Volume 11 • Number 4 • 2008
VALUE IN HEALTH

The KIDSCREEN-52 Quality of Life Measure for Children and Adolescents: Psychometric Results from a Cross-Cultural Survey in 13 European Countries

Ulrike Ravens-Sieberer, Prof,[1] Angela Gosch, PhD,[2] Luis Rajmil, Prof,[3] Michael Erhart, DiplPsych,[1] Jeanet Bruil, PhD,[4] Mick Power, Prof,[5] Wolfgang Duer, PhD,[6] Pascal Auquier, Prof,[7] Bernhard Cloetta, Prof,[8] Ladislav Czemy, PhD,[9] Joanna Mazur, PhD,[10] Agnes Czimbalmos, PhD,[11] Yannis Tountas, Prof,[12] Curt Hagquist, Prof,[13] Jean Kilroe, PhD,[14] the KIDSCREEN Group

[1]School of Public Health, WHO Collaborating Center for Child and Adolescent Health Promotion, University of Bielefeld, Germany; [2]Munich University of Applied Sciences, Munich, Germany; [3]Agency for Quality, Research and Assessment in Health (AQuRAHealth), Barcelona, Spain; [4]TNO, Prevention and Health, Leiden, The Netherlands; [5]Department of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK; [6]Ludwig Boltzmann-Institute for Sociology of Health and Medicine, University of Vienna, Vienna, Austria; [7]Department of Public Health, University Hospital of Marseille, Marseille, France; [8]Social and Behavioral Health Research, Department of Social and Preventive Medicine, University of Berne, Berne, Switzerland; [9]Prague Psychiatric Center, Prague, Czech Republic; [10]Department of Epidemiology, National Research Institute of Mother & Child, Warsaw, Poland; [11]Health Promotion and Development Center, Child Health Department, Budapest, Hungary; [12]Institute of Social and Preventive Medicine, Athens, Greece; [13]Karlstad University, Karlstad, Sweden; [14]Knowledge Officer Child Health HeBE, Programme of Action for Children, Dublin, Ireland

ABSTRACT

Objective: This study assesses the reliability and validity of the European KIDSCREEN-52 generic health-related quality of life (HRQoL) questionnaire for children and adolescents.
Research Design: The KIDSCREEN-52, which measures HRQoL in 10 dimensions, was administered to a representative sample of 22,827 children and adolescents (8 to 18 years) in 13 European countries. Psychometric properties were assessed using the Classical Test Theory approach, Rasch analysis, and structural equation modeling (SEM). A priori expected associations between KIDSCREEN scales and sociodemographic and health-related factors were examined. Test–retest reliability was assessed in 10 countries.
Results: For the overall sample, Cronbach's alpha values ranged from 0.77 to 0.89. Scaling success (Multitrait Analysis Program) was >97.8% for all dimensions and Rasch analysis item fit (INFITmsq) ranged from 0.80 to 1.27. The intraclass correlation coefficients ranged from 0.56 to 0.77. No sizeable differential item functioning (DIF) was found by age, sex or health status. Four items showed DIF across countries. The specified SEM fitted the data well (root mean square error of approximation: 0.06, comparative fit index: 0.98). Correlation coefficients between Pediatric Quality of Life Inventory, Child Health and Illness Profile-Adolescent Edition, and Youth Quality of Life Instrument scales and KIDSCREEN dimensions assessing similar constructs were moderate for those ($r = 0.44$ to 0.61). Statistically significant differences between children with and without physical and mental health problems (Children with Special Health Care Needs screener: $d = 0.17$ to 0.42, Strengths and Difficulties Questionnaire: $d = 0.32$ to 0.72) were found in all dimensions. All dimensions showed a gradient according to socioeconomic status.
Conclusions: The KIDSCREEN-52 questionnaire has acceptable levels of reliability and validity. Further work is needed to assess longitudinal validity and sensitivity to change.
Keywords: child and adolescents health, cultural sensitivity, measurement, quality of life, research methodology.

Address correspondence to: Ulrike Ravens-Sieberer, School of Public Health, University of Bielefeld, PB 100131, D-33501 Bielefeld, Germany. E-mail: U.Ravens-Sieberer@uni-bielefeld.de
10.1111/j.1524-4733.2007.00291.x

KIDSCREEN international co-ordinator in chief:
Ulrike Ravens-Sieberer, Robert Koch Institute, Berlin, Germany.
Members of the KIDSCREEN Group:
Austria: Wolfgang Duer, Kristina Fuerth; Czech Republic: Ladislav Czemy; France: Pascal Auquier, Marie-Claude Simeoni, Stephane Robitail; Germany: Ulrike Ravens-Sieberer, Michael Erhart, Jennifer Nickel, Bärbel-Maria Kurth, Angela Gosch, Ursula von Rüden; Greece: Yannis Tountas, Christina Dimitrakakis; Hungary: Agnes Czimbalmos, Anna Aszman; Ireland: Jean Kilroe, Celia Keenaghan; The Netherlands: Jeanet Bruil, Symone Detmar, Eric Veripps; Poland: Joanna Mazur, Ewa Mierzejeswka; Spain: Luis Rajmil, Silvina Berra, Cristian Tebé, Michael Herdman, Jordi Alonso; Sweden: Curt Hagquist; Switzerland: Thomas Abel, Corinna Bisegger, Bernhard Cloetta, Claudia Farley; United Kingdom: Mick Power, Clare Atherton, Katy Phillips.
Advisors: Jordi Alonso, Jacob Bjorner; Stef van Buuren, Michael Rigby, Alan Tennant, John Ware, Liz Waters.

 1098-3015/08/645 645–658

Introduction

The measurement of health-related quality of life (HRQoL) in adults has seen rapid advances in recent years, with the development of many generic and disease-specific instruments [1]. Interest in developing similar instruments for children and adolescents has also increased, and several generic and disease-specific measures are now available for use in these groups [2]. Generic HRQoL measures can be useful in identifying subgroups of children and adolescents who are at risk for health problems, and can assist in determining the burden of a particular disease or disability [3]. This type of measure can also be used to evaluate health service needs and thereby influence public policy decisions, promote policies and legislation related to child and adolescent health, and aid in the allocation of health-care resources [3]. Furthermore, monitoring population health status should allow us to track health trends and identify inequalities in health, to plan preventive strategies and, consequently, help to improve population health.

To date, the majority of HRQoL questionnaires for children and adolescents has been generated within one country and has subsequently translated into other languages [4,5]. An alternative to this approach is to elicit questionnaire content simultaneously in several different countries, thereby helping to provide a broad perspective on lay interpretations of HRQoL across different countries [6,7]. It has been recommended that the simultaneous approach should include the exploration of the relevant dimensions of HRQoL for inclusion in the questionnaire for different age groups and across different countries, and that initial scale content should be based on content generated by focus groups in different countries [8].

The KIDSCREEN project was funded by the European Commission (EC) to produce the first generic instrument for use in children and adolescents to be developed using this methodology [9]. Modern techniques for the development and cross-cultural validation of HRQoL, such as structural equation modeling (SEM) and Item Response Theory were applied. The result of this project was the generic KIDSCREEN-52 HRQoL questionnaire [10]. The KIDSCREEN project also cooperated closely with the EC-funded DISABKIDS project, which aimed to develop a chronic-generic and disease-specific instrument using a similar approach [11].

The aim of this article is to provide an overview of the development and psychometric characteristics of the KIDSCREEN-52 HRQoL questionnaire. It presents the results of tests using the Rasch measurement model as well as tests for internal consistency and test–retest reliability. The instrument's structural, known-groups, and convergent validity are discussed, and we examine the degree to which the instrument meets the requirements of instrument validation proposed by the Medical Outcomes Trust Scientific Committee [12].

Methods

Subjects and Settings

The study was carried out in 13 European countries. Three different approaches to sample selection were used: in six countries (Austria [AT], Germany [DE], Switzerland [CH], Spain [ES], France [FR], The Netherlands [NL]), address sampling was conducted via computer-assisted telephone interviews. Questionnaires were sent by post to families who agreed to participate. The questionnaires were completed in the respondent's home and sent back to the national centers in a prepaid envelope. In five countries (Greece [EL], Hungary [HU], Ireland [IE], Poland [PL], Sweden [SE]), the samples were obtained in schools that were representative of the country as a whole in terms of school type (private vs. public, rural vs. urban). The pupils completed the questionnaires during class time and took a copy of the parent questionnaire home with them (in PL, the parent version of the questionnaire was administered by post). Parents were asked to fill in the questionnaires and to send them back to the appropriate national center using a prepaid envelope. The UK combined both telephone and school administration, and Czech-Republic (CZ) carried out multistage random sampling of communities and households. The national samples were representative in terms of age and sex structure, though they are not proportional to population size. Sample representativity, study design, sampling methods, and administration procedures by country are described in detail elsewhere, together with an assessment of external validity [13]. Additional material is also provided in the appendix on the Value in Health web page. Test–retest reliability was determined by administering the KIDSCREEN-52 HRQoL questionnaire for children and adolescents on two separate occasions, 2 weeks apart, to 10% of the initial sample in 10 countries.

Measures

Respondents completed a series of measures to be able to test the convergent validity of the new questionnaire.

KIDSCREEN-52 HRQoL questionnaire for children and adolescents. Content for the KIDSCREEN-52 generic questionnaire was generated through literature reviews, expert consultation using the Delphi method [14], and focus groups with children and adolescents in all participating countries. This enabled HRQoL dimensions and items which were relevant to respondents in all countries to be identified [15]. The statements derived from focus groups were reduced after

EUROHIS guidelines [16], by excluding redundant items and items which were difficult to understand or poorly accepted during cognitive debriefing [15]. The resulting pilot measure included 179 items and was translated from English into the languages of the participating countries according to international guidelines using a forward-backward-forward translation technique and international harmonization sessions [17,18]. For the item reduction analysis, a multinational pilot study was carried out in a sample of more than 3900 children, adolescents and their parents, in seven European countries. In addition to applying standard psychometric analyses, item response theory and SEM were used to determine the measure's optimal item and scale characteristics.

After item reduction, the final questionnaire consisted of 52 items to be answered by the child/adolescent and which assessed HRQoL in 10 dimensions: Physical Well-Being explores the level of the child's/adolescent's physical activity, energy and fitness; Psychological Well-Being examines the psychological well-being of the child/adolescent, including positive emotions and satisfaction with life; Moods and Emotions covers the experience of depressive moods and emotions, and stressful feelings; Self-Perception explores whether respondents perceive their bodily appearance positively or negatively and body image is explored in questions dealing with satisfaction with looks as well as with clothes and other personal accessories; Autonomy looks at the respondents' opportunities to shape their social and leisure time; Relations with Parents and Home Life examines the child's relationship with his or her parents and the atmosphere at home; Social Support and Peers examines the nature of the respondents' relationships with other children/adolescents; School Environment explores the child's/adolescent's perceptions of his/her cognitive capacity, learning and concentration, and his/her feelings about school; Social Acceptance (Bullying) covers the aspect of feeling rejected by peers at school; and Financial Resources assesses the respondents' perceptions of their financial resources.

The KIDSCREEN items use 5-point Likert-type scales to assess either the frequency (never-seldom-sometimes-often-always) of certain behaviors/feelings or, in a smaller number of cases, the intensity of an attitude (not at all-slightly moderately very-extremely). The recall period is 1 week. Rasch scores are computed for each dimension and are transformed into *T*-values with a mean of 50 and a standard deviation of 10; higher scores indicate better HRQoL and well-being. KIDSCREEN-52 T scores refer to the mean values and standard deviations from a multinational European sample described later in this article.

The Pediatric Quality of Life Inventory. The Pediatric Quality of Life Inventory (PedsQL) 4.0 generic child self-reported core scales are scales for measuring HRQoL in healthy and ill children and adolescents [19]. The PedsQL 4.0 generic core scales consist of 23 items which cover the Physical, Emotional, Social, and School-related aspects of HRQoL. The validated UK version of the PedsQL [20] was included in the UK and IE survey. Strong associations were expected between the PedsQL-Physical dimension and the KIDSCREEN Physical Well-Being dimension, the PedsQL-Emotional domain and the KIDSCREEN Psychological Well-Being and Moods and Emotions dimension, the PedsQL-Social domain and the KIDSCREEN Social Acceptance dimensions, and between the PedsQL-School domain and the KIDSCREEN School Environment dimension.

Child Health and Illness Profile-Adolescent Edition. The Satisfaction domain of the Child Health and Illness Profile-Adolescent Edition (CHIP-AE), a generic measure of health status [21], was answered by adolescents aged 12 years or older in all countries. For the analysis, the scores were transformed to a scale on which higher values indicate better QoL. The strongest associations between the CHIP-AE Satisfaction domain and the KIDSCREEN were expected for the KIDSCREEN Psychological Well-Being and Physical Well-Being dimensions. For the KIDSCREEN project, the CHIP-AE was adapted from the US version after internationally recommended procedures [17,18] for use in all participating countries.

The Youth Quality of Life Instrument-Surveillance Version. The Youth Quality of Life Instrument-Surveillance Version (YQOL-S) is a 13-item generic QoL questionnaire designed to assess QoL among adolescents aged 11 to 18 years [22,23]. The YQOL-S provides one perceptual QoL score in which higher scores indicate better QoL. The YQOL-S was completed by adolescents aged 12 years and older, in all countries. Strong associations between the YQOL-S and the KIDSCREEN dimensions of Psychological Well-Being, Moods and Emotions and Self-Perception were expected based on previously published information about the development and validation of the YQOL [23] and also after evaluating the content of the YQOL-S questionnaire. The YQOL-S was adapted from the US version for use in all participating countries.

Physical and emotional health status. The Children with Special Health Care Needs (CSHCN) screener was included in all participating countries except Ireland as a valid measure of physical and emotional chronic health problems [24,25]. The CSHCN contains five sequences of questions: the first question in each section is followed by two additional questions which ask about the presence and duration of any

health conditions. The five principal questions address the use of or need for prescription medication; the use of or need for medical, mental health or educational services; functional limitations; use of and need for specialized therapies (for example, occupational therapy, physiotherapy, speech therapy); and treatment or counseling for emotional or developmental problems. The CSHCN screener has been field-tested in several studies and 12% to 20% children in population-based samples were identified as having chronic or special health-care needs [26]. Based on existing literature, children with a chronic condition were expected to have poorer HRQoL [2]. A chronic health condition was expected to be associated with lower scores in all KIDSCREEN dimensions, but particularly for the dimension of Physical Well-Being.

Mental health status. The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioral screening questionnaire aimed at 3- to 16-year-olds, which asks about children's and teenagers' symptoms and positive attitudes [27]. The SDQ had been previously adapted and validated in all participating countries (http://www.sdqinfo.com/b3.html). The instrument asks about positive or negative attributes using 25 items covering emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. Responses for items are summed to generate a total difficulties score. The SDQ parent report version was included as a screening instrument for mental health in all countries except Ireland. The presence of mental disorders was expected to manifest itself through lower scores in all KIDSCREEN dimensions but particularly in the Psychological Well-being, Moods and Emotions and Self-Perception dimensions.

Family socioeconomic status. The Family Affluence Scale (FAS), a socioeconomic indicator for use in child populations, includes questions on family car ownership, having one's own unshared room, the number of computers in the home, and the time children spent on holiday in the previous 12 months [28]. The FAS was collected in seven categories (from 0, the lowest, to 7, the highest FAS category) and was recoded into three groups for analysis (low FAS [0 to 3], intermediate [4 to 5], and high FAS [6 to 7]). Low family affluence was expected to be associated with lower scores on all KIDSCREEN dimensions, but particularly on Financial Resources.

Statistical and Psychometric Analysis

Several psychometric characteristics were studied, including reliability, convergent validity, and construct validity. A previous study had showed that there were no significant (>15%) floor effects in any KIDSCREEN-52 dimensions, though some ceiling effects were observed in three dimensions (15.5%, 24.5%, and 49.1% of ceiling effect in the dimensions Relations with Parents and Home Life, Financial Resources, and Social Acceptance). Missing value rates were negligible (<2.9%) for all dimensions [29].

Measurement model and reliability. The internal consistency of the KIDSCREEN-52 dimensions was calculated using Cronbach's alpha [30]. Alpha coefficients of 0.7 or higher were considered acceptable. Scaling success was assessed by examining the percentage of items which had a higher correlation with the scale to which they hypothetically belonged than with other scales in the questionnaire. Scaling success was examined using the Multitrait Analysis Program [31]. Test–retest stability was determined by administering the KIDSCREEN-52 on two separate occasions, 2 weeks apart, to 10% of the initial sample in 10 countries. Intraclass correlation coefficients (ICC) were computed for the dimension scores [32]. A coefficient of 0.6 or higher was considered evidence of adequate test–retest stability [33].

A confirmatory factor analysis was conducted by specifying an SEM according to the 10-dimension KIDSCREEN-52 model. The categorical nature and non-normal distribution of the KIDSCREEN items was taken into account by calculating the matrix of polychoric correlations using the PRELIS 2.72 software [34] as the database for estimating the model parameters using the LISREL 8.72 software [35]. Because each observed variable only loads on one respective latent exogenous construct, the identifiability of the parameters of the measurement model was ensured [36]. The variance of the latent constructs was fixed at one during parameter estimation, but correlation between the latent constructs was allowed. The subsequent complete standardization of the model enabled parameters to be estimated correctly [36]. The model parameters were estimated according to the maximum likelihood criterion. The global goodness of fit (GoF) of the model was assessed via the root mean square error of approximation (RMSEA) [37] and the comparative fit index (CFI) [38]. A CFI larger than 0.95 (0.90) and an RMSEA lower than 0.06 (0.8) were considered to constitute an excellent (adequate) fit between the specified model and the data [39]. The statistical significance and the magnitude of the estimated factor pattern coefficients were analyzed, as well as the largest cross-loading (structure coefficient). The analyses were repeated separately for children (8 to 11 years) and adolescents (12 to 18 years) and the results were compared by comparing an unconstrained model (separate parameter estimates) with a constrained model in which factor pattern coefficients and error terms were set to be equal for both age groups. Where the chi-square value of the difference between the constrained and unconstrained model were signifi-

cant, the standardized pattern coefficients of the unconstrained model were transformed to Fisher-*Z*-values [40] and the differences were computed to assess the magnitude of the deviation.

The test data were subsequently analyzed using the partial credit model (PCM), from the Rasch family of models. The PCM tries to explain the actual behavior of the responders in the testing situation by means of the estimated person parameter and the location of the item-answer-category thresholds. The PCM assumes that item thresholds and persons parameter values can be ordered on the same unidimensional latent trait continuum [41]. Item threshold parameters were estimated using a marginal maximum likelihood procedure and person parameters were estimated using Warm's modified weighted maximum likelihood (WLE) method. The GoF was tested for every item by calculating the infit mean square. An item functions well if the infit mean square is between 0.80 (0.70) and 1.20 (1.30) [42]. Item and person parameter estimates were calculated using WINMIRA software [43]. Infit mean square values were calculated using MULTIRA software [44]. Among the various item response theory (IRT) models, the PCM was chosen because it provides powerful tests of item fit, enables interval-scaled measurement and invariant comparison of item and persons. Nevertheless, within the pilot item-reduction analysis, the parametric generalized PCM and a nonparametric functional approach were conducted beforehand as well.

Zumbo's ordinal logistic regression approach [45] was used to identify items displaying Differential item functioning (DIF) across countries, age groups (8 to 11 years vs. 12 to 18 years), sexes, and children with or without special health-care needs as determined by the CSHCN screener [24]. A DIF effect size of $R^2 = 0.035$ was considered to be evidence of noticeable DIF [46].

Validity. Convergent validity was assessed by correlating KIDSCREEN-52 dimension scores with scores on other validated questionnaires measuring similar concepts. A priori hypotheses about the relative magnitude of the correlations were specified.

Pearson correlation coefficients were computed to analyze convergent and discriminant validity between the KIDSCREEN-52 and the other instruments. Convergent validity was considered to be demonstrated when correlations between comparable dimensions were significantly higher than correlations between theoretically different dimensions. Correlation coefficients between 0.1 and 0.3 were considered low, those between 0.31 and 0.5 moderate, and those over 0.5 were considered high [47].

Construct validity was further evaluated based on previously developed hypotheses regarding expected differences between healthy children and adolescents and those with physical or mental health problems, and between those reporting high and low family socioeconomic status. Construct validity was assessed by calculating between group effect sizes (ES, d) [48]. Effect sizes of 0.2 to 0.5 were considered small; those between 0.51 and 0.8 moderate, and those over 0.8 were considered large.

Results

Sample Characteristics

The final sample included 22,827 children and adolescents. Response rates varied across countries, from 24.2% to 91.2% depending on the sampling approach taken, with a higher response rate in countries which used school-based sampling. Table 1 shows the sociodemographic characteristics of the final sample, overall and by country. The mean ages for the overall child and adolescent samples were 9.7 and 14.4 years, respectively. There were slightly more girls than boys in both samples. In terms of age and sex, the child and adolescent samples were broadly similar across all participating countries. The most notable differences between countries occurred in socioeconomic status (FAS) with, for example, 49.5% of the Czech child sample reporting low FAS as compared with only 7.5% in the French sample. A similar pattern was seen in the adolescent sample.

Measurement Model and Reliability

The specified 10-dimensional structural equation model fitted the data well, with an RMSEA of 0.062 and a CFI of 0.976. Table 2a shows that all items displayed loadings (pattern coefficients) which were greater than 0.62. The largest cross-loadings were always lower than the corresponding pattern coefficient. The correlation between the dimensions of the measurement model were estimated as model parameters and ranged between 0.16 and 0.76 (see Table 2b). The separate models for children and adolescents displayed similar GoF (RMSEA = 0.062 and 0.062; CFI = 0.974 and 0.973), although a statistically significant difference was observed between younger children and adolescents, in terms of pattern coefficients, latent factor correlation, and error terms ($\chi^2_{(df104)} = 3303.75$; $P < 0.001$). Nevertheless, the actual differences between corresponding pattern coefficients were quite small and, after Fisher-Z transformation, ranged between 0.002 and 0.278, indicating marginal to medium effects. The mean Fisher-Z difference was 0.11, indicating a small effect [47].

Table 3 shows the psychometric characteristics of the KIDSCREEN-52 dimensions for the overall sample, and for some statistics the range across different countries. Cronbach alphas for the overall sample ranged from 0.77 (Social Acceptance and Bullying) to

Table 1 Response rate and sociodemographic characteristics of the KIDSCREEN sample

	Total (n = 22,827)	Austria (n = 1,475)	Switzerland (n = 1,701)	Czech Republic (n = 1,592)	Germany (n = 1,723)	Spain (n = 876)	France (n = 1,049)	Greece (n = 1,174)	Hungary (n = 3,237)	Ireland (n = 1,240)	The Netherlands (n = 1,885)	Poland (n = 1,715)	Sweden (n = 3,283)	United Kingdom (n = 1,877)
Response rate (%)	68.9	35.3	40.2	71.5	40.6	24.2	26.4	72.0	90.0	82.5	68.0	59.6	91.2	42.4
Children (N)	6,882	518	580	556	610	321	394	—	1,378	321	678	565	—	961
Mean age years (SD)	9.7 (1.1)	9.7 (1.1)	9.8 (1.0)	9.6 (1.0)	9.7 (1.1)	9.7 (1.1)	9.5 (1.1)	—	9.5 (1.1)	10.4 (0.7)	9.6 (1.1)	9.9 (1.0)	—	9.5 (1.0)
Age range	8–11	8–11	8–11	8–11	8–11	8–11	8–11	—	8–11	8–11	8–11	8–11	—	8–11
Female (%)	51.3	53.5	52.4	50.5	50.0	46.1	50.3	—	55.2	52.6	49.3	53.3	—	47.3
Socioeconomic status*														
Low FAS (%)	20.0	14.4	10.9	49.5	10.1	17.7	7.5	—	26.7	18.2	11.2	35.7	—	11.2
Medium FAS (%)	45.4	49.2	44.7	41.6	46.5	47.6	44.4	—	47.2	48.9	49.0	48.8	—	36.7
High FAS (%)	34.6	36.4	44.4	8.9	43.4	34.7	48.1	—	26.1	32.9	39.9	15.5	—	52.2
Adolescents (N)	15,945	957	1,121	1,036	1,113	555	655	1,174	1,859	919	1,207	1,150	3,283	916
Mean age years (SD)	14.4 (1.7)	14.5 (1.8)	14.5 (1.8)	14.9 (1.9)	14.6 (1.9)	14.7 (1.9)	14.6 (1.9)	14.6 (1.7)	14.6 (1.8)	14.6 (1.4)	14.6 (1.8)	14.8 (1.9)	13.7 (1.0)	14.1 (1.6)
Age range	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–18	12–15	12–18
Female (%)	53.8	53.9	54.2	48.9	52.0	50.8	52.8	59.7	60.8	62.2	52.2	55.5	49.0	49.7
Socioeconomic status*														
Low FAS (%)	23.3	14.1	11.3	48.9	12.5	21.7	9.0	37.3	32.3	14.4	9.4	39.1	—	14.2
Medium FAS (%)	46.5	50.2	47.1	41.4	49.3	51.3	44.0	45.1	46.5	44.4	49.2	48.0	—	41.0
High FAS (%)	30.3	35.7	41.7	9.7	38.2	27.0	47.0	17.6	21.2	41.2	41.4	12.9	—	44.8

*FAS, Family Affluence Scale (0–3 = low; 4–5 = medium; 6 to 7 = high).

Table 2 Confirmatory factor analysis with LISREL—(a) standardized pattern coefficients and largest cross-loadings (structure coefficient); (b) confirmatory latent factor intercorrelation (N = 19,612)

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
(a) Item labels										
Physical Well-being										
In general, how would you say health is?	**0.62**	0.44	0.35	0.38	0.32	0.31	0.29	0.33	0.10	0.20
Have you felt fit and well?	**0.80**	0.57	0.46	0.49	0.41	0.41	0.37	0.42	0.13	0.27
Have you been physically active?	**0.66**	0.47	0.38	0.40	0.34	0.33	0.30	0.35	0.11	0.22
Have you been able to run well?	**0.70**	0.50	0.40	0.43	0.36	0.36	0.32	0.37	0.11	0.23
Have you felt full of energy?	**0.78**	0.56	0.45	0.48	0.40	0.40	0.36	0.41	0.13	0.26
Psychological Well-being										
Has your life been enjoyable?	0.61	**0.86**	0.65	0.57	0.52	0.59	0.50	0.56	0.19	0.34
Have you felt pleased that you are alive?	0.60	**0.83**	0.63	0.55	0.51	0.58	0.49	0.54	0.18	0.33
Have you felt satisfied with your life?	0.62	**0.87**	0.66	0.57	0.53	0.60	0.51	0.56	0.19	0.34
Have you been in a good mood?	0.53	**0.74**	0.56	0.49	0.45	0.51	0.43	0.48	0.16	0.29
Have you felt cheerful?	0.55	**0.77**	0.58	0.51	0.47	0.53	0.45	0.50	0.17	0.30
Have you had fun?	0.52	**0.72**	0.55	0.48	0.44	0.50	0.42	0.47	0.16	0.28
Moods and Emotions										
Have you felt that you do everything badly?	0.39	0.52	**0.68**	0.47	0.36	0.40	0.30	0.37	0.29	0.25
Have you felt sad?	0.43	0.57	**0.76**	0.52	0.40	0.45	0.33	0.42	0.32	0.27
Felt so bad that didn't want to do anything?	0.43	0.56	**0.74**	0.51	0.39	0.44	0.32	0.41	0.31	0.27
Felt that everything in life goes wrong?	0.47	0.63	**0.83**	0.57	0.43	0.49	0.36	0.45	0.35	0.30
Have you felt fed up?	0.42	0.55	**0.73**	0.51	0.38	0.43	0.32	0.40	0.31	0.27
Have you felt lonely?	0.42	0.55	**0.73**	0.50	0.38	0.43	0.32	0.40	0.31	0.26
Have you felt under pressure?	0.38	0.50	**0.66**	0.46	0.34	0.39	0.29	0.36	0.28	0.24
Self-Perception										
Have you been happy with way you are?	0.47	0.51	0.53	**0.76**	0.42	0.48	0.30	0.43	0.23	0.28
Have you been happy with your clothes?	0.42	0.45	0.47	**0.68**	0.38	0.43	0.26	0.39	0.20	0.25
Been worried about the way you look?	0.43	0.46	0.48	**0.69**	0.38	0.43	0.27	0.39	0.21	0.26
Felt jealous of way other girls/boys look?	0.44	0.48	0.50	**0.72**	0.40	0.45	0.28	0.41	0.21	0.27
Like to change something about your body?	0.45	0.49	0.51	**0.74**	0.41	0.46	0.29	0.42	0.22	0.27
Autonomy										
Have you had enough time for yourself?	0.38	0.45	0.38	0.41	**0.74**	0.43	0.42	0.35	0.14	0.35
Able to do things—want to do in free time?	0.42	0.50	0.43	0.45	**0.82**	0.48	0.46	0.39	0.15	0.39
Had enough opportunity to be outside?	0.42	0.49	0.42	0.45	**0.80**	0.47	0.46	0.38	0.15	0.38
Have you had enough time to meet friends?	0.40	0.47	0.40	0.43	**0.77**	0.45	0.44	0.37	0.14	0.37
Able to choose what to do in free time?	0.36	0.43	0.37	0.39	**0.70**	0.41	0.40	0.33	0.13	0.33
Parent Relation and Home Life										
Have your parent(s) understood you?	0.42	0.58	0.49	0.52	0.49	**0.83**	0.40	0.52	0.19	0.39
Have you felt loved by your parent(s)?	0.42	0.57	0.49	0.52	0.49	**0.83**	0.40	0.51	0.19	0.39
Have you been happy at home?	0.42	0.57	0.49	0.52	0.48	**0.83**	0.40	0.51	0.19	0.38
Parent(s) had enough time for you?	0.39	0.53	0.45	0.48	0.45	**0.77**	0.37	0.47	0.17	0.36
Have your parent(s) treated you fairly?	0.41	0.56	0.48	0.51	0.47	**0.81**	0.39	0.50	0.18	0.38
Aable to talk to parent(s) when wanted to?	0.40	0.54	0.46	0.49	0.46	**0.78**	0.38	0.48	0.18	0.36
Social Support and Peers										
Have you spent time with your friends?	0.33	0.42	0.31	0.28	0.41	0.35	**0.72**	0.31	0.22	0.28
Done things with other girls and boys?	0.29	0.37	0.28	0.24	0.36	0.30	**0.63**	0.27	0.20	0.24
Have you had fun with your friends?	0.39	0.49	0.36	0.32	0.47	0.40	**0.83**	0.36	0.26	0.32
You and your friends helped each other?	0.36	0.46	0.34	0.30	0.44	0.38	**0.78**	0.34	0.24	0.30
Able to talk about everything with friends?	0.33	0.42	0.31	0.27	0.40	0.34	**0.71**	0.30	0.22	0.28
Have you been able to rely on your friends?	0.35	0.44	0.33	0.29	0.42	0.36	**0.75**	0.32	0.23	0.29
School Environment										
Have you been happy at school?	0.43	0.52	0.44	0.46	0.39	0.50	0.35	**0.81**	0.15	0.31
Have you got on well at school?	0.40	0.49	0.42	0.43	0.36	0.47	0.33	**0.76**	0.15	0.29
Been satisfied with your teachers?	0.40	0.49	0.42	0.43	0.36	0.47	0.33	**0.76**	0.15	0.29
Have you been able to pay attention?	0.38	0.47	0.40	0.41	0.34	0.45	0.31	**0.72**	0.14	0.28
Have you enjoyed going to school?	0.40	0.49	0.41	0.42	0.36	0.46	0.32	**0.75**	0.14	0.29
Got along well with your teachers?	0.41	0.50	0.42	0.44	0.37	0.48	0.33	**0.77**	0.15	0.30
Social Acceptance and Bullying										
Been afraid of other girls and boys?	0.11	0.15	0.30	0.21	0.13	0.16	0.22	0.14	**0.71**	0.19
Have other girls and boys made fun of you?	0.14	0.18	0.36	0.25	0.16	0.19	0.26	0.16	**0.85**	0.23
Have other girls and boys bullied you?	0.14	0.19	0.37	0.26	0.16	0.20	0.27	0.17	**0.88**	0.23
Financial Ressources										
Enough money to do things as friends?	0.30	0.36	0.33	0.34	0.43	0.42	0.35	0.35	0.24	**0.91**
Had enough money for your expenses?	0.29	0.35	0.32	0.32	0.42	0.41	0.34	0.34	0.23	**0.87**
Enough money to do things with friends?	0.29	0.35	0.32	0.33	0.42	0.41	0.34	0.34	0.24	**0.88**
(b) Latent factors										
F1 Physical Well-being										
F2 Psychological Well-being	0.72									
F3 Moods and Emotions	0.57	0.76								
F4 Self-Perception	0.62	0.66	0.69							
F5 Autonomy	0.52	0.61	0.52	0.55						
F6 Parent Relation and Home Life	0.51	0.69	0.59	0.63	0.59					
F7 Social Support and Peers	0.46	0.59	0.44	0.39	0.57	0.48				
F8 School Environment	0.53	0.65	0.55	0.57	0.48	0.62	0.43			
F9 Social Acceptance and Bullying	0.16	0.22	0.42	0.30	0.19	0.22	0.31	0.19		
F10 Financial Resources	0.33	0.40	0.36	0.37	0.48	0.47	0.39	0.39	0.27	

Values in bold, sectandardized pattern coeffizient (SE = 0.006–0.007); Values in normal, cross loadings: structure coefficients (SE = 0.007).
SE of Correlation coefficients = 0.004–0.008.

Table 3 Internal consistency, scaling success, Rasch measurement itemfit and ordinal logistic regression DIF-analysis (Zumbo, 1999) of the KIDSCREEN-52 dimensions

Scale	No. of items	No. of cases	Internal consistency reliability		Scaling success (MAP) %	Retest reliability ICC	Rasch item fit Infit MNSQ Min–max†	Age DIF§ Delta R^2 Min–max	Sex DIF¶ Delta R^2 Min–max	Country DIF** Delta R^2 Min–max	Health DIF†† Delta R^2 Min–max
			Cronbach's alpha	(Range)* Delta R^2 Min–max							
Physical Well-Being	5	22,295	0.80	(0.75–0.86)	100	0.65	0.888–1.125	0.001–0.007	0.001–0.001	0.009–0.041	0.001–0.015
Psychological Well-Being	6	22,528	0.89	(0.85–0.91)	100	0.62	0.945–1.138	0.001–0.004	0.001–0.002	0.012–0.027	0.001–0.001
Moods and Emotions	7	22,421	0.85	(0.80–0.89)	100	0.58	0.813–1.232	0.001–0.015	0.001–0.005	0.006–0.027	0.001–0.001
Self-Perception	5	22,529	0.79	(0.71–0.85)	97.8	0.69	0.877–1.072	0.002–0.005	0.001–0.009	0.011–0.037	0.001–0.001
Autonomy	5	22,548	0.84	(0.79–0.86)	100	0.56	0.895–1.081	0.001–0.012	0.001–0.002	0.005–0.017	0.001–0.001
Parent Relation and Home Life	6	22,354	0.88	(0.85–0.90)	100	0.72	0.886–1.088	0.001–0.005	0.001–0.002	0.005–0.027	0.001–0.001
Social Support and Peers	6	22,348	0.84	(0.81–0.87)	100	0.61	0.795–1.270	0.001–0.007	0.001–0.004	0.015–0.041	0.001–0.001
School Environment	6	22,274	0.87	(0.81–0.88)	100	0.77	0.899–1.132	0.001–0.010	0.001–0.002	0.008–0.015	0.001–0.001
Social Acceptance (Bullying)	3	22,546	0.77	(0.61‡-0.83)	100	0.57	0.919–1.104	0.001–0.002	0.001–0.001	0.016–0.021	0.001–0.001
Financial Resources	3	22,226	0.89	(0.82–0.91)	100	0.68	0.966–1.021	0.001–0.003	0.001–0.001	0.002–0.007	0.001–0.001

*Range across countries.
†Minimum-maximum across items.
‡France.
§8–11 vs. 12–18.
¶Girls vs. boys.
**Germany vs. Spain, The Netherlands, Austria, UK; France, Switzerland, Hungaria, Greece, Czech-Republic, Ireland, Poland, Sweden.
††Healthy chidren vs. children with special health care needs (CSHCN screener, Bethell et al. [24]).
CSHCN, Children with Special Health Care Needs; DIF, differential item functioning; MAP, Multitrait Analysis Program; MNSQ, mean square residual.

Table 4 Convergent validity of the KIDSCREEN-52

	PedsQL				CHIP	YQOL-S
KIDSCREEN-52 dimensions	Physical functioning* *r*	Emotional functioning† *r*	Social functioning‡ *R*	School functioning§ *r*	Satisfaction domain¶ *r*	Perceptual scale** *r*
Physical Well-Being	***0.44***	0.33	0.29	0.31	***0.60***	*0.41*
Psychological Well-Being	0.32	*0.43*	0.37	0.32	*0.58*	***0.61***
Moods and Emotions	0.34	***0.53***	0.41	0.37	0.52	*0.56*
Self-Perception	0.32	0.47	0.36	0.31	0.57	0.51
Autonomy	0.29	0.36	0.35	0.23	0.44	0.40
Parent Relation and Home Life	0.29	0.39	*0.34*	0.33	0.48	0.60
Social Support and Peers	0.22	0.24	*0.36*	0.14	0.40	0.37
School Environment	0.23	0.29	0.23	***0.45***	0.45	0.47
Social Acceptance (Bullying)	0.26	0.37	***0.50***	0.18	0.24	0.24
Financial Resources	0.26	0.28	0.35	0.24	0.37	0.37

*N = 2836–2907.
†N = 2796–2861.
‡N = 2800–2865.
§N = 2770–2839.
¶N = 11,561–11,754.
**N = 10,635–10,790.
Pearson correlation coefficients.
All correlations are statistically significant at $P < 0.01$ level; SE (r) = 0.007.
Values in bold indicate the highest correlations; values in italics indicate the a priori expected highest correlations.
CHIP, Child Health and Illness Profile; PedsQL, Pediatric Quality of Life Inventory; YQOL-S, Youth Quality of Life Instrument-Surveillance Version.

0.89 (Financial Resources and Psychological Well-Being). The ICC for test–retest reliability ranged from 0.56 (Autonomy) to 0.77 (School Environment) and was only slightly below the a priori specified threshold of 0.60 in three dimensions.

Scaling success ranged from 97.8% to 100% for the different scales. All corrected item-total correlations were above $r = 0.42$. The results of the Rasch measurement analysis indicated reasonable item fit for all items with infit mean square statistics between 0.80 and 1.27. Only four items displayed slight DIF across countries, with the delta-R^2 effect size slightly above the threshold of 0.035 (the highest score was 0.041). None of the items displayed sizeable DIF by age group, sex or health status.

The analyses of the measurement model and reliability were repeated for children (8 to 11 years) and adolescents (12 to 18 years) and can be viewed in Table 2aa, 2ab, 3a, and 3b of the appendix on the Value in Health web page. Overall, the results were similar for both groups with slightly higher alpha values for the adolescents.

Validity

Convergent and discriminant validity. Table 4 shows the results of the convergent and discriminant validity analysis. KIDSCREEN-52 HRQoL and PedsQL dimensions generally displayed a small to medium level of correlation for the expected relationships. As anticipated, the PedsQL Physical Functioning scale correlated highest with the KIDSCREEN-52 Physical Well-Being dimension ($r = 0.44$), although the PedsQL Emotional Functioning Scale correlated highest with Moods and Emotions ($r = 0.53$). Contrary to expectations, its correlation with KIDSCREEN-52 Psychological Well-Being ($r = 0.43$) was lower than with the Self-Perception dimension ($r = 0.47$). The PedsQL Social Functioning Scale correlated highest with the KIDSCREEN-52 Social Acceptance (Bullying) dimension ($r = 0.50$). Finally, the PedsQL School Functioning Scale correlated highest with the KIDSCREEN-52 School Environment dimension ($r = 0.45$) as hypothesized.

As also hypothesized, the CHIP satisfaction domain correlated most highly with the KIDSCREEN-52 Physical Well-Being dimension ($r = 0.60$), followed by the Psychological Well-Being ($r = 0.58$) and Self-Perception ($r = 0.57$) dimensions. The YQOL Perceptual scale correlated highest with Psychological Well-Being ($r = 0.61$), as expected, followed by Relations with Parents and Home Life ($r = 0.60$), and Moods and Emotions ($r = 0.56$). Higher correlations were expected a priori with Moods and Emotions and Self-Perception, than with Relations with Parents and Home Life.

Expected differences for physical health and mental health status. Table 5 shows differences in KIDSCREEN-52 dimension scores by Physical and Mental Health Status. The statistically significant differences between healthy and ill children, as measured by the CSHCN screener, that were expected a priori were found for all KIDSCREEN-52 dimensions, but an effect size of 0.42 was seen only for the Physical Well-Being domain. Children and adolescents with scores on the SDQ indicating poorer mental health scored lower on all KIDSCREEN-52 dimensions than children with better mental health, with effect sizes between the two groups ranging from 0.32 to 0.72.

Table 5 Differences in KIDSCREEN-52 dimension scores by chronic health conditions and mental health status

	CSHCN					SDQ parents						
	No condition*		Some conditions†			Normal‡		Borderline§		Abnormal¶		
KIDSCREEN-52 dimensions	Mean *T*-value**	SD	Mean *T*-value††	SD	Effect size *d*‡‡	Mean *T*-value**	SD	Mean *T*-value§§	SD	Mean *T*-value¶¶	SD	Effect size *d*‡‡
Physical Well-Being	51.11	9.89	46.88	10.06	0.42	51.15	9.85	48.02	9.72	47.12	10.76	0.40
Psychological Well-Being	50.66	9.76	47.57	9.93	0.32	51.02	9.56	46.91	9.98	45.21	10.40	0.59
Moods and Emotions	50.46	9.68	48.09	9.76	0.24	51.02	9.57	46.01	8.85	44.02	9.28	0.72
Self-Perception	50.41	9.83	48.58	9.75	0.19	50.73	9.76	47.62	9.44	46.58	9.69	0.42
Autonomy	50.23	9.93	48.56	9.77	0.17	50.50	9.83	47.63	9.74	47.34	10.18	0.32
Parent Relation and Home Life	49.94	9.72	47.61	9.80	0.24	50.42	9.54	45.85	9.48	44.78	10.21	0.58
Social Support and Peers	49.99	9.89	46.97	10.07	0.30	50.29	9.75	46.51	9.68	45.22	10.76	0.51
School Environment	50.29	9.93	48.13	9.98	0.22	50.76	9.83	46.27	9.33	45.10	9.92	0.57
Social Acceptance (Bullying)	49.82	9.74	47.13	10.83	0.27	50.27	9.45	46.11	10.82	43.62	11.66	0.67
Financial Resources	49.81	9.98	47.69	10.34	0.21	50.27	9.81	46.29	9.83	44.81	10.81	0.54

*Range of N = 13,575–13,929.
†Range of N = 1630–1664.
‡Range of N = 13,222–13,554.
§Range of N = 1018–1054.
¶Range of N = 1091–1121.
**SE = 0.09.
††SE = 0.25.
‡‡Effect size (d) is calculated dividing the mean difference by the overall standard deviation.
§§SE = 0.31.
¶¶SE = 0.30.
All mean differences are statistically significant at $P < 0.01$ level.
Effect sizes in the SDQ column are for comparisons between the highest and lowest categories.
CSHCN, Children with Special Health Care Needs; SDQ, Strengths and Difficulties Questionnaire.

Table 6 Differences in KIDSCREEN-52 dimension scores by socioeconomic categories

	Familial affluence (FAS)						
	Low*		Medium†		High‡		Effect size *d* (high vs. low)
KIDSCREEN-52 dimensions	Mean *T*-value§	SD	Mean *T*-value¶	SD	Mean *T*-value**	SD	
Physical Well-Being	48.13	10.14	50.21	9.81	51.56	9.85	0.34
Psychological Well-Being	47.64	10.21	50.13	9.78	51.33	9.66	0.37
Moods and Emotions	47.86	9.82	49.87	9.67	50.94	9.67	0.32
Self-Perception	47.94	9.97	49.99	9.91	51.05	9.86	0.31
Autonomy	47.94	10.37	49.86	10.01	50.77	9.76	0.28
Parent Relation and Home Life	47.38	10.20	49.83	9.91	50.69	9.52	0.33
Peers and Social Support	48.07	10.26	49.85	9.82	50.97	9.92	0.29
School Environment	47.62	9.96	49.85	10.02	50.74	9.97	0.31
Bullying	48.79	10.47	49.48	10.04	49.78	9.89	0.10
Financial Resources	45.19	10.11	49.73	9.68	52.69	9.33	0.75

*Range of N = 3394–3452.
†Range of N = 7038–7239.
‡Range of N = 4784–4925.
§SE = 0.15.
¶SE = 0.11.
**SE = 0.13.
All mean differences are statistically significant at $P < 0.01$ level.
FAS, Family Affluence Scale.

Other expected differences in HRQoL: socioeconomic status. Table 6 shows mean *T*-values for the KIDSCREEN-52 dimensions stratified by FAS. A gradient was observed on all KIDSCREEN-52 dimensions, with children from families in the lower socioeconomic categories having poorer HRQoL. Effect sizes between those in the high and low FAS categories ranged from 0.11 on the dimension of Social Acceptance (Bullying) to 0.75 for the Financial Resources dimension.

Discussion

The KIDSCREEN-52 HRQoL questionnaire is the first instrument for children and adolescents to be developed simultaneously in several different countries with subsequent testing in a large number of children and adolescents in a representative sample. This method of simultaneous development makes it possible to incorporate potential content for the instrument from informants in different countries from the outset and can provide insight into similarities and differences between countries as regards perceptions of the concept of HRQoL [17]. In principle at least, the method should also help to ensure that instrument content will be relevant across the countries in which focus groups are performed. Clearly, the fact that the instrument was developed within European countries means that the items and dimensions of the KIDSCREEN-52 HRQoL questionnaire, although relevant to European children and adolescents, may not be equally relevant to children and adolescents outside Europe.

Few studies have applied a combination of classical test theory with modern techniques, such as item response theory and SEM, within the process of instrument development. The combination of both classical and modern methods has contributed to building a sound measure with good psychometric properties, which meet most of the MOT Scientific Committee's recommendations [12]. Nevertheless, three of the dimensions (autonomy, social acceptance, moods, and emotions) failed to meet generally accepted standards for test–retest reliability. Further investigation of the instrument's test–retest reliability is required together with its ability to detect changes over time.

The KIDSCREEN-52 measurement model covers aspects of children's and adolescents' HRQoL beyond the existing HRQoL measurement armamentarium for this population [49], for example by assessing both positive affect and negative affect, autonomy, social acceptance (bullying) and financial resources, aspects not addressed by most comparable instruments.

The 10 dimensions of the KIDSCREEN-52 HRQoL questionnaire enable true cross-cultural measurement on an interval-scaled level by fulfillling the assumptions of the Rasch model and displaying no to slight DIF. The use of these techniques has furthermore shown that the items work equally well across all language versions and across age and sex. In other words, users of the instrument can be reasonably confident that any differences detected between countries or by age and sex are likely to be true differences. Given that all dimensions met the conditions of the Rasch model, the summation of item scores is justified [41]. Furthermore, confirmatory factor analysis corroborated the validity of the dimensional structure in the KIDSCREEN measurement model. To a certain extent factorial invariance across age groups was also demonstrated in separate analyses by age group. In spite of statistically significant deviations, the differences between children and adolescents—in terms of

factor loading coefficients—were generally small, thus confirming the previous finding of no sizeable DIF by age group. This means that the KIDSCREEN items and dimensions are suitable for comparing children (8 to 11 years) and adolescents (12 to 18 years).

Analyses of convergent validity generally indicated that the KIDSCREEN measurement model displayed a reasonable pattern of associations. Correlations between PedsQL and KIDSCREEN dimensions demonstrate convergent and discriminant validity if it is considered that there is evidence of convergent validity when medium to large effects (e.g., >0.30) are present, although discriminant validity consists of small correlations (e.g., 0.10 to 0.29) [49]. For example, the physical functioning dimensions correlated 0.44. The emotional functioning dimensions correlated 0.53. The school functioning dimensions correlated 0.45. Finally, an analysis of the PedsQL social functioning scale reveals that it is measuring other kids teasing the child and not getting along with other kids. This scale correlates 0.50 with the KIDSCREEN social acceptance scale, and is thus conceptually consistent with convergent validity.

The KIDSCREEN-52 HRQoL questionnaire discriminated well and in the hypothesized direction between children and adolescents in good health and those with poorer physical or mental health as measured by the CSCHN and SDQ. The fact that effect sizes were larger when children were classified using the SDQ in comparison with the CSCHN might be due to the fact that the former focuses more on psychological and emotional aspects of health, and many of the KIDSCREEN dimensions also focus on these aspects, whereas the CSCHN places a slightly greater emphasis on problems of physical health. More research using the KIDSCREEN-52 HRQoL questionnaire in different clinical conditions is required to determine response patterns associated with those conditions to be identified. In analyses reported here and elsewhere [29], the KIDSCREEN scales also discriminated well and in the anticipated direction between children classified according to age group, sex, and socioeconomic category. These results suggest that the instrument is likely to be useful in both epidemiological and clinical settings.

One limitation of the study was the identification of subgroups with impaired physical and mental health status using screening instruments, without further clinical information. Thus the information on health status may lack reliability and validity. From a test-theoretical point of view, this might lead to an underestimation of the strength of association and the validity coefficients [33]. Future testing of the KIDSCREEN-52 should be performed in settings where clinical diagnoses and information about the severity of conditions is available. Another limitation of the study was the impossibility of testing the KIDSCREEN-52's sensitivity to change because of the cross-sectional study design. The instrument's sensitivity to change should be tested within a longitudinal interventional study using a randomized control design. Finally, differences in response rates could affect study results. Nevertheless, a comparison of the present sample with a representative sample for each country obtained from the EuroStat database indicated only very small differences in terms of age and sex between the KIDSCREEN sample and the EuroStat reference sample. The use of resampling techniques such as bootstrapping indicated that these slight differences had no appreciable impact on KIDSCREEN-52 scores. Likewise, the large sample size and the spread of characteristics of children and adolescents in each of the countries guarantees a diversity of responses which are useful for validity studies such as this.

The international, collaborative nature of the KIDSCREEN project provided many challenges in terms of producing an instrument which was conceptually and linguistically appropriate for use in many different countries. By giving each country the opportunity to be involved as early on as the item construction phase, the KIDSCREEN project provided an opportunity to include the perspective of children from many different backgrounds at the instrument development stage.

Using the KIDSCREEN-52 HRQoL questionnaire as a starting point, two shorter versions of the instrument have been produced (the KIDSCREEN-27 and the KIDSCREEN-10 Index), and the corresponding versions for parents are available for all three instruments (results not reported here). All of these versions could be useful in informing European policies by providing data on HRQoL and its distribution across Europe. By covering aspects of HRQoL which go beyond those included in existing instruments, the KIDSCREEN-52 measurement model, can also contribute to a better understanding of perceived health in children and adolescents and can help to identify at risk populations.

The authors would like to thank all the families who participated in the KIDSCREEN study as well as everyone who contributed to the project.

Source of financial support: The KIDSCREEN Project was funded by a grant from the European Commission (contract No. QLG-CT-2000-00751) within the EC 5th framework-Programme "Quality of Life and Management of Living Resources". There is no conflict of interest for any of the authors. There are no other persons who made substantial contributions to the work but do not met the criteria for Authorship.

References

1 McHorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. Ann Intern Med 1997;127:743–50.

2 Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. Arch Dis Child 2001;84:205–11.
3 Committee on Evaluation of Children's Health NRC. Measuring Children's Health. Children's Health the Nation's Wealth: Assessing and Improving Child Health. Washington: The National Academies Press, 2004.
4 Felder-Puig R, Frey E, Proksch K, et al. Validation of the German version of the Pediatric Quality of Life Inventory (PedsQL) in childhood cancer patients off treatment and children with epilepsy. Qual Life Res 2004;13:223–34.
5 Landgraf JM, Maunsell E, Speechley KN, et al. Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. Qual Life Res 1998; 7:433–45.
6 Bullinger M, Anderson R, Cella D, et al. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. In: Shumaker SA, Berzon RA, eds. The International Assessment of Health-related Quality of Life. New York: Oxford Rapid Communications, 1995.
7 The WHOQOL Group. The development of the World Health Organization Quality of Life Assessment Instrument (the WHOQoL). In: Orley J, Kuyken W, eds. Quality of Life Assessment: International Perspectives. Heidleberg: Springer-Verlag, 1994.
8 Skevington SM, Sartorius N, Amir M. Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL instruments. Soc Psychiatry Psychiatr Epidemiol 2004;39:1–8.
9 Ravens-Sieberer U, Gosch A, Abel T, et al. Quality of life in children and adolescents: a European public health perspective. Soc Prev Med 2001;46:297–302.
10 Ravens-Sieberer U, Erhart M, Power M, et al. Item-response-theory analyses of child and adolescent self-report quality of life data: the European cross-cultural research instrument KIDSCREEN. Qual Life Res 2003;12:1793.
11 Bullinger M, Schmidt S, Petersen C, the DISABKIDS Group. Assessing quality of life of children with chronic health conditions and disabilities: a European approach. Int J Rehabil Res 2002;25:197–206.
12 Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and health-related quality of life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.
13 Ramil L, Berra S, Erhart M, et al. Survey methods and representativeness of European national surveys of the KIDSCREEN study. BMC Public Health 2007;7: 182.
14 Herdman M, Rajmil L, Ravens-Sieberer U, et al. Expert consensus in the development of a European health-related quality of life measure for children and adolescents: a Delphi study. Acta Pediatr 2002;9: 1385–90.
15 Detmar SB, Bruil J, Ravens-Sieberer U, et al. A focus group approach in constructing the KIDSCREEN HRQOL questionnaire for children and adolescents. Qual Life Res 2006;15:1345–53.
16 Nosikov A, Gudex C. EUROHIS: Developing Common Instruments for Health Surveys. Biomed Health Res 2003:57.
17 WHOQOL Group. Study protocol for the World Health Organisation project to develop a quality of life assessment instrument (WHOQOL). Qual Life Res 1993;2:153–9.
18 Bullinger M, Alonso J, Apolone G, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998; 51:913–23.
19 Varni JW, Seid M, Kurtin PS. PedsQL 4.0. reliability and validity of the Pediatric Quality of Life Inventory version 4.0. generic core scales in healthy and patient populations. Medical Care 2001;39:800–12.
20 Upton P, Eiser C, Cheung I, et al. Measurement properties of the UK-English version of the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales. Health Qual Life Outcomes 2005;3: 1–7.
21 Starfield B, Riley AW, Green BF, et al. The adolescent CHIP: a population-based measure of health. Med Care 1995;33:553–66.
22 Edwards TC, Huebner CE, Connell FA, et al. Adolescent quality of life, part I. Conceptual and measurement model. J Adolesc Health 2002;25:275–86.
23 Patrick DL, Edwards TC, Topolski TD. Adolescent Quality of Life, Part II: initial validation of a new instrument. J Adolesc Health 2002;25:287–300.
24 Bethell CD, Read D, Stein RE, et al. Identifying children with special health care needs: development and evaluation of a short screening instrument. Ambul Pediatr 2002;2:38–48.
25 Bethell CD, Read D, Neff J, et al. Comparison of the children with special health care needs screener to the questionnaire for identifying children with chronic conditions-revised. Ambul Pediatr 2002;2:49–57.
26 Davidoff AJ. Identifying children with special health care needs in the National Health Interview Survey: a new resource for policy analysis. Health Serv Res 2004;39:53–71.
27 Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997;38:581–6.
28 Currie CE, Elton RA, Todd J, et al. Indicators of socioeconomic status for adolescents: the WHO Health Behaviour in School-aged Children Survey. Health Educ Res 1997;12:385–97.
29 Ravens-Sieberer U, Gosch A, Rajmil L, et al. The KIDSCREEN-52 Quality of life measure for children and adolescents: development and first results from a European survey. Exp Rev Pharmacoeconom Res 2005;5:353–64.
30 Cronbach LJ. Coefficient alpha and the internal structure of test. Psychometrika 1951;16:297–334.
31 Hays RD, Hayashi T, Carson S, et al. User's guide for the Multitrait Analysis Program (MAP). Santa Monica, CA: RAND, 1988.

32 Chinn S, Burney P. On measuring repeatability of data from self-administered questionnaires. Int J Epidemiol 1987;16:121–7.
33 Nunnally J, Bernstein IJ. Psychometric Theory (3rd ed.). New York: Mc-Graw-Hill, 1994.
34 Jöreskog KG Sörbom D, eds. PRELIS 2: Users's Reference Guide. Lincolnwood: Scientific Software International, Inc, 2002.
35 Jöreskog KG Sörbom D, eds. LISREL 8: Users's Reference Guide. Lincolnwood: Scientific Software International, Inc 2001.
36 McDonald RP Ho MHR. Principals and practice in reporting structural equation analyses. Psychol Methods 2002;7:64–82.
37 Browne MW Cuddeck R. Alternative ways of assessing model fit. In: Bollen KA Long JS, eds. Testing Structural Equation Models. Thousand Oaks, CA: Sage, 1993.
38 Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1999;107:238–46.
39 Hu L Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives Struct Eq Model 1999;6:1–55.
40 De Goede M Spruit E Iedema J, et al. How do vocational and relationship stressors and identity formation affect adolescent mental health? J Adolesc Health 1999;25:14–20.
41 Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
42 Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahwah, NJ: L. Erlbaum; 2001.
43 von Davier M. WINMIRA program description and recent enhancements. Methods Psychol Res Online 1997;2:25–8.
44 Rost J, Carstensen CH. Multidimensional Rasch measurement via item component models and faceted designs. Appl Psychol Meas 2002;26:42–56.
45 Zumbo BDA. Handbook on Theory and Methods of Differential Item Functioning (DIF). Ottawa: National Defense Headquarters, 1999.
46 Jodoin MG, Gierl MJ. Evaluating type I error and power using an effect size measure with the logistic regression procedure for DIF detection. Appl Meas Educ 2001;14:329–49.
47 Bortz J, Döring N. Forschungsmethoden und Evaluation (3rd ed.). Berlin: Springer, 2002.
48 Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum, 1988.
49 Campbell DT, Fiske DW. Convergent and discriminant validation by the Multitrait-Multimethod Matrix. Psychol Bull 1959;56:81–105.