

## Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire

Stephane Robitail · Ulrike Ravens-Sieberer · Marie-Claude Simeoni · Luis Rajmil · Jeanet Bruil · Mick Power · Wolfgang Duer · Bernhard Cloetta · Ladislav Czemy · Joanna Mazur · Agnes Czimbalmos · Yannis Tountas · Curt Hagquist · Jean Kilroe · Pascal Auquier · the KIDSCREEN Group

Accepted: 28 June 2007 / Published online: 1 August 2007  
© Springer Science+Business Media B.V. 2007

### Abstract

**Objectives** The aim of this study is to assess the structural and cross-cultural validity of the KIDSCREEN-27 questionnaire. **Methods** The 27-item version of the KIDSCREEN instrument was derived from a longer 52-item version and was administered to young people aged 8–18 years in 13 European countries in a cross-sectional survey. Structural and cross-cultural validity were tested using multitrait multi-item analysis, exploratory and confirmatory factor analysis, and Rasch analyses. Zumbo's logistic regression method was applied to assess differential item functioning (DIF) across countries. Reliability was assessed using Cronbach's alpha. **Results** Responses were obtained from  $n = 22,827$  respondents (response rate 68.9%). For the combined sample from all countries, exploratory factor analysis with pro-

crustean rotations revealed a five-factor structure which explained 56.9% of the variance. Confirmatory factor analysis indicated an acceptable model fit (RMSEA = 0.068, CFI = 0.960). The unidimensionality of all dimensions was confirmed (INFIT: 0.81–1.15). Differential item functioning (DIF) results across the 13 countries showed that 5 items presented uniform DIF whereas 10 displayed non-uniform DIF. Reliability was acceptable (Cronbach's  $\alpha = 0.78$ –0.84 for individual dimensions).

**Conclusions** There was substantial evidence for the cross-cultural equivalence of the KIDSCREEN-27 across the countries studied and the factor structure was highly replicable in individual countries. Further research is needed to correct scores based on DIF results. The KIDSCREEN-27 is a new short and promising tool for use in clinical and epidemiological studies.

Members of the KIDSCREEN group are: Austria: Wolfgang Duer, Kristina Fuerth; Czech Republic: Ladislav Czemy; France: Pascal Auquier, Marie-Claude Simeoni, Stephane Robitail, Germany: Ulrike Ravens-Sieberer (international coordinator in chief), 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, Eimear Flannery; The Netherlands: Jeanet Bruil, Symone Detmar, Eric Veripps; Poland: Joanna Mazur, Ewa Mierzejewska; 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.

L. Rajmil  
Agency for Quality, Research and Assessment in Health (AQuRAHealth), Barcelona, Spain

J. Bruil  
Child Health Unit, Prevention and Physical Activity, TNO Quality of Life, Leiden, The Netherlands

M. Power  
Department of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, Scotland

W. Duer  
Ludwig Boltzmann-Institute for Sociology of Health and Medicine, University of Vienna, Vienna, Austria

B. Cloetta  
Unit for Health Research, Institute for Social and Preventive Medicine, University of Bern, Bern, Switzerland

L. Czemy  
Prague Psychiatric Center, Prague, Czech Republic

S. Robitail · M.-C. Simeoni · P. Auquier (✉)  
EA 3279, School of Medicine, Perceived Health Research Unit, Marseille, France  
e-mail: pascal.auquier@medecine.univ-mrs.fr

U. Ravens-Sieberer  
School of Public Health, WHO Collaborating Center for Child and Adolescent Health Promotion, University of Bielefeld, Bielefeld, Germany

**Keywords** Cross-cultural equivalence · Health-related Quality of Life · Item response theory · Pediatric · Questionnaire

## Introduction

The assessment of health-related quality of life (HRQOL) plays an increasingly important role in the assessment of population health and health care services in Europe [1] and has received increasing attention in pediatrics and adolescent care. Several instruments for children and adolescents have been developed in different European countries, including Germany [2], the United Kingdom [3], the Netherlands [4], and, France [5]. Some of these instruments have been adapted and validated for use in other European languages [6, 7].

Since these instruments were developed, however, standards in the field of HRQOL measurement have evolved. First, there has been increasing consensus regarding the concept of HRQOL, which is most commonly defined as a multidimensional concept covering the social, physical and psychological domains of health [8, 9]. Second, it is generally accepted that instrument content should be largely derived from the population whose QOL is to be measured and that it should reflect their interests and concerns. This focus has been increasingly incorporated into the assessment of QOL in pediatric population in the past 10 years [5]. Third, psychometric and other methods used in instrument development and validation have become more sophisticated [10], and the importance of determining an instrument's suitability for use in cross-cultural settings has been recognized [11].

Given that none of the existing instruments to measure HRQOL in children and adolescents have been developed cross-culturally, it was considered important to construct a measure that would reflect the concerns and interests of children and adolescents from different European countries, and which would take into account possible cultural differences between countries. The KIDSCREEN project

was funded by the European Commission within the Fifth Framework program [12], to develop an instrument to assess HRQOL in children and adolescents using input from children in different European countries.

The first instrument developed from the KIDSCREEN project was a 52-item questionnaire (KIDSCREEN-52) covering 10 dimensions of HRQOL and which has been shown to have good psychometric properties [13]. Two shorter versions of the instrument (the KIDSCREEN-27 questionnaire and the KIDSCREEN-10 Index) were developed during later stages of the project in order to provide instruments that might be useful in clinical practice and/or as screening tools. Given the nature and objectives of the KIDSCREEN project, it was considered important to examine the structural validity of the measures both globally and by individual country, and to establish their validity for use in cross-cultural research.

The objectives of this analysis were to examine the structural and cross-cultural validity of the KIDSCREEN-27 self-reported questionnaire by analyzing its factor structure, the unidimensionality and internal consistency of its individual dimensions, as well as differential item functioning (DIF) across countries. All of these analyses were performed for the sample as a whole and by individual country. The results of analyzing the measure's external validity are reported in a companion paper to the current article [13].

## Population and methods

### Development of the KIDSCREEN-27

The KIDSCREEN-27 was derived from the longer, 52-item version of the instrument using two approaches.

Item reduction was carried out using exploratory factor analysis, Mokken scale analysis [14–16], Rasch partial credit modeling [17, 18], DIF (using Zumbo's logistic regression method [19], MAP analyses [20], and confirmatory factor analysis [21]. Item reduction was an

---

J. Mazur  
Department of Epidemiology, National Research Institute of Mother & Child, Warsaw, Poland

A. Czimbalmos  
Health Promotion and Development, Centre Child Health Department, Budapest, Hungary

Y. Tountas  
Centre for Health Services Research and Hellenic Society for Health Promotion and Education, Athens, Greece

C. Hagquist  
Karlstad University, Karlstad, Sweden

J. Kilroe  
HSE, Programme of Action for Children, Dublin, Ireland

P. Auquier  
Service de Santé Publique. Faculté de Médecine, 27, bd Jean Moulin, Marseille 13385, France

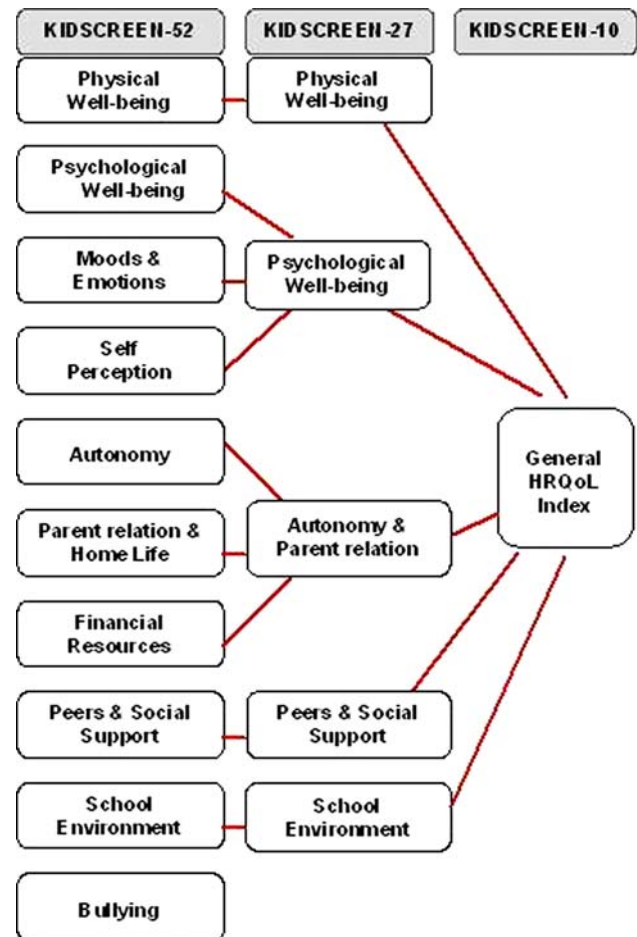
P. Auquier  
Department of Public Health, Timone University Hospital, Marseilles, France

iterative process and at each step item reduction was carried out in 50% of the sample chosen at random and the results were then confirmed in the other half of the sample. A similar process was used in parallel to develop an even shorter, index version of the KIDSCREEN. As the idea of the project team was that the 27-item version should include the index in its entirety, a further stage of the development process was to combine the two versions. At the end of the reduction process there was an expert review of the whole process in which there was opportunity to decide on whether additional items were necessary.

The resulting KIDSCREEN-27 questionnaire consisted of 27 items which are used to assess HRQoL across 5 dimensions: Physical Well-Being (5 items) explored the level of the child's/adolescent's physical activity, energy and fitness; Psychological Well-Being (7 items) included items on positive emotions, satisfaction with life, and feeling emotionally balanced; Parent Relations & Autonomy (7 items) examined relationships with parents, the atmosphere at home, and feelings of having enough age-appropriate freedom, as well as degree of satisfaction with financial resources; Social Support & Peers (4 items) examined the nature of the respondent's relationships with other children/adolescents, and; School Environment (4 items) explored the child's/adolescent's perceptions of his/her cognitive capacity, learning and concentration, and their feelings about school. The items assess either the frequency of behavior/feelings or, in fewer cases, the intensity of an attitude and are answered on a five-point scale with a timeframe of one week. Scores were computed for each dimension using the person estimation based on the Rasch model. Figure 1 shows the relationship between the different versions of the KIDSCREEN questionnaires.

In order to test the structural and cross-cultural validity of the 27-item version, results were used from a large cross-sectional, observational study in 13 European countries: Austria (AT), Czech Republic (CZ), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IE), Poland (PL), Spain (ES), Sweden (SE), Switzerland (CH), the Netherlands (NL), and the United Kingdom (UK). The target population for this study was children and adolescents aged 8–18, and the sample was designed to be representative by age, sex, and region.

In this study, three approaches to sample selection and administration were followed: (1) telephone sampling followed by mail survey (AT, CH, DE, ES, FR, and NL), (2) school sampling and administration (EL, HU, IE, and SE), or school sampling and mail administration (PL), and (3) multistage random sampling of communities and households (CZ). In the UK, a combination of telephone and school sampling methods was used.



**Fig. 1** Relationships between the different versions of the KIDSCREEN questionnaires.

Fieldwork was carried out between May and September 2003 except in IE, where data was collected in 2005. Some data was collected on those who refused to participate. All procedures were carried out following the data protection requirements of the European Parliament (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data). Each country was asked to respect ethical and legal requirements in their country for this type of survey and to obtain signed informed consent from participants. A more detailed description of the KIDSCREEN sampling methods is provided elsewhere, together with a detailed analysis on sample representativeness based on Eurostat data [22].

#### Statistical analyses

The statistical analyses aimed to check the internal and cross-cultural validity of the measurement model of the

European KIDSCREEN-27 questionnaire across countries by testing its multi-dimensional structure, the uni-dimensionality of its five dimensions, its internal consistency, and the existence of any cross-cultural differences in item functioning. Several methods were used. As a first step, missing values and floor and ceiling effects were calculated by dimension. Missing values under 5%, and floor and ceiling effects lower than 15% were considered as acceptable [30]. Exploratory factor analyses using principal component analyses (PCA) were performed on the sample as a whole and in each country to test the generalizability of the hypothesized five-dimensional structure of the KIDSCREEN-27 instrument. Components were rotated orthogonally using the varimax rotation method and the invariance of the factorial structure across countries was assessed using EFA with procrustean rotations [21, 23]. A factor loading higher than 0.40 was considered as relevant. Confirmatory factor analysis (CFA) of the five factor structure obtained with EFA on international data was performed using the LISREL model [24]. Since the data were ordinal and non-normally distributed, polychoric correlation matrices were computed in addition to the usual maximum likelihood (ML) CFA [24]. The fit of international and national data to the model was analyzed using global indices (chi-squared and root mean square error of approximation [RMSEA] and incremental indices (Normed Fit Index [NFI] and comparative fit index [CFI]) [25]. NFI and CFI values over 0.90 were considered to represent an adequate fit; a RMSEA under 0.05 indicated a good fit and a RMSEA between 0.05 and 0.08, a reasonable fit [21, 26]. When the national sample size exceeded 1000, the weighted least square estimation (WLS) method using polychoric coefficients and the corresponding asymptotic covariance matrix was performed to corroborate the results [27]. Multisample confirmatory factor analyses were also conducted across the countries to evaluate the invariance of the factor structure with regard to RMSEA, NFI, and CFI [28].

Item-internal consistency (IIC) was assessed by correlating each item with its corresponding scale after correcting for overlap. IIC is considered satisfactory if 90% of the possible item-scale correlations are greater than 0.40 after correction for overlap [20, 29, 30]. Item discriminant validity (IDV) was assessed by determining the extent to which items correlated more highly with the dimensions they were hypothesized to represent than with other dimensions. It was measured as the percentage of item-scale correlations in which the correlation between an item and its hypothesized scale was at least 2 standard errors greater than the correlations between the item and other scales. The IDV standard is met if this condition is fulfilled in at least 80% of cases [20, 29, 30]. Internal consistency reliability was determined by computing

Cronbach's alpha coefficients [31] for all dimensions using combined data and for national sub-samples. A reliability of at least 0.70 is recommended to compare groups of patients [32].

The uni-dimensionality of the five pre-identified constructs was assessed using Rasch analyses with the Partial Credit Model (PCM) and unconditional maximum likelihood estimation method [17, 18, 29, 33]. The information mean square fit statistic (INFIT) was used to evaluate the fit to the model. An acceptable range for the INFIT was considered to be 0.7–1.3. Items with lower INFIT were considered redundant and those with an INFIT over 1.3 as not measuring the same dimension [17]. Analyses were performed on the whole sample and by country.

Differential item functioning (DIF) analyses based on IRT modeling were used to determine whether items behaved in the same way in the different countries. Zumbo's ordinal logistic regression [19] was used to test uniform and non-uniform DIF. Uniform DIF exists when the probability of giving a certain answer given a certain level of health is greater in one group (country) than the other, uniformly over all levels of health, i.e. there is no interaction between level of health and group membership. Non-uniform DIF exists when the probability of giving a particular answer at a given level of health varies both by country and by levels of health, i.e. interaction exists between level of health and group membership [34]. For the present study, a cut-off point of 2% for DIF was used [35]. The magnitude of DIF was assessed by using the pseudo- $R^2$  difference ( $\Delta-R^2$ ), which measures the increase in explained item variance when the variable country is included in the model. DIF was assessed in 78 head to head comparisons across the 13 countries.

Analyses were performed using WINSTEPS 3.42 [36], LISREL 8.52 [37], PRELIS software [38], MAP [20] and SPSS software 10.1.3 (SPSS, Chicago, IL, USA).

## Results

### Sample characteristics

The final sample consisted of 22,827 children and adolescents. The overall response rate was (68.9%). Sample characteristics overall and by country are shown in Table 1.

Missing values ranged from 1.6% (0.8–2.9% by country) for Peers and Social Support to 3.8% (0.9–9.0%) for Autonomy and Parents. Floor effects ranged from 0.0% (0.0–0.1%) for Psychological Well-being to 0.4% (0.0–0.9%) for Autonomy and Parents; and ceiling effects ranged from 5.1% (2.7–9.2%) for Physical Well-being to 15.0% (6.8–22.8%) for Peers and Social Support.

**Table 1** Characteristics of the KIDSCREEN sample

Country <sup>a</sup>	Total	AT	CH	CZ	DE	ES	FR	GR	HU	IE	NL	PL	SE	UK
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
N (valid cases)	22,827	1,475	1,701	1,592	1,723	876	1,049	1,174	3,237	1,240	1,885	1,715	3,283	1,877
<i>Children</i>														
Age range: 8–11														
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)
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
<i>Adolescents</i>														
Age range: 12–18														
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)
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

<sup>a</sup> Countries: AT = Austria; CH = Switzerland; CZ = Czech Republic; DE = Germany; ES = Spain; FR = France; GR = Greece; HU = Hungary; IE = Ireland; NL = Netherlands; PL = Poland; SE = Sweden; UK = United Kingdom

## Exploratory factor analysis

Results of EFA with procrustean rotations on international data and across the countries are presented in the Table 2. Using the combined data from the whole sample, a five-factor solution explained the greatest amount of variance (56.9%). At country level, the same five-factor solution explained between 51.8% (Switzerland) and 58.6% (Sweden) of the variance. Using combined data, one item (Have you had fun?) loaded similarly on the Psychological Well Being dimension and the Peers and Social Support dimension. One other item (Have you had enough time for yourself?) showed similar item loadings on the Autonomy & Parents and Psychological Well Being dimensions.

In general, the factor structure was replicated across all the participating countries with few exceptions. For example the item “had fun?” showed similar item loadings on the dimensions Psychological Well-being and Peers and social support in 9 out of the 13 countries, and “Have you had enough time for yourself?” loaded equally on the dimensions of Psychological Well-being and Autonomy & Parents in 10 out of the 13 countries.

## Confirmatory factor analysis

Confirmatory factor analysis showed a reasonable fit (RMSEA = 0.069) for the five-factor solution using combined data from all countries (Table 3). At the country level, the lowest fit for this model was for Greece (RMSEA = 0.082) and Poland (RMSEA = 0.080), while the best fit was observed using the Austrian data (RMSEA = 0.059). Results based on WLS estimation were consistent with those based on the ML method (data not shown). Multiple sample CFA results indicated acceptable factor invariance across the countries, with RMSEA at the acceptable upper threshold (RMSEA = 0.084) (Fig. 2).

## Psychometric properties

Overall, IIC and IDV were very satisfactory, with 100% scaling success using combined data and over 80% for all dimensions in all countries. The lowest level of scaling success was seen for one item in the Physical Well-being dimension (How would you say your health is?) and one item in the Autonomy and Parents dimension (Have you had enough time for yourself?) (Table 4). Cronbach’s alpha coefficients were over 0.78 for all dimensions using combined data (Table 4).

## Uni-dimensionality

Rasch analysis showed all five dimensions of the KIDSCREEN-27 to be uni-dimensional. All INFIT statistics

**Table 2** Exploratory factor analyses on international data & procrustean rotations by country: factor loadings

	Psy. WB	Phy. WB	Peers & SS	Au. & Parents	School
In general, how would you say your health is?	0.22 [0.08,0.59]	<b>0.56</b> [0.41,0.67]	0.06 [-0.01,0.17]	0.14 [0.02,0.21]	0.13 [0.04,0.22]
felt fit and well?	0.27 [0.14,0.37]	<b>0.71</b> [0.62,0.77]	0.08 [0.01,0.14]	0.10 [0.03,0.14]	0.17 [0.07,0.27]
been physically active (e.g. running, climbing, biking)?	0.06 [-0.11,0.2]	<b>0.74</b> [0.56,0.77]	0.13 [0.07,0.21]	0.08 [-0.07,0.13]	0.06 [-0.03,0.12]
been able to run well?	0.07 [0.02,0.21]	<b>0.75</b> [0.66,0.77]	0.09 [-0.01,0.16]	0.11 [0.02,0.2]	0.10 [0.02,0.13]
felt full of energy?	0.27 [0.15,0.35]	<b>0.67</b> [0.54,0.73]	0.17 [0.13,0.24]	0.10 [0.03,0.16]	0.19 [0.08,0.29]
your life been enjoyable?	<b>0.55</b> [0.42,0.61]	0.33 [0.24,0.36]	0.22 [0.12,0.31]	0.17 [0.01,0.24]	0.27 [0.2,0.33]
been in a good mood?	<b>0.50</b> [0.34,0.54]	0.31 [0.21,0.37]	0.29 [0.2,0.36]	0.11 [-0.04,0.16]	0.25 [0.16,0.31]
had fun?	<b>0.43</b> [0.28,0.54]	0.31 [0.22,0.35]	<b>0.43</b> [0.36,0.52]	0.12 [-0.06,0.16]	0.18 [0.05,0.23]
felt sad?	<b>0.74</b> [0.46,0.81]	0.15 [0.07,0.3]	0.07 [-0.04,0.32]	0.07 [-0.07,0.29]	0.08 [-0.02,0.38]
felt so bad that you didn't want to do anything?	<b>0.66</b> [0.29,0.8]	0.16 [0.07,0.39]	0.04 [-0.11,0.26]	0.10 [-0.04,0.28]	0.15 [0.03,0.4]
felt lonely?	<b>0.68</b> [0.45,0.76]	0.09 [-0.01,0.25]	0.19 [0.13,0.38]	0.15 [0.04,0.34]	0.06 [-0.05,0.31]
been happy with the way you are?	<b>0.48</b> [0.4,0.59]	0.32 [0.2,0.37]	0.17 [0.09,0.23]	0.23 [0.16,0.29]	0.31 [0.22,0.37]
had enough time for yourself?	0.36 [0.23,0.51]	0.18 [-0.06,0.29]	0.17 [0.08,0.31]	0.36 [0.22,0.47]	0.18 [-0.25,0.39]
been able to do the things that you want to do in your free time?	0.34 [0.26,0.49]	0.18 [0.02,0.25]	0.22 [0.17,0.3]	<b>0.44</b> [0.28,0.54]	0.12 [-0.2,0.31]
your parent(s) had enough time for you?	0.34 [0.21,0.58]	0.07 [-0.09,0.15]	0.14 [0,0.19]	<b>0.58</b> [0.39,0.64]	0.22 [0.11,0.3]
your parent(s) treated you fairly?	0.33 [0.22,0.59]	0.03 [-0.08,0.09]	0.09 [-0.06,0.16]	<b>0.54</b> [0.34,0.62]	0.30 [0.18,0.43]
been able to talk to your parent(s) when you wanted to?	0.37 [0.23,0.6]	0.04 [-0.11,0.16]	0.11 [-0.03,0.19]	<b>0.55</b> [0.32,0.63]	0.24 [0.14,0.34]
had enough money to do the same things as your friends?	-0.03 [-0.35,0.14]	0.15 [0.11,0.28]	0.14 [0.02,0.43]	<b>0.78</b> [0.48,0.87]	0.05 [-0.14,0.42]
had enough money for your expenses?	-0.01 [-0.34,0.12]	0.16 [0.13,0.3]	0.10 [-0.04,0.39]	<b>0.79</b> [0.48,0.86]	0.11 [-0.06,0.47]
spent time with your friends?	0.15 [0.12,0.32]	0.18 [0.11,0.23]	<b>0.73</b> [0.62,0.78]	0.08 [-0.02,0.25]	0.01 [-0.15,0.04]
had fun with your friends?	0.21 [0.18,0.32]	0.13 [0.08,0.24]	<b>0.78</b> [0.7,0.81]	0.14 [0.03,0.18]	0.09 [-0.08,0.12]
you and your friends helped each other?	0.04 [-0.02,0.19]	0.09 [0.04,0.17]	<b>0.78</b> [0.63,0.82]	0.14 [0.04,0.19]	0.16 [0.04,0.21]
been able to rely on your friends?	0.14 [0,0.24]	0.05 [-0.02,0.14]	<b>0.70</b> [0.59,0.74]	0.17 [0.09,0.21]	0.15 [-0.03,0.18]
been happy at school?	0.23 [0.09,0.34]	0.16 [0.09,0.3]	0.20 [0.09,0.25]	0.11 [-0.01,0.16]	<b>0.71</b> [0.5,0.79]
got on well at school?	0.18 [0.07,0.3]	0.18 [0.09,0.33]	0.12 [-0.09,0.23]	0.15 [0.11,0.23]	<b>0.73</b> [0.63,0.82]
been able to pay attention?	0.17 [0.12,0.25]	0.19 [0.12,0.27]	0.07 [-0.18,0.1]	0.22 [0.15,0.26]	<b>0.70</b> [0.62,0.72]
got along well with your teachers?	0.07 [0.05,0.19]	0.09 [-0.04,0.15]	0.09 [-0.11,0.14]	0.17 [0.1,0.26]	<b>0.77</b> [0.58,0.8]

[Range of factor loading across countries]; In bold: factor loadings &gt;0.40

Psy. WB: Psychological Well Being; Phy. WB: Physical Well Being; Peers &amp; SS: Peers and Social Support; Au. &amp; Parents: Autonomy and Parents; School: School Environment

**Table 3** Confirmatory factor analyses on whole sample and across countries (df = 314)

Unisample	$\chi^2$	df	NFI	CFI	RMSEA [95% confidence interval]
Whole	31001.5	314	0.958	0.959	0.069 [0.068,0.070]
<i>By countries</i>					
Germany	1941.0	314	0.957	0.964	0.058 [0.055,0.060]
Spain	1406.6	314	0.947	0.958	0.068 [0.064,0.071]
Netherlands	2780.6	314	0.937	0.944	0.068 [0.065,0.070]
Austria	1680.0	314	0.952	0.961	0.059 [0.056,0.062]
UK	3017.1	314	0.949	0.954	0.073 [0.070,0.075]
France	1704.9	314	0.935	0.946	0.072 [0.068,0.075]
Switzerland	2338.6	314	0.938	0.947	0.066 [0.063,0.068]
Hungary	5334.7	314	0.950	0.953	0.072 [0.071,0.074]
Greece	2652.3	314	0.914	0.924	0.082 [0.079,0.085]
Czech Republic	2956.7	314	0.940	0.947	0.074 [0.072,0.077]
Ireland	2566.6	314	0.931	0.939	0.079 [0.076,0.082]
Poland	3522.0	314	0.937	0.942	0.080 [0.078,0.082]
Sweden	5261.7	314	0.96	0.963	0.074 [0.072,0.076]
Multisample	$\chi^2$	df	NFI	CFI	RMSEA
By countries	52622.3	4850	0.92	0.927	0.084 [0.084,0.085]

$\chi^2$ : chi-square value; df: degrees of freedom for  $\chi^2$ ; NFI: Normed Fit Index; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation

were between 0.70 and 1.30 except for one item (In general, how would you say your health is?), which did not present a satisfactory fit for the PCM model (IN-FIT = 1.33) in the Czech Republic sample (Table 5).

#### Cross-cultural item functioning

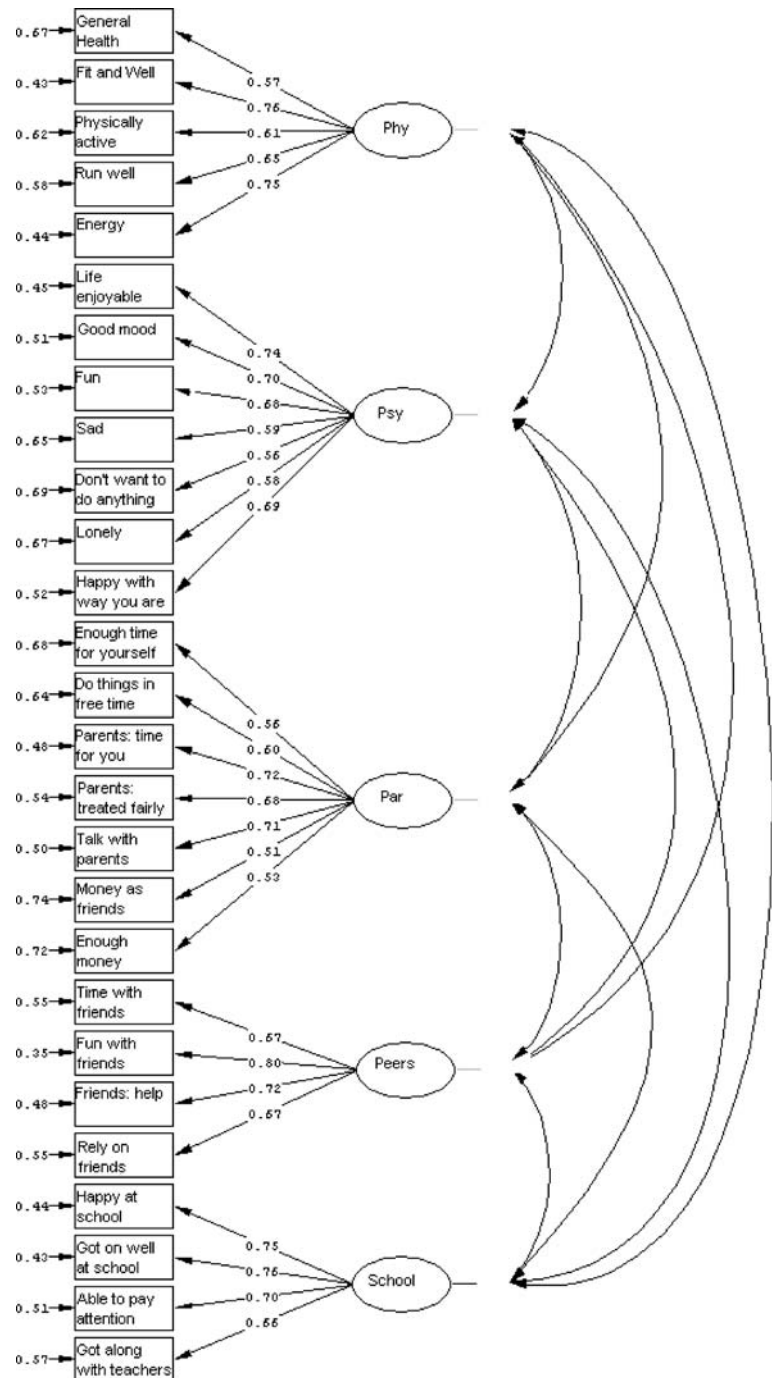
Results of DIF analyses were satisfactory considering the number of the countries and the sample sizes (Table 5). Across the 13 countries, uniform DIF was observed for only 5 items whereas non-uniform DIF was found for 10 items, mostly in the dimensions Psychological Well Being and Parents and Autonomy. Considering these 10 items, 78 head to head comparisons between the 13 countries have led to different results considering each couple of countries (e.g. Spain vs. Germany). No DIF was detected for 15 head to head comparisons. One or two items presented DIF for 24 comparisons whereas three or four items displayed DIF for 26 comparisons and five or six for 12 head to head comparisons. Only one pair of countries presented 7 out of the 10 items showing DIF.

#### Discussion

As far as we are aware, the KIDSCREEN project is the first to develop and validate an HRQOL instrument simultaneously in several different countries. As the first instrument (the KIDSCREEN-52) produced during the project included a total of 52 items [39], it was felt necessary to

produce shorter versions which would be useful in studies or situations in which the number of items that could be administered was limited, for example, in large-scale population health surveys, or for use in clinical practice. The study group decided it would be useful to have at least two shorter versions, one—an intermediate length version—which would provide a profile of children's HRQOL, and another—the 10-item Index—which would provide a single global score. The 27-item version was intended to be the profile measure, though it was also decided that it should incorporate all of the items from the KIDSCREEN-10 Index. Given the multinational nature of the KIDSCREEN project and its fundamental desire to produce a cross-culturally valid and useful instrument, several validation strategies were considered necessary to determine how far the project's aims had been achieved. As well as describing the process of item reduction, in the present paper, we have also examined the instrument's structural and cross-cultural integrity, as well as some psychometric characteristics such as ceiling and floor effects and internal consistency. The companion paper looks at other aspects of validity, including criterion and construct validity. Differences in response rates could affect study results. However, 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. Likewise, the large sample size and the spread of characteristics of children and adolescents in each of the countries

**Fig. 2** Confirmatory Factor Analysis Path Diagram (Whole sample RMSEA = 0.069). Phy: Physical Well Being; Psy: Psychological Well Being; Par: Autonomy and Parents; Peers: Peers and Social Support; School: School Environment



guarantees a diversity of responses which are useful for validity studies such as this.

Item reduction was carried out using a mixed approach which combined techniques and methods of classical psychometric and factor analysis together with newer techniques for item reduction from Item Response Theory. Although almost all items were selected on the basis of results from factor analysis and IRT, expert opinion regarding which items should be

included was also taken into account at various stages of questionnaire development. For example, the decision to include the item regarding overall health (In general, how would you say your health is?) was based on expert opinion, even though psychometrically the item did not produce optimal results. This decision was based on the fact that similar items are widely used in health-care surveys, and have demonstrated their usefulness as a reference item.



**Table 4** Psychometric properties for the whole sample and by country: Cronbach's Alpha, and multi-item multi-trait analysis

	Phy. WB	Psy. WB	Au. & Parents	Peers & SS	School
Alpha	0.78 [0.75–0.86]	0.84 [0.79–0.88]	0.81 [0.74–0.82]	0.81 [0.76–0.84]	0.81 [0.76–0.82]
<i>IIC/IDV</i>					
Global	100/100	100/100	100/100	100/100	100/100
Germany	80/100	100/100	100/96.4	100/100	100/100
Spain	100/95	100/100	100/96	100/100	100/100
Netherlands	100/100	100/96	100/96	100/100	100/100
Austria	80/100	100/100	100/96	100/100	100/100
United Kingdom	100/100	100/96	100/100	100/100	100/100
France	100/95	100/100	100/96	100/100	100/100
Switzerland	80/100	100/100	100/93	100/100	100/100
Hungary	100/100	100/100	100/96	100/100	100/100
Greece	100/100	100/93	86/86	100/100	100/94
Czechoslovakia	100/100	100/96	100/96	100/100	100/100
Ireland	100/100	100/100	100/96	100/100	100/100
Poland	100/95	100/100	100/96	100/100	100/100
Sweden	100/80	100/100	100/96	100/100	100/94

Alpha: Cronbach's alpha on whole sample [range across countries]; IIC/IDV: item internal consistency/item discriminant validity expressed in terms of range of Item-scale correlation (% Scaling success); Phy. WB: Physical Well Being; Psy. WB: Psychological Well Being; Au. & Parents: Autonomy and Parents; Peers & SS: Peers and Social Support; School: School Environment

In terms of structural validity, we observed that the 5-factor structure of the KISDCREEN-27 was confirmed at the international as well as at the national level, and the uni-dimensionality of each scale was demonstrated. EFA was also carried out in all of the individual countries, rather than having a model imposed from above. The amount of variance explained (56.9% using the entire, combined data set) was also acceptable and in line with the amount of variance explained by other, similar instruments, such as the Child Health Questionnaire [40] and the Pediatrics Quality of Life Inventory [41].

In the present study, EFA was performed using principal components analysis, which assumes that the different dimensions are totally explained by a linear combination of items. It has been argued that common factor analysis, which includes an error term in its model, should be preferred in psychometrics [32]. However, as pointed out by Gorsuch, in large samples these two methods lead to the same results [42]. Given the large sample size in the present study, results should be quite similar between the two methods.

Confirmatory factor analysis supported the results found with EFA and both have demonstrated that the instrument conforms to conceptual and theoretical considerations regarding what should be measured in HRQOL, i.e. a broad range of physical, psychological and social aspects of quality of life. Items loadings were generally satisfactory and the multidimensional structure has been demonstrated using confirmatory factor analysis. Significant CFA chi-

square were also observed which might indicate a misfit of the 5-dimensional structure across the countries. However, this is more likely to be due to the fact that the CFA chi-square statistic alone is not appropriate in large data sets, because of its statistical power under those conditions, whereby any discrepancy, however trivial, between the data and the model leads to a significant result [43].

One of the strengths of the present study is that we used a combination of methods to test the instrument's 'internal' psychometric properties. Psychometric analyses such as item-scale correlations, and analysis of floor and ceiling effects and internal consistency, suggest that the instrument is psychometrically very sound across all countries, and easily meets accepted standards. Floor and ceiling effects were below the recommended cut point of 15% [44], and internal consistency was also acceptable and well above the usually accepted cut point of 0.70 [45].

Finally, cross-cultural validity results across the 13 countries showed that 5 items displayed uniform DIF whereas 10 showed non-uniform DIF. When items did not display DIF between countries, the scores are comparable without any bias due to the measurement instrument. When DIF is detected for an item, calibrating item parameters using IRT analysis separately in appropriate countries based on the head to head results allows us to calculate person Rasch scores adjusted for DIF [46]. By comparing the unadjusted and adjusted Rasch scores for each item it is possible to assess the impact of DIF on the QoL measure. Several methods have been already tested for correcting for

**Table 5** Rasch analyses: FIT statistics and DIF analyses

	FIT results INFIT	DIF results across 13 countries (%)	
		$\Delta$ -R <sup>2</sup> (U)	$\Delta$ -R <sup>2</sup> (NU)
In general, how would you say your health is? felt fit and well?	1.13 [0.97,1.33]	<b>2.84</b>	<b>4.00</b>
been physically active (e.g. running, climbing, biking)?	0.89 [0.83,0.97]	0.29	0.75
been able to run well?	1.06 [0.95,1.27]	1.06	1.39
felt full of energy?	0.96 [0.84,1.06]	0.43	1.06
your life been enjoyable?	0.95 [0.79,1.14]	0.53	1.51
been in a good mood?	0.86 [0.81,0.95]	1.03	1.58
had fun?	0.93 [0.8,1.02]	1.48	<b>2.70</b>
felt sad?	1.01 [0.88,1.14]	1.79	<b>3.09</b>
felt so bad that you didn't want to do anything?	0.98 [0.88,1.02]	<b>2.36</b>	<b>3.57</b>
felt lonely?	1.09 [0.98,1.29]	<b>2.03</b>	<b>2.96</b>
been happy with the way you are?	1.1 [0.92,1.25]	1.02	1.61
had enough time for yourself?	1.02 [0.94,1.16]	1.94	<b>2.20</b>
been able to do the things that you want to do in your free time?	1.13 [1.05,1.21]	1.69	<b>2.01</b>
your parent(s) had enough time for you?	1.03 [0.94,1.1]	1.40	1.63
your parent(s) treated you fairly?	0.89 [0.85,0.95]	1.28	1.48
been able to talk to your parent(s) when you wanted to?	0.93 [0.86,1]	1.41	1.70
had enough money to do the same things as your friends?	0.9 [0.83,0.99]	1.28	1.37
had enough money for your expenses?	1.09 [1.01,1.19]	1.69	<b>2.13</b>
spent time with your friends?	1.02 [0.98,1.13]	<b>2.19</b>	<b>2.92</b>
had fun with your friends?	1.15 [0.96,1.2]	<b>2.79</b>	<b>2.94</b>
and your friends helped each other?	0.81 [0.74,0.89]	1.50	1.79
been able to rely on your friends?	0.92 [0.86,0.99]	1.09	1.32
been happy at school?	1.08 [1.02,1.22]	0.78	1.08
got on well at school?	0.97 [0.87,1.18]	0.30	0.78
been able to pay attention?	0.91 [0.82,0.98]	1.05	1.49
got along well with your teachers?	1.04 [0.85,1.18]	0.95	1.67
	1.06 [0.96,1.18]	0.92	1.10

FIT statistics: INFIT on whole sample [range by countries]; DIF Zumbo's ordinal logistic regression analyses results on whole sample across the thirteen countries (U: uniform; NU: non uniform), in bold: values >2%

DIF on the KIDSCREEN-27 and have shown promising results [47].

In conclusion, the KIDSCREEN-27 appears to be psychometrically robust instrument which would be suitable for use in national or international clinical and epidemiological studies.

**Acknowledgements** Source of support: The KIDSCREEN project was financed by a grant from the European Commission (QLG-CT-2000-00751) within the EC 5th Framework-Programme "Quality of Life and Management of Living Resources".

## References

- World Health Organization Regional Office for Europe (1996). *The Ljubljana charter on reforming health care*. Paper presented at the Ljubljana Conference, Ljubljana.
- Ravens-Sieberer, U., & Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: First psychometric and content analytical results. *Quality of Life Research*, 7, 399–407.
- Eiser, C., Havermans, T., Craft, A., & Kernahan, J. (1995). Development of a measure to assess the perceived illness experience after treatment for cancer. *Archives of Disease in Childhood*, 72, 302–307.
- Vogels, T., Verrips, G. H. W., Verloove-Vanhorick, S. P., et al. (1998). Measuring health-related quality of life in children: the development of the TACQOL parent form. *Quality of Life Research*, 7, 457–465.
- Simeoni, M. C., Auquier, P., Antoniotti, S., Sapin, C., & San Marco, J. L. (2000). Validation of a French health-related quality of life instrument for adolescents: The VSP-A. *Quality of Life Research*, 9, 393–403.
- Rajmil, L., Serra-Sutton, V., Fernandez-Lopez, J. A., et al. (2004). [The Spanish version of the German health-related quality of life questionnaire for children and adolescents: The Kindl]. *Anales de Pediatría (Barcelona)*, 60, 514–521.
- Serra-Sutton, V., Herdman, M., Rajmil, L., et al. (2002). [Cross-cultural adaptation to Spanish of the Vecu et Sante Percue de l'Adolescent (VSP-A): A generic measure of the quality of life of adolescents]. *Revista Española de Salud Pública*, 76, 701–712.

8. Rajmil, L., Herdman, M., De Sanmamed, M. J. F., et al. (2004). Generic health-related quality of life instruments in children and adolescents: A qualitative analysis of content. *Journal of Adolescent Health, 34*, 37–45.
9. Patrick, D. L., & Chiang, Y. P. (2000). Measurement of health outcomes in treatment effectiveness evaluations: Conceptual and methodological challenges. *Medical Care, 38*, II14–25.
10. Schmidt, S., & Bullinger, M. (2003). Current issues in cross-cultural quality of life instrument development. *Archives of Physical Medicine and Rehabilitation, 84*, S29–S34.
11. Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. *Quality of Life Research, 7*, 323–335.
12. Ravens-Sieberer, U., Gosch, A., Abel, T., et al. (2001). Quality of life in children and adolescents: A European public health perspective. *Sozial-Und Praventivmedizin, 46*, 294–302.
13. Ravens-Sieberer, U., Auquier, P., Erhart, M., et al. The KIDSCREEN-27 quality of life measure for children and adolescents – psychometric results from a cross-cultural survey in 13 European countries. *Quality of Life Research*. doi: [10.1007/s11136-007-9240-2](https://doi.org/10.1007/s11136-007-9240-2).
14. Mokken, R. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton, Berlin De Gruyter.
15. Mokken, R., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6*, 417–430.
16. Molenaar, I. (1982). Mokken slaing revisited. *Kwantitatieve Methoden, 3*, 145–164.
17. Wright, B., & Masters, G. (1982). *Rating scale analysis*. Chicago: MESA Press.
18. Wright, B., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.
19. Zumbo, B. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (Ordinal). Item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
20. Ware, J., Harris, W., & Gandek, B. (1997). *MAP-R for Windows: Multitrait-multi-item analysis program—Revised user's guide*. Boston, MA: Health Assessment Laboratory.
21. Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. American Psychological Association.
22. Rajmil, L., Berra, S., von-Rueden, U., Tebe, C., Erhart, M., Gosh, A., et al. (2004). Representativity of 12 national surveys of children and adolescents 8–18 years old included in the KIDSCREEN HRQOL study. *Quality of Life Research, 13*(9), 1576.
23. Veldman, D. (1978). *Fortran programming for the behavioural sciences*. New York: Holt, Rinehart, & Winston.
24. Jöreskog, K. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity, 24*, 387–404.
25. Maruyama, G. (1998). *Basic of structural equation modeling*. Thousand Oaks, CA: Sage.
26. Bollen, K., & Long, J. (1993). Alternative ways of assessing model fit. Testing structural equation models. Sage Publications.
27. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491.
28. Jöreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.
29. Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
30. McHorney, C. A., Ware, J. E. Jr., Lu, J. F., & Sherbourne, C. D. (1994). The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care, 32*, 40–66.
31. Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
32. Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill, Inc.
33. Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.
34. Mellenbergh, G. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7*, 105–107.
35. Martin, M., Blaisdell, B., Kwong, J. W., & Bjorner, J. B. (2004). The Short-Form Headache Impact Test (HIT-6) was psychometrically equivalent in nine languages. *Journal of Clinical Epidemiology, 57*, 1271–1278.
36. Linacre, J. (2003). *A user guide to Winsteps. Rasch model computer program*. Chicago, IL: MESA edition.
37. Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
38. Jöreskog, K., & Sörbom, D. (1988). *PRELIS – A program for multivariate data screening and data summarization. A preprocessor for LISREL* (2nd ed.). Chicago, IL: Scientific Software, Inc.
39. Ravens-Sieberer, U., Gosch, A., Rajmil, L., et al. (2006). The KIDSCREEN-52 quality-of-life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Expert Review of Pharmacoeconomics & Outcomes Research, 5*(3), 353–364.
40. Warschburger, P., Landgraf, J., Petermann, F., & Freidel, K. (2003). Health-related quality of life in children assessed by their parents: Evaluation of the psychometric properties of the CHQ-PF50 in two German clinical samples. *Quality of Life Research, 12*, 291–301.
41. Varni, J., Seid, M., & Kurtin, P. (2001). 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, 39*, 800–812.
42. Gorsuch, R. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
43. Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, Issues, and Applications* (pp. 76–99). Thousand Oaks, CA: Sage.
44. McHorney, C., & Tarlov, A. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4*, 293–307.
45. Scientific Advisory Committee of the Medical Outcomes Trust (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193–205.
46. Crane, P., Hart, D., Gibbons, L., & Cook, K. (2006). A 37-item shoulder functional status item pool had negligible differential item functioning. *Journal of Clinical Epidemiology, 59*, 478–484.
47. Erhart, M., Ravens-Sieberer, U., Hagquist, C., Robitail, S., & van Buuren, S. (2006). Does correcting for differential item functioning (DIF) using two different techniques enhance the validity and diagnostic quality of adolescents HRQoL test scores? *Quality of Life Research, 15*, A113–A114.