarchy, Threshold, Person reliability index	Construct validity, correlation with with the summary score from the PedsQ Generic Core scales to examine whether daily experience was representative of HRQL	Person reliability Index (internal consistency)	+

(Continued)

Table 3 (Continued).

No.	Author (Year)	Rasch Software	Mathematical Derivation of Rasch Model	Rasch Parameters (Type)	Validity (Rasch and Classical Methods)*	Reliability (Rasch and Classical Methods)*	Mixed Results (Rasch and Classical Methods)
21	Strong et al (2017)[49]	WINSTEPS	PCM	Infit, Outfit, DIF, Unidimensionality, Item hierarchy, Threshold	CFA, Concurrent validity, Known-groups validity	Internal consistency (Cronbach's alpha), ICC	+
22	Vélez et al (2016)[33]	WINSTEPS	PCM	Infit, Outfit, DIF, Unidimensionality, Threshold, Person and Item separation,	Internal scale validity	Internal consistency (Person and Item separation)	-
23	Ravens-Sieberer et al (2014)[34]	Reported elsewhere	PCM	DIF, Infit, Unidimensionality	CFA, Convergent validity, Known-groups validity, Criterion validity	Cronbach's alpha, ICC	+
24	Huang et al (2012)[44]	WINSTEPS	Not reported	Infit, Outfit, Unidimensionality, Item difficulty and threshold (item-person map), Item and Person reliability, Item and Person separation	CFA, known-groups validity, Floor- ceiling effects	Item separation, PSI, Item reliability, Person reliability, Precision, Inter-item correlation	+
25	Park et al (2012)[38]	WINSTEPS	PCM	Infit (MnSq), Unidimensionality, Average item calibration	Floor-ceiling effect, Item separation (by individual inter-item difference or average item calibration)	No parameter was reported as reliability	-

No.	Author	Software	Model	Item analysis	Validity	Reliability	Rating
26	Huang et al (2011)[43]	WINSTEPS	Not reported	Item hierarchy of difficulty, Item and Person Separation Index	Item-domain convergent and discriminant validity, known-groups validity, construct validity	Internal consistency (Cronbach's alpha), Item Separation Index, Person Separation Index	+
27	Erhart et al. (2010)[35]	WINMIRA	PCM	Item fit (two methods: Infit and Q-Index as a conditional item-fit), Unidimensionality, Cross-cultural DIF, Threshold, Local independence	Structural validity (CFA), known-group validity, Cross-cultural validity, Relative validity for socio-demographic and health-related factors (ES)	Corrected item-total correlation and Cronbach's alpha (maximizing)	+
28	Khadka et al. (2010)[53]	WINSTEPS	RSM	Infit, Outfit, ZSTD, Unidimensionality, Item hierarchy, DIF, Threshold	Person and item estimates, Item reduction and calibration, Content validity, Ceiling effect	PSI, Item Separation Reliability, ICC	+
29	Erhart et al. (2009)[39]	WINMIRA	PCM	Item fit (two methods: Infit and Q-index), DIF, Threshold, Unidimensionality, Local dependence	CFA	No parameter was reported as reliability	+
30	Robitail et al (2007)[36]	WINSTEPS	PCM	Infit, DIF, Unidimensionality	EFA, CFA, IIC, IDV, Floor-ceiling effect	Cronbach's alpha	+
31	Bower et al (2006)[50]	WINSTEPS	PCM	Infit, Unidimentionality	EFA, Internal validity (INFIT), Floor-ceiling effect	Cronbach's alpha	+

Abbreviations: PCM, Masters Partial Credit Model; RSM, Andrich Rating Scale Model; DIF, differential item functioning; PSI, Person Separation Index; PSR, person separation ratio; ZSTD, standardized fit statistics; PCA, principal component analysis; PCAR, principal component analysis of residuals; SEM, structural equation modeling; CFA, confirmatory factor analysis; EFA, explanatory factor analysis; ICC, intra-class correlation analysis; ICC, intra-class correlation coefficient; IIC, item-internal consistency; IDV, item discriminant validity; GLLRM, graphical log linear Rasch model; CLR, conditional likelihood ratio test; SRM, standardised response mean; ES, effect size; MnSq, mean square; CPC, category probability curve.

Table 4 Evaluation of the Quality of Reporting Rasch Parameters by the QI Checklist in Studies (n = 31)

Rasch Parameters		Number of Studies (n)	Percentage (%)
1	Software Program[a]	30	96.7
	WINSTEPS	23	74.2
	WINMIRA	3	9.7
	RUMM 2030	3	9.7
	Others	2	6.4
	Not reported[b]	1	3.2
2	Mathematical derivation of Rasch model	22	70.1
	Partial Credit Model (PCM)	15	48.4
	Rating Scale Model (RSM)	6	19.4
	Both Methods	1	3.2
	Not reported	9	29.0
3	Threshold (for Polytomous Item Response)	18	58.1
4	Test of Item Fit	29	93.5
5	Test of Person Fit	8	25.8
6	DIF[c]	17	54.8
7	Reliability	12	38.7
	PSI[d]	10	32.2
	Item Separation Ratio	6	19.3
	Two methods (PSI, Item Separation Ratio)	4	12.9
	Cronbach's alpha[e]	1	3.2
	Not reported	3	9.6
8	Response Dependency	5	16.1
9	Unidimensionality	25	80.6
10	Transformation Table	0	0.0

Notes: [a]BIGSTEPS was classified in the category of the WINSTEPS, so it is a free DOS precursor to WINSTEPS; One study applied two software programs, WINMIRA and MULTIRA, this study was classified by the name of the first software; [b]One study reported the name of the software elsewhere, therefore it was classified in the category of "not reported"; [c]DIF, Differential Item Functioning; [d]PSI, Person Separation Index; [e]Before RUMM 2030, reliability was reported by Cronbach alpha coefficient.

(10–18 years of age). This suggests that there is a place for further development and exploration of other QoL instruments in the future. Furthermore, most QoL instruments involve a wide range of ages, so they cover two stages of life, childhood and adolescence, such as KIDSCREEN, PedsQL, CFQ. In some of these instruments, it is not possible to distinguish adolescence from the childhood period, which may reduce their efficiency. In the present study, 12 types of instruments involved a wide range of ages, including childhood and adolescence (ISYQOL, KIDSCREEN, DCGM, PedsQL, CFQ, CHAQ, Sizing Me Up, KIDCAT, HRQOL, PODCI, CVAQC, and PinQ).

Evidence shows that there are a limited number of critical appraisal tools and available guidelines for writing papers in the field of Rasch[21,22] and this can probably have an influence on the reporting of Rasch studies. The results of our study indicate that authors have reported a various number of Rasch parameters in their studies. Some parameters, such as application of the software program and test of item fit to the Rasch model were reported in more than 93% of studies. The type of identified mathematical Rasch model was described in more than 74% of studies, and unidimensionality, threshold and DIF were reported in around 80%, 58% and 54% of studies, respectively. But, parameters, such as response dependency, were reported less (around 16%), and the transformation table had not been reported in the selected studies. This is not a surprising result, so the results of an earlier systematic review in the field of rheumatology showed that the application of the transformation table was limited. From 1991 to 2012 in two periods, it was reported as static with a proportion of 11.4% and 12.2%, respectively.[21] As the transformation table is a report about converting ordinal scales to the interval for the estimation of responsiveness and clinical changes in outcome variables, this limit can influence the application of Rasch transformed interval measurement in clinical practice. But, according to the study design and type of the QoL instrument, there is a possibility that it has not been in the objective of those studies. However, this can be an area for further development and evaluation of outcome measures in clinical practice in future Rasch QoL studies.

Furthermore, it should be noted that, the QI checklist was used in the present study as a critical appraisal tool to evaluate the quality of reporting of Rasch parameters in those studies,[21] but the number of critical appraisal tools for Rasch papers is scarce. Following an extensive search by authors, only the QI checklist and the Rasch Box with 4 items at three levels in the COSMIN (Consensus-based Standards for selection of health status Measurement Instruments) were found.[59] Although, there is a guide with five headings, 22 items and a number of useful references for writing Rasch papers in the Journal of Applied Measurement,[58] it has been presented as a general guide and not in the format of a critical appraisal tool. In this study, the QI checklist was applied for assessment of the Rasch parameters in the studies, because of the variety of items and ease of use. QI checklist seems to be a good tool

for criticizing Rasch analysis studies, but it is not widely used, so its validity and reliability has not been entirely confirmed. One of the criticisms of this 10-item checklist is an unclear response level for items, and another one is about the method of scoring. However, it seems that there is a requirement for more rigorous critical appraisal tools and guidelines to evaluate the quality of Rasch papers and Rasch parameters. Evidence shows that, few studies have used the QI checklist, and it is recommended to conduct further research on this tool and to identify the strengths and weaknesses of it. Therefore, application of the Rasch model in adolescents' QoL studies needs a criterion for using necessary parameters, as available evidence indicates that, some papers have only reported several parameters in adolescents' QoL instruments.

Limitations

There were some limitations in this study. The current systematic review was limited to English language papers, thus information in other languages was not reported in this study. Furthermore, Gray literature, consisting of dissertations and congress abstracts, was not included in the study. Also, we assessed only the presence of Rasch parameters according to the QI checklist without evaluating the quality of parameters' analysis.

Conclusions

Application of the Rasch model for psychometric assessment of adolescents' QoL instruments has increased in recent decades. There were 20 types of adolescents' QoL instruments which were evaluated by the Rasch models. The most commonly used instrument was, KIDSCREEN, administered in different countries. Six Rasch parameters were reported with a higher percentage in the studies, including application of the software program (96.7%), test of item fit to the Rasch model (93.5%), unidimensionality (80.6%), type of the identified mathematical Rasch model (74.1%), threshold (58%) and DIF (54.8%). Our results indicate that although Rasch analysis is a robust method in adolescents' QoL studies due to documented instruments' measurement functioning, no access to the strong critical appraisal tools may influence the high power application of it.

Author Contributions

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed on the journal to which the article will be submitted; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no competing interests.

References

1. Huang Y, Zhong X-N, Li Q-Y, et al. Health-related quality of life of the rural-China left-behind children or adolescents and influential factors: a cross-sectional study. *Health Qual Life Outcomes*. 2015;13(1):29. doi:10.1186/s12955-015-0220-x
2. World Health Organization (WHO). Recognizing adolescence. WHO; 2014. Available from: https://apps.who.int/adolescent/second-decade/section2/page1/recognizing-adolescence.html. Accessed October 10, 2020.
3. Adamson P, Bradshaw J, Hoelscher P, Richardson D Child poverty in perspective: an overview of child well-being in rich countries; 2007. Available from: https://www.unicef-irc.org/publications/pdf/rc7_eng.pdf. Accessed September 20, 2020.
4. World Health Organization (WHO). Adolescent friendly health services: an agenda for change. World Health Organization. 2002. Available from: https://apps.who.int/iris/handle/10665/67923. Accessed October 10, 2020.
5. Solans M, Pane S, Estrada MD, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. *Value Health*. 2008;11(4):742–764. doi:10.1111/j.1524-4733.2007.00293.x
6. Temin M, Levine R Start with a girl: a new agenda for global health: center for Global Development; 2009. Availale from: https://www.cgdev.org/sites/default/files/1422899_file_Start_with_a_Girl_FINAL_0.pdf. Accessed September 20, 2020.
7. Dijkhuis RR, Ziermans TB, Van Rijn S, Staal WG, Swaab H. Self-regulation and quality of life in high-functioning young adults with autism. *Autism*. 2017;21(7):896–906. doi:10.1177/1362361316655525
8. World Health Organization (WHO). WHOQOL: measuring quality of life. Available from: https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/. Accessed September 20, 2020.
9. O'Connor RM. *Measuring Quality of Life in Health*. Edinburgh: Churchill Livingstone; 2004.
10. Bakas T, McLennon SM, Carpenter JS, et al. Systematic review of health-related quality of life models. *Health Qual Life Outcomes*. 2012;10(1):134. doi:10.1186/1477-7525-10-134
11. Netuveli G, Blane D. Quality of life in older ages. *Br Med Bull*. 2008;85(1):113–126. doi:10.1093/bmb/ldn003
12. Reich M, Lesur A, Perdrizet-Chevallier C. Depression, quality of life and breast cancer: a review of the literature. *Breast Cancer Res Treat*. 2008;110(1):9–17. doi:10.1007/s10549-007-9706-5
13. Moslehi S, Atefimanesh P, Farid AA. The relationship between problem-focused coping strategies and quality of life in schizophrenic patients. *Med J Islam Repub Iran*. 2015;29:288.
14. Kagawa-Singer M, Padilla GV, Ashing-Giwa K, editors. *Health-Related Quality of Life and Culture. Seminars in Oncology Nursing*. Elsevier; 2010.
15. Grant M, Sun V, editors. *Advances in Quality of Life at the End of Life. Seminars in Oncology Nursing*. Elsevier; 2010.

16. Sun A, Wong-Kim E, Stearman S, Chow EA. Quality of life in Chinese patients with breast cancerm. *Cancer*. 2005;104 (S12):2952–2954. doi:10.1002/cncr.21516

17. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. Lippincott Williams & Wilkins; 2008.

18. Wilson M, Hoskens M. *Multidimensional Item Responses: Multimethod-Multitrait Perspectives. Applied Rasch Measurement: A Book of Exemplars*. Springer; 2005.

19. Wang W-C, Yao G, Tsai Y-J, Wang J-D, Hsieh C-L. Validating, improving reliability, and estimating correlation of the four sub-scales in the WHOQOL-BREF using multidimensional Rasch analysis. *Qual Life Res*. 2006;15(4):607–620. doi:10.1007/s11136-005-4365-7

20. Moher D, Liberati A, Tetzlaff J, Altman DG; The PRISMA Group. Preferred reporting items for systematic reviews and meta- analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed1000097

21. Leung -Y-Y, Png M-E, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease-A special interest group report of OMERACT 11. *J Rheumatol*. 2014;41(1):159–164. doi:10.3899/jrheum.130814

22. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res (Hoboken)*. 2007;57(8):1358–1362.

23. Kelderman H. Multidimensional Rasch models for partial-credit scoring. *Appl Psychol Meas*. 1996;20(2):155–168. doi:10.1177/014662169602000205

24. Kook SH, Varni JW. Validation of the Korean version of the pediatric quality of life inventory™ 4.0 (PedsQL™) generic core scales in school children and adolescents using the Rasch model. *Health Qual Life Outcomes*. 2008;6(1):41. doi:10.1186/1477-7525-6-41

25. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health Qual Life Outcomes*. 2017;15(1):1–8. doi:10.1186/s12955-017-0755-0

26. Andrich D. *The Rasch Model Explained. Applied Rasch Measurement: A Book of Exemplars*. Springer; 2005.

27. Linacre JM. *User's Guide to WINSTEPS*. Chicago, IL: MESA Press; 2011.

28. Sapin C, Simeoni M-C, El Khammar M, Antoniotti S, Auquier P. Reliability and validity of the VSP-A, a health-related quality of life instrument for ill and healthy adolescents. *J Adolesc Health*. 2005;36 (4):327–336. doi:10.1016/j.jadohealth.2004.01.016

29. Ravens-Sieberer U, Gosch A, Rajmil L, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Value Health*. 2008;11(4):645–658. doi:10.1111/j.1524-4733.2007.00291.x

30. Ng JY, Burnett A, Ha AS, Sum KW. Psychometric properties of the Chinese (Cantonese) versions of the KIDSCREEN health-related quality of life questionnaire. *Qual Life Res*. 2015;24(10):2415–2421. doi:10.1007/s11136-015-0973-z

31. Jafari P, Bagheri Z, Safe M. Item and response-category functioning of the Persian version of the KIDSCREEN-27: Rasch partial credit model. *Health Qual Life Outcomes*. 2012;10(1):127. doi:10.1186/1477-7525-10-127

32. Aires MT, Auquier P, Robitail S, Werneck GL, Simeoni M-C. Cross-cultural adaptation and psychometric properties of the Brazilian-Portuguese version of the VSP-A (Vécu et Santé Perçue de l'Adolescent), a health-related quality of life (HRQoL) instrument for adolescents, in a healthy Brazilian population. *BMC Pediatr*. 2011;11(1):8. doi:10.1186/1471-2431-11-8

33. Vélez C-M, Lugo-Agudelo L-H, Hernández-Herrera G-N, García-García H-I. Colombian Rasch validation of KIDSCREEN-27 quality of life questionnaire. *Health Qual Life Outcomes*. 2016;14 (1):67. doi:10.1186/s12955-016-0472-0

34. Ravens-Sieberer U, Herdman M, Devine J, et al. The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. *Qual Life Res*. 2014;23(3):791–803.

35. Erhart M, Hagquist C, Auquier P, et al. A comparison of Rasch item-fit and Cronbach's alpha item reduction analysis for the development of a quality of life scale for children and adolescents. *Child Care Health Dev*. 2010;36(4):473–484.

36. Robitail S, Ravens-Sieberer U, Simeoni M-C, et al. Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Qual Life Res*. 2007;16(8):1335–1345. doi:10.1007/s11136-007-9241-1

37. Chae S, Park E-Y, Choi Y-I. The psychometric properties of the Childhood Health Assessment Questionnaire (CHAQ) in children with cerebral palsy. *BMC Neurol*. 2018;18(1):151. doi:10.1186/s12883-018-1154-9

38. Park MS, Chung CY, Lee KM, et al. Rasch analysis of the pediatric outcomes data collection instrument in 720 patients with cerebral palsy. *J Pediatr Orthop*. 2012;32(4):423–431. doi:10.1097/BPO.0b013e31824b2a1f

39. Erhart M, Ravens-Sieberer U, Dickinson HO, Colver A. Rasch measurement properties of the KIDSCREEN quality of life instrument in children with cerebral palsy and differential item functioning between children with and without cerebral palsy. *Value Health*. 2009;12 (5):782–792. doi:10.1111/j.1524-4733.2009.00508.x

40. Gothwal VK, Bharani S, Mandal AK. Parent-child agreement on health-related quality of life in congenital glaucoma. *Transl Vis Sci Technol*. 2018;7(4):15. doi:10.1167/tvst.7.4.15

41. Jervaeus A, Kottorp A, Wettergren L. Psychometric properties of KIDSCREEN-27 among childhood cancer survivors and age matched peers: a Rasch analysis. *Health Qual Life Outcomes*. 2013;11(1):96. doi:10.1186/1477-7525-11-96

42. Robitail S, Simeoni M-C, Erhart M, et al. Validation of the European proxy KIDSCREEN-52 pilot test health-related quality of life questionnaire: first results. *J Adolesc Health*. 2006;39(4):596.e1–e10. doi:10.1016/j.jadohealth.2006.01.009

43. Huang I-C, Wen P-S, Revicki DA, Shenkman EA. Quality of life measurement for children with life-threatening conditions: limitations and a new framework. *Child Indic Res*. 2011;4(1):145–160. doi:10.1007/s12187-010-9079-x

44. Huang I-C, Quinn GP, Wen P-S, et al. Using three legacy measures to develop a health-related quality of life tool for young adult survivors of childhood cancer. *Qual Life Res*. 2012;21(8):1437–1450. doi:10.1007/s11136-011-0055-9

45. Caronni A, Sciumè L, Donzelli S, Zaina F, Negrini S. ISYQOL: a Rasch-consistent questionnaire for measuring health-related quality of life in adolescents with spinal deformities. *Spine J*. 2017;17 (9):1364–1372. doi:10.1016/j.spinee.2017.05.022

46. Simeoni M-C, Schmidt S, Muehlan H, Debensason D, Bullinger M, Group D. Field testing of a European quality of life instrument for children and adolescents with chronic conditions: the 37-item DISABKIDS chronic generic module. *Qual Life Res*. 2007;16 (5):881–893. doi:10.1007/s11136-007-9188-2

47. Landfeldt E, Mayhew A, Straub V, Lochmüller H, Bushby K, Lindgren P. Psychometric analysis of the pediatric quality of life inventory 3.0 neuromuscular module administered to patients with duchenne muscular dystrophy: a Rasch analysis. *Muscle Nerve*. 2018;58(3):367–373. doi:10.1002/mus.26109

48. Bray P, Bundy AC, Ryan MM, North KN. Can in-the-moment diary methods measure health-related quality of life in Duchenne muscular dystrophy? *Qual Life Res*. 2017;26(5):1145–1152. doi:10.1007/s11136-016-1442-z

49. Strong C, Lin Y-C, Tsai M-C, Lin C-Y. Factor structure of sizing me up, a self-reported weight-related quality of life instrument, in community children across weight status. *Child Obes*. 2017;13 (2):111–119. doi:10.1089/chi.2016.0259

50. Bower W, Wong E, Yeung C. Development of a validated quality of life tool specific to children with bladder dysfunction. *Neurourol Urodyn*. 2006;25(3):221–227. doi:10.1002/nau.20171

51. Caronni A, Zaina F, Negrini S. Improving the measurement of health-related quality of life in adolescent with idiopathic scoliosis: the SRS-7, a Rasch-developed short form of the SRS-22 questionnaire. *Res Dev Disabil*. 2014;35(4):784–799. doi:10.1016/j.ridd.2014.01.020

52. Henry B, Aussage P, Grosskopf C, Goehrs J-M. Development of the Cystic Fibrosis Questionnaire (CFQ) for assessing quality of life in pediatric and adult patients. *Qual Life Res*. 2003;12(1):63–76. doi:10.1023/A:1022037320039

53. Khadka J, Ryan B, Margrain TH, Woodhouse JM, Woodhouse JM. Development of the 25-item Cardiff visual ability questionnaire for children (CVAQC). *Br J Ophthalmol*. 2010;94(6):730–735. doi:10.1136/bjo.2009.171181

54. Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. *Qual Life Res*. 2017;26(8):2025–2039.

55. Smidt AC, Lai J-S, Cella D, Patel S, Mancini AJ, Chamlin SL. Development and validation of Skindex-Teen, a quality-of-life instrument for adolescents with skin disease. *Arch Dermatol*. 2010;146 (8):865–869. doi:10.1001/archdermatol.2010.161

56. Basra M, Salek M, Fenech D, Finlay AY. Conceptualization, development and validation of T-QoL©(Teenagers' Quality of Life): a patient-focused measure to assess quality of life of adolescents with skin diseases. *Br J Dermatol*. 2018;178(1):161–175. doi:10.1111/bjd.15853

57. Nielsen JB, Kyvsgaard JN, Sildorf SM, Kreiner S, Svensson J. Item analysis using Rasch models confirms that the Danish versions of the DISABKIDS® chronic-generic and diabetes-specific modules are valid and reliable. *Health Qual Life Outcomes*. 2017;15(1):44. doi:10.1186/s12955-017-0618-8

58. Smith R, Linacre J, Smith E. Guidelines for manuscripts. *J Appl Meas*. 2003;4(2):198–204.

59. Mokkink LB, Prinsen CA, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther*. 2016;20(2):105–113. doi:10.1590/bjpt-rbf.2014.0143