Developing the DCGM-12:
A short-form of the DISABKIDS condition-generic module assessing health related quality of life in children and adolescents with chronic conditions

Doctoral Thesis

Holger Muehlan
University of Hamburg

1. Dissertation supervisor (main):	Prof. Dr. Silke Schmidt (Ernst-Moritz-Arndt-University of Greifswald)
2. Dissertation supervisor (second):	PD Dr. Susanne Fricke (University of Hamburg)
1. Disputation supervisor (main):	Prof. Dr. Ulrike Ravens-Sieberer (University Clinic of Hamburg-Eppendorf)
2. Disputation supervisor (second):	Prof. Dr. Berhhard Dahme (University of Hamburg)
Date of Disputation:	21. October 2010
Place of Disputation:	Hamburg
Year of Printing:	2010

Printed in Hamburg.

Acknowledements

This thesis has a long history. Inspired by Monika Bullinger at the time I was joining the “international quality of life” working group, being employed at the university hospital of Hamburg-Eppendorf in 2003 to 2005, the work was ultimately finalized between 2005 and 2008 in various places in Germany. Step by step, sometimes just one step forwards and two steps backwards, the thesis slowly grew.

There are many people I would like to include in my acknowledgements, especially my former colleagues in Hamburg and Mainz. Amongst the most important there are: First my dissertation supervisors Prof. Dr. Monika Bullinger (Hamburg) and Prof. Dr. Silke Schmidt (Greifswald), which supported me with enough of their “TripleS” (“supervisor’s special skills”), that is: support, trust, and patience. Second, all of my supervisors during my two-years-appointment in Mainz, namely Prof. em. Gernot Huppmann (Oppenheim), PD Dr. Holger Schulz (Hamburg), and PD Dr. Jochen Hardt (Mainz). All of them provided me opportunity to contuinue work on my thesis while I was not especially engaged for doing so. Moreover, I would also like to thank the latter two colleagues for some fruitful methodological talks and inspirations. This also holds for Prof. Dr. Matthias Morfeld (Stendal), Dr. Carsten Maurischat (Kiel), Dr. Friedrich Funke (Jena), and Dr. Sabine Rehmer (Jena). I’m thankful to Dr. Karl-Andrew Woltin (Brussels), who spends hours with correcting the language and provide me with a lot of comments regarding the thesis. I also would like to thank Prof. Dr. Rüdiger Trimpop (Jena), who provided me with facilities in terms of office rooms and technical utilities for preparing my thesis at the end of 2007. Some of my friends have to point out here for their very helpful “support work” during my years in the Main area, namely Heike and Janko Zoellner. I guess this place is truely the best to thank both of them for turning the private side of these two years in Frankfurt/Main into a great pleasure to me. I also would like to thank my parents, “Irmchen” and Werner. For everything else, not to be described in more detail, I would like to thank Sabine.

I could not have done it without every single one of you. Thank you all.

Content

Part I

1 Quality of life assessment

1.1 Introduction

Quality of life (hereafter *QOL*) as applied to the health care system has been understood in a variety of ways, ranging from more general conceptualizations, such as the definition provided by the WHOQOL Group (Power et al., 1998), to more pragmatic definitions (Bullinger, 1991), which allow the respective construct to be translated into a measurement model for QOL assessment. Similarly, the range of dimensions included within QOL assessments depends on the conceptual model implemetend (e.g. Bowling, 1997, 2001). Thus, in a pragmatic sense, irrespective of questions about which facets and which structure of QOL are to be assumed, QOL can be more simply defined by the way it is measured.

The history of developing measures assessing QOL regarding health-related issues is relatively recent and still ongoing. Overall, this progress can be divided into three main phases (Steinbüchel et al., 2005): (I) A first phase took place at the end of the 1980s and is characterised by the *construction of generic measurements* and their validation at the beginning of the 1990s. (II) In the following phase, starting by the beginning of the 1990s, *developing and validating disease/condition specific measurements* has been of most interest. (III) Finally, since the middle of the 1990s, more research endeavour has been directed to the *integration of different measurement models into one general model*.

In addition to those historical developments, some actual *"trends"* in constructing QOL assessments can be identified: (I) *Shortening measures or developing short-forms and indices* of QOL assessments. (II) *Examining the cross-cultural comparability* of QOL assessments. (III) Addressing more *specific target populations* with respect to different age groups and regarding varying health status or specific health conditions. (IV) Developing measures allowing for the assessment of QOL at *different levels of generality*.

These four selected issues in QOL assessment development will be addressed in the following sections. Much work has been published focussing on the last three points, namely covering the characteristics of cross-cultural approaches in QOL assessment development, assessing QOL

in different populations, as well as implementing different levels of generality in QOL measurement (see Bowling, 1997 & 2001 for an overview). The second chapter of this thesis emphasizes the aspect of economy in terms of shortening measures and developing short-forms. Some basic ideas and common procedures in shortening measurements in general are outlined. A selection of prominent examples of short-form development as applied to QOL assessment by means of such current developments is given prior to the "short-form chapter". The aim is to give a brief overview of the field and to add illustrative and prominent examples from the area of QOL assessment.

1.2 Some general trends in QOL assessment

1.2.1 Assessing QOL more economically

In the past few years, there has been a growing demand for the construction of measures and diagnostic tools that are economic in nature. This is documented by the increasing number of short-forms, indices, single-item indicators, or computer-adaptive tests developed in the QOL area.

The general trend for developing short-forms of psychological assessment is noticeable. *Popular examples* from the area of clinical psychology are various short-forms derived from the Symptom-Checklist (SCL-90-R; e.g. Franke, 1995, 2002), originally developed by Derogatis (1992). Even though there is a translated Brief Symptom Inventory in German (*BSI*; Franke, 2000; original version by Derogatis, 2003), within the last few years a series of different short-forms of the SCL-90-R have been published, varying with respect to the number of items included and the dimensional structure identified (*SCL-9*: Klaghofer & Brähler, 2001; *SCL-14*: Harfst et al., 2002; *SCL-27*: Hardt et al., 2004).

Since the demand for more economical measures in QOL assessment is still considerable, the construction of *"conventional"* short-forms – referring to fixed item pools, paper-and-pencil completion, and without computer assisted administration – is undampened. Although computer-adaptive testing and computer-assisted administration of measurements for assessing and quantifying QOL represent very economical and efficient diagnostic tools, they do not completely replace brief questionnaires that assess QOL the conventional way.

1.2.2 QOL assessment from a cross-cultural perspective

The need for *cultural equivalence* in QOL measurement has become more important in the last few years (e.g. Bullinger, 2003; Bullinger et al., 1995; Bullinger & Schmidt, 2007; Schmidt & Bullinger, 2003, 2007). Developing QOL assessments in a cross-cultural context emerged, to meet the need for cross-cultural comparability. Since the beginning of the 1990s, more attention has been paid to the importance of cross-cultural adaptation and test score comparability in constructing measures for the assessment of QOL in large multinational studies (e.g. *IQOLA project*; Aaronson et al., 1992).

Some special prerequisites for using QOL assessments in cross-cultural studies are to be considered; first and foremost various kinds of *equivalence* of measurements, especially *functional equivalence*, as a precondition for cross-cultural comparability of test scores from populations of different cultural backgrounds (e.g. Harkness, 1998). Recently, several strategies and procedures for establishing, ensuring, and thus achieving cross-cultural equivalence have been developed, adopted, and conducted within the field of QOL assessment. For reasons of simplicity, those strategies and procedures can be roughly clustered with respect to their location within the process of test development into two groups (construction step, selection step):

- Within the *construction step* of a test a number of strategies have been proposed for achieving equivalence. With respect to general steps in test construction, one can distinguish at least three typical *approaches* (e.g. Bullinger et al., 1996): the sequential approach, the parallel approach, and the simultaneous approach. The *simultaneous approach* requires that within cross-cultural projects, tests should be developed by simultaneously identifying and harmonizing items and tests in different languages and for different cultural settings.
- Within the process of *item selection*, the question of eliminating items with substantial lack of *equivalence* across countries or cultures arises.

The demand for establishing and ensuring cross-cultural equivalence in assessment has been voiced by institutions of health policy (e.g. World Health Organization, WHO; European Commission, EU; National Institute of Health U.S.A., NHI). In parallel, financial support for various projects in the area of international QOL research has become available. These projects are working on developing cross-cultural validated measures not only by *evaluating* equivalence but also by applying a simultaneous approach by the means of cross-cultural test *construction* (e.g. The DISABKIDS Group Europe, 2006; The KIDSCREEN Group Europe, 2006; WHOQOL Group: Skevington, 2004; Skevington et al., 2004).

1.2.3 QOL assessment in specific populations

The growing interest in and a decisive focus on QOL as well as subjective health status in specific populations, has led to a change in the emphasis from adults to more *specific age groups*, such as children and adolescents (e.g. Bullinger & Ravens-Sieberer, 1995; Landgraf et al., 1997; Ravens-Sieberer & Bullinger, 1997). Recently, more emphasis has been placed on QOL of older people and the "oldest old" (e.g. Farquhar, 1995; Haywood et al., 2005; Stewart et al., 1996).

In addition, QOL assessment has become more sensitive to specific populations with respect to *different health status or various diseases*. As a consequence, appropriate measures which are sensitive to specific burdens for a wide range of diseases are available, and the number of questionnaires developed for QOL assessment specific to various diseases increased rapidly (Bowling, 2001).

Another facet of differentiation related to specific target populations concerns the construction of measures within a *modular approach*. In addition to a core QOL assessment, various project groups have developed additional measures, appropriate for specific groups or for specific areas of QOL. For instance, included with the core measure for QOL assessment in patients with cancer, the EORTC Group provides many modules for disease specific QOL assessment, sensitive to burdens of the respective type of cancer (Aaronson et al., 1994). The EORTC Group applied a parallel approach to cross-cultural development of specific assessments (Bullinger et al., 1996), newly developed modules were and still are sequentially added.

1.2.4 QOL assessment at different levels of generality

As QOL assessment becomes more sensitive to specific target populations, QOL conceptualization and assessment has been differentiated regarding various levels of generality (Rose, 2003):

- *General/global quality of life (QOL)*: Measurements assessing general/global QOL collect information about general QOL. QOL measures are often highly integrated single-measures or generic indices which are sensitive tends to be low. They are used in epidemiological and health care research as well as in population-based studies in the field of social sciences as applied to medicine (e.g. medical sociology). *Example:* The EUROHIS-8 index (Power, 2003) is a measure for brief assessment of general QOL in large surveys.
- *(Generic) Health-related quality of life (HRQOL)*: HRQOL measures assess information about general (subjective) health status at a population level. They usually provide a multidimensional profile of characteristics, including at least physical, mental, and social aspects of subjective health, according to a bio-psycho-social understanding of health as defined by the WHO. HRQOL measures are sensitive to change primarily for health-related events. General comparisons can be made between populations with different levels of health status (e.g.

healthy vs. unhealthy people). In addition, clinical comparison can be made between different diseases. *Example:* The SF-36 Health Survey measure (Ware, 1993) is usually referred to as the most widely used and popular HRQOL questionnaire.

- *Disease/Condition generic quality of life (CGQOL)*: CGQOL measures assess information about common burdens by various diseases or treatments (especially medications). They usually provide a multidimensional profile of characteristics and are sensitive to change following specific interventions. Clinical comparisons are possible between different diseases regarding disease consequences. *Example:* The DCGM-37 (Petersen et al., 2005; Schmidt et al., 2006; Simeoni et al., 2007) is a newly developed multidimensional measure, and provides for age-appropriated assessment of CGQOL in younger populations (children and adolescents, 8-16 years old) with various chronic health conditions.
- *Condition/disease specific/related quality of life (CSQOL): "*CSQOL measures" (e.g. Rockwood, 2007, p. 74) assess information about burden of specific diseases and the impact of condition specific treatments (medications). CSQOL measures usually provide a multi-dimensional profile of QOL characteristics. They should be sensitive to change for specific interventions and allow for clinical comparison between different therapies. *Example:* The HAEMO-QOL modules (Bullinger et al., 2002; Mackensen et al., 2004; Pollak et al., 2006) are age-appropriate questionnaires for assessing CSQOL in children with haemophilia.

- *Utility measures:* A unique approach to health assessment is represented by a class of measures referred to as "utility measures". As compared to the QOL assessment described above, utility measures differ regarding their focus on patient preferences. Thus, subjective health as seen from the perspective of utility is conceptualized from a "consumer" perspective. Within this group various kinds of measures are subsumized, which all have in common that they are highly integrated single-measures, assessing the personal self-rated values of health states (Bowling, 1997). Methods differ with respect to design, procedure, and scoring of the respective assessment. Utility measures are most widely used in the area of health economics. *Example:* The EQ-5D is a short measure with 5 items assessing HRQOL using a preference-based index-score (Brooks et al., 2003).

Table 1: Conceptual levels in quality-of-life research (adopted and translated from Rose, 2003)

Conceptual level	Content	Properties	Primary area of application	Example	Special Age *(Kids)*
General/ global quality of life (QOL)	Information (statements) about general quality of life	Highly integrated single-measure; sensitive to change only for incisive life-events	Basic research in public health, medical sociology and medical psychology	EUROHIS-QOL	
Health-related quality of life (HRQOL)	Information (statements) about general health status	Multidimensional profile of characteristics; sensitive to change only for health-related events	Comparison between different health status groups (general population) and different diseases (clinical population)	SF-36	KIDSCREEN measures
Condition/Disease generic quality of life (CGQOL)	Information (statements) about common burdens by different diseases or treatments (medication)	Multidimensional profile of characteristics; sensitive to change for interventions	Clinical comparison between different diseases	EORTC-CORE Module	DISABKIDS DCGM
Disease specific/ related quality of life (CSQOL)	Information (statements) about the specific burden by specific diseases or treatments (medications)	Multidimensional profile of characteristics; sensitive to change only for specific interventions	Clinical comparison between different therapies	EORTC Modules	DISABKIDS DSM-Modules
Utility measures	Information (statements) about impact of the *specific* disease on *general* quality of life	Highly integrated single- measure for disease *and* quality of life	Health economics; clinical comparison between different diseases or therapies	EUROQOL EQ-5D	Child EUROQOL (EQ-5D)

1.3 Selected examples

In the previous section, some examples of measures were given which provide short-forms and indices of original versions and assess QOL on different levels of generality and with respect to various populations (*original measures*: WHOQOL-100, SF-36, KIDSCREEN-52, HAEMO-QOL). These measures were developed or adopted for use in cross-cultural studies within large international multicentre studies (*WHOQOL*: www.who.int; *IQOLA*: www.iqola.org*; KIDSCREEN*: www.kidscreen.com; *HAEMO-QOL*: www.haemo-qol.de). Different levels of QOL generality will be outlined in detail in what follows and will be referred to as life satisfaction or "general QOL" (*QOL*) respectively, "health-related QOL" (*HRQOL*), "disease/condition generic QOL" (*CGQOL*), and "disease/condition specific QOL" (*CSQOL*). Regarding their use in specific populations, measures presented in this section vary according to their application to various age groups (adults/ children) and according to the health status they were designed for (healthy people/general population vs. unhealthy people/clinical population). Special emphasis is put on the linkage between the different versions; this is to say how the respective shorter version represents the original measure regarding conceptual or structural issues.

1.3.1 Generic quality of life (QOL) in the general population (adults): The WHOQOL measures

General information: The WHOQOL measures are questionnaires for assessing QOL in the general population. The measures were developed on behalf of the World Health Organization (WHO). The WHOQOL measures include the original "long" version WHOQOL-100 and the short version WHOQOL-BREF. In addition, an 8 item QOL index was derived from the WHOQOL measures within the EUROHIS project (Nosikov & Gudex, 2003), referred to as "EUROHIS-QOL" (Schmidt et al., 2006-b).

Original "long-form" measure (WHOQOL-100): The "WHOQOL-100" questionnaire is a generic multidimensional measurement, assessing subjective general QOL of adults. The WHOQOL-100 module comprises 100 items pertaining to 6 domains: (I) „Physical QOL", (II) „Mental QOL", (III) „Independence", (IV) „Social Relationships", (V) „Environment ", and (VI) „Spirituality". Items are assigned to 24 factors (4 items each), with the 4 remaining items being global judgements (single-item measurements) of general QOL and health. Psychometric performance of the WHOQOL-100 (Power et al., 1998) was evaluated as acceptable to very good with respect to reliability in terms of internal consistency of the scales (α = .59 - .91).

Short-form measure (WHOQOL-BREF): The short-form of the WHOQOL-100 measure is referred to as "WHOQOL-BREF" (The WHOQOL Group, 1998). It is also a generic multidimensional measurement, assessing subjective general QOL of adults. As compared to the WHOQOL-100 with 6 domains, the 26 items of the WHOQOL-BREF pertain to 4 domains only: (I) "Physical QOL" (7 items), (II) "Mental QOL" (6 items), (III) "Social Relationships" (3 items), and (IV) „En-

vironment“ (8 items). The excluded domains from the WHOQOL-100 have been merged into one of these four remaining domains (“Independence” into “Physical QOL” and “Spirituality” into “Mental QOL”). Two items are global judgements on general QOL and overall health status. Every domain comprises the same relative amount of items as compared to the original domains of the WHOQOL-100. Tthus, the item number of each domain is equivalently lowered by roughly 75%. The short-form approach can be referred to as “deductive approach”, as the conceptual model and dimensional structure of the WHOQOL-BREF has been derived from the WHOQOL-100. However, it is not fully equivalent due to the merging (and reduction) of domains. Nevertheless, this approach has been chosen to establish a measure that could be treated as a brief version of the WHOQOL-100. Psychometric performance of the WHOQOL-BREF was found to be acceptable to good with respect to reliability in terms of internal consistency of the scales (α = .57 - .88).

Index (EUROHIS-QOL): An index derived from the WHOQOL measures was developed and validated within the EUROHIS project (Nosikov & Gudex, 2003), and accordingly is referred to as “EUROHIS-QOL” (Power, 2003; Schmidt et al., 2006-b). The EUROHIS-QOL serves as a generic measurement, assessing subjective global QOL in the general population. Although it aimed to assess one dimension with respect to psychometric properties, it conceptually represents items derived from different domains of QOL represented within the original WHOQOL measures. This index consists of 8 items, selected from the WHOQOL-BREF. Important to know, all items of the WHOQOL-BREF originally are chosen from the WHOQOL-100 (see above). The EUROHIS-QOL index was developed combining various item selection strategies (for details see Power, 2003). Psychometric performance of the 8 item measure was found to be good with respect to reliability in terms of internal consistency of the scale (α = .86).

Additional notes: Within the WHOQOL Group, the WHOQOL approach also had been applied to developing a QOL module for older and oldest old people (*“WHOQOL-OLD”*; Power et al., 2005). This module is intended to be used in addition to generic WHOQOL measures in surveys within populations of older people (> 60 years), but it may also be used by itself. A short-form measure of the WHOQOL-OLD module is in progress (Schmidt, personal communication).

1.3.2 Health-related quality of life (HRQOL) in the general population (adults): The SF Health Survey measures

General information: The SF-36 Health Survey measure is a questionnaire for assessing HRQOL in healthy and ill populations. Since the SF-36 was originally derived from the Medical Outcomes Study Questionnaire (MOS), the SF-36 questionnaire is a short-form itself. Nevertheless the SF-36 can be considered as an original version, as it has been frequently used and is very well documented in published studies (Coste et al., 1997). In addition, conceptual model and measurement approach of the SF-36 Health Survey measure have been applied as a template in test construction to a varitey of measures, with the SF-12 questionnaire

as the most well-known derivate amongst other ones. On the contrary, the original MOS Health Survey measure is not commonly used.

Original measure (SF-36): The "*SF-36 Health Survey*" (*SF-36*; Ware, 1993) is a generic and multidimensional measurement, assessing subjective HRQOL of adults. It comprises a profile of 8 facets leading to 2 summary scores of HRQOL. This measure is a questionnaire with 36 items, which have to be answered on scales with varying answering categories and different response choice wordings (anchors). The 36 Items pertaining to the 8 facets can be combined into 2 overall sum scores: (I) *"Physical sum scale":* (i) "Physical Functioning", (ii) "Bodily Role Functioning", (iii) "Pain", and (iv) "Global Health Perception"; (II) *"Mental Sum Scale":* (v) "Vitality", (vi) "Social Function", (vii) "Emotional Role Functioning", and (viii) "Mental well-being". The SF-36 has been cross-culturally adopted by the IQOLA project group within a large multinational study (Aaronson et al., 1992)..

Short-form (SF-12): The "*SF-12 Health Survey*" (*SF-36*; Ware et al., 1996) is a short-form of the SF-36 and is recommended for use within the same populations. It is also a generic and multi-dimensional measurement, assessing subjective HRQOL of adults. This questionnaire comprises 12 items, which were derived from the SF-36 item pool using regression analysis guided by a representative approach. The structure model of the SF-12 was established to represent the SF-36 structure at the level of both sum scales. Thus, the SF-12 is considered to reproduce the domain structure of the SF-36 structure model, also resulting in two summary scales. Sample items from all respective factors of the SF-36 were chosen for the SF-12. Precision of sum scores in terms of explained variance with respect to SF-36 sum scales varies between 80%-90%.

Index (SF-8): The SF-8 index (Ware et al., 1999) is a single-item factor measurement. Items included are not directly derived from the SF-36/-12 Health Survey measures. Each item is considered to represent the main content of one of the eight subscales from the SF-36; thus, every item serves as indicator for factor of the original scales. Scoring procedures also recommend the additional computation of scores for both summary scales (physical/mental). The index was established using a representative approach of short-form/index development.

Additional notes: There are several further developments, for instance a *preference-based application* (see above: "utility measures") derived from the SF-36 Health Survey measurement approach, referred to as *"SF-6D"* (Brazier et al., 1998).

1.3.3 General quality of life (QOL) in specific age groups (children & adolescents): The KIDSCREEN measures

General information: The European KIDSCREEN project was designed as a cross-cultural epidemiological project, which aimed to develop a survey instrument on QOL of children and adolescents in the general population (Ravens-Sieberer et al., 2001; The KIDSCREEN Group Europe, 2006). The project was funded by the European Union (EU) and was closely related to the DISABKIDS project.

Original measure (KIDSCREEN-52): The core instrument of the KIDSCREEN project (*KIDSCREEN-52*; The KIDSCREEN Group Europe, 2006) is an age-appropriate, multidimensional measurement, assessing general subjective QOL of children and adolescents (8 to 16 years old). This questionnaire comprises 52 items pertaining to 10 facets: (1) “Physical Well-being” (5 items), (2) “Psychological Well-being” (6 items), (3) „Moods & Emotions“ (5 items), (4) „Self-Perception“ (5 items), (5) “Autonomy“ (5 items), (6) „Parent Relation & Home Life“ (6 items), (7) “Financial Resources” (3 items), (8) “Social Support & Peers” (6 items), (9) “School Environment” (6 items), and (10) “Social Acceptance/Bullying” (3 items). Psychometric performance of the original 52 item version can be judgedto be good to very good with respect to reliability in terms of internal consistency of the scales (α = .77 - .89).

Short-form measure (KIDSCREEN-27): The „KIDSCREEN-27“ (KIDSCREEN-27; Ravens-Sieberer et al., 2007; Robitail et al., 2007) is an age-appropriate, multidimensional measurement, assessing general subjective QOL of children and adolescents (8 to 16 years old). This module is a questionnaire comprising 27 items which belong to 5 subscales: (1) “Physical Well-being” (5 items), (2) “Psychological Well-being” (7 items), (3) “Autonomy & Parent Relation” (7 items), (4) “Social Support & Peers” (4 items), and (5) “School Environment” (4 items). Regarding conceptual representation, the dimensional structure of the KIDCREEN-27 measure follows a largely deductive approach, as some factors of the KIDCSREEN-52 are merged together. However, as not all facets are represented by the short-form measure (“Bullying” was excluded), this approach is also linked to “refinement” intentions and differs from unique “short-form”-approaches. Therefore it can be referred to as a “merging” approach, even though this denotation might not fully cover the specifities of this procedure. Psychometric performance of the final 27 item version can judged to be good with respect to reliability in terms of internal consistency of the scales (α = .80 - .84).

Index (KIDSCREEN-10): The KIDSCREEN-10 is a unidimensional index, containing 10 items, taken from the KIDSCREEN-27 measure (thus, all items are also included in the KIDSCREEN-52). The KIDSCREEN-10 index assesses the main content areas of the longer KIDSCREEN measures. The items provide a composite score, indicating global QOL (The KIDSCREEN Group Europe, 2006).

1.3.4 Condition specific quality of life (CSQOL) in specific age groups (children): The HAEMO-QOL measures

General information: The Haemo-QOL project was a European cross-cultural study in six countries, aimed at developping disease specific measures for age appropriated assessment of CSQOL in children and adolescents with haemophilia applying a cross-cultural perspective.

The „*HAEMO-QOL*“ (HAEMO-QOL; Bullinger et al., 2002; von Mackensen et al., 2004) is a age-appropriate, condition specific as well as multidimensional measurement, assessing subjective health-related quality of life of children and adolescents (4 to 16 years old) with haemophilia. This measure is available in three different age-appropriated versions, each of them with different numbers of items and facets as well as varying response choice formats. Items (statements) have to be answered with respect to their degree of appropriatedness on three-point (HAEMO-QOL I) or five-point (HAEMO-QOL II & III) Likert-like scales. The three versions of the HAEMO-QOL quality of life measure are the following: *HAEMO-QOL I* (age group 4 - 7 years, 29 items, 8 facets; 3-point response choice format), *HAEMO-QOL II* (age group 8 - 12 years, 84 items, 9 facets; 5-point response choice format) and *HAEMO-QOL III* (age group 13 - 16 years, 91 items, 11 facets; 5-point response choice format). All items of the three HAEMO-QOL versions are assigned to one summary score (comprising 29, 84, and resp. 91 items) and pertain to at least 8 (“shared”) facets (*below in parenthesis*: indication of number of items per facet for the HAEMO-QOL I, resp. HAEMO-QOL II & III): (a) „Physical health“ (5 resp. 9 items), (b) „Emotions“ (3 resp. 6 items), (c) „Views“ (3 resp. 7 items), (d) „Family“ (5 resp. 11 items), (e) „Friends“ (2 resp. 8 items), (f) „Other people“ (5 resp. 13 items), (g) “Sports and school” (4 resp. 11 items), and “Treatment” (2, resp. 9 items). Some items of the HAEMO-QOL version II and III are assigned to the additional “Coping” facet (11 items each). Other specific items of the HAEMO-QOL version III are assigned to the additional “Future” facet (4 items) or the “Relationships” facet (3 items). Psychometric performance of the final versions were analyzed using data from the HAEMO-QOL field study sample (N = 339 participants; von Mackensen et al., 2004) and varies on the one hand regarding the three age-appropriated versions of the HAEMO-QOL and on the other hand regarding to the different facets comprising each version. With respect to internal consistency, the values indicating reliability (Cronbach’s alpha) ranged from .45 to .82 for the subscales of the HAEMO-QOL I, from .60 - .79 for the HAEMO-QOL II, and from .52 to .87 for the HAEMO-QOL III. Thus, for all three versions reliability can be judged to be suffcient to very good.

Index: The HAEMO-QOL 8 item index (Pollak et al., 2006) represents the core content of the original long version. According to preliminary analyses, the index's psychometric performance concerning reliability and convergent validity is satisfactory. Nevertheless, the data do not support the uni-dimensionality of the measurement model.

Additional notes: Each of the three age-appropriate HAEMO-QOL versions is also available as a self-report and as proxy-report measure. The HAEMO-QOL was originally translated in all six languages spoken by the HAEMO-QOL project participants (Dutch, English, French, German, Italian, and Spanish). Within the ESCHQOL project, there are now 22 language versions.

1.4 Concluding remarks

Constructing short-forms and indices is an important task in QOL assessment development. Moreover, presenting a shorter and more economic alternative of the original long-form, has become an important for QOL research.

For both of the probably most widely used measurements in the area of QOL research, the WHOQOL-100 measure (Power et al., 1998) and the SF-36 (Ware, 1993), short-forms and indices exist, which are closely related to the conceptual approach of the original measures. Therefore, they provide useful tools for epidemiological screening surveys and health research. The short-forms of the WHOQOL-100 and SF-36 measures were developed later than their respective originals, in the meantime, some short-forms and indices were developd within a simultaneous approach. This approach has already been applied, for example within a multinational study on QOL assessment development. The KIDSCREEN Group (2006) also provided short-form measures within their modular tool set of QOL assessment.

From a general perspective on QOL assessment, the development of short measures has two perspectives. As QOL assessment has become more differentiated, the need for short measures results from this development. If needed and feasible, even short measures should fulfil the purposes of cross-cultural comparability, age-appropriate measurement, and sensitivity to the specifics of a respective target population (e.g. diseases). Moreover, some measures also allow for a choice between a short-form providing multidimensional QOL profiles, and a global index, providing a global unidimensional assessment of QOL. Profiles and indices each have their own advantages and disadvantages, depending on the purpose they need to fulfill (Bullinger, 1993).

2 Short-form development

The intentions for constructing "conventional" short-forms are manifold. Generally speaking, there is still a *necessity* for developing economical assessments with *fixed* item pools (as opposed to computer-*adaptive* testing) in all areas of application for which the possibility of computer-assisted presentation is not available or only available with numerous limitations. Moreover, there is a *need* for economical assessment in some special areas of application (e.g. clinical routine trials, screening, evaluating, monitoring, etc.) In particular when conducting surveys there are difficulties with the application of long measures in census and survey research due to the occurrence of fatigue, loss of motivation, and higher drop-out rates resulting from too long tests. Notwithstanding this, the demand for economical and at the same time reliable and valid assessment methods has so far not been fully met.

2.1 Rationale for "shortening" measures: From test refinement to short-form development

In general, the core *rationale* for shortening a test is to provide a more economical assessment.

Furthermore, psychometric deficiencies of measurements initiate *"refinement"* motivations to improve insufficient measurements by developing a revised version of the original measure (Smith & McCarthy, 1995). As enhancing psychometric performance of a measure often is realized by item selection, "refinement" and "shortening" a measure are intertwined.

In contrast, just "*shortening"* a measurement (Coste et al., 1997) could also serve as an explicit aim in test construction, with the goal to economize a measure. As compared with "short-form" development in particular, "shortening" a measure is sometimes guided to a smaller degree by the conditions of the original measure in terms of its conceptual representation. Developing *"short" forms instead of "short-forms"* does not necessary mean the same. Thus, the difference between shortening a measure and short-form development of a measure should be kept in mind.

Answers on how to define a *"short-form"* have been put forth recently (Smith et al., 2000): In the broadest sense of the word, "short-form" refers to all those measurements, which were developed or generated on the basis of an item pool of a given original version. As already mentioned, refinement by shortening existing tests is not equivalent to developing a short-form even though in both cases item selection results in a reduced item pool. "Short-form" refers to an assessment derived from a shared item pool provided by a measurement tool to which the short-form measure should be strongly related in terms of conceptual representation, statistical

association, and comparability regarding psychometric performance. As compared to the original “long-form” version of a questionnaire, such a short-form version should provide an even more economical option for assessing QOL by the same conceptual approach regarding respective domains and populations.

To *summarize*, with respect to these different purposes in developing shortened measures, one can differentiate between: (I) *„Refinement“*, referring to improving deficiencies of psychometric performance at item-, scale-, or test-level. (II) *„Shortening“,* referring to reducing both the item pool and test length respectively (often oscillating in the blurred area between „refinement“, and “short-form development“). (III) “*Short-form development*“*,* referring to the construction of an economical alternative assessment.

2.2 Strategies and procedures in short form development

Based on a literature review spanning the years from 1985 to 1995, Coste et al. (1997) provide an overview of common *special procedures* of short-form development. In general, they differentiate between (I) *expert-based*, (II) *statistical*, and (III) *combined* approaches. *Expert-based approaches* mainly differ with regard to the degree of involvement (authors, other experts in the field, patients/volunteers). *Statistical methods* are manifold. The most widely used strategies are checking for correlation between long and short form scores (simple or adjusted for common error), internal consistency of dimensions (Cronbach’s alpha), correlation of items and the composite scores (item-total, item-remainder), and factor analysis.

All these *statistical procedures are linked with different approaches in short-form development*, as each one aims to ensure different purposes. Correlations between long and short-form scores are intended to prove a high *association* between both measures. Internal consistency is used to establish reliability in terms of *homogeneity* of the short-form measure. Correlations of items and the composite score are interpreted as indicator of the amount to which an item represents the *content* of the respective dimension (this is also indicated by the composite score of all items). At minimum, (exploratory) factor analysis aims at identifying the structure of the item pool, thus it is primarily linked to the question of *conceptual and dimensional representation.*

A very unique approach in developing short-forms or indices is represented by the construction of *“single-item factors”.* Whereas the intention in constructing single-item factors is the same as for “conventional” short-forms, it differs in its construction approach as the question of item selection is not of greatest interest. On the contrary, new items can be constructed, based on facets and dimensionality conveyed by the original structure or conceptual model through a highly semantic aggregation of item contents belonging to one common dimension. Within the scope of QOL research the possibility of constructing *single-item factors,* particularly as outlined

above, has not widely occurred so far. This is rather surprising given that single-item measurements in general are of common use in QOL research.

Single-item measures usually are *not* short-forms of the original "long" versions and in addition not the same as "single-item factors" as outlined above. Nevertheless, "single-item factors" are usually used as "single-item" measures and if appropriate, basically every item of a questionnaire can be treated as a "single-item measurement". A multitude of global single-item indicators of QOL and well-being as applied to heath-related issues already exist, especially for assessing subjective health status. Furthermore, many studies have critically reflected the application of single-item measures and investigated their psychometric properties compared to conventional assessments (e.g. Barofsky, 2004; Cunny & Perry, 1991; Hurny et al., 1996; Sloan et al., 2002; Youngblut & Casper, 1993).

2.3 Psychometric performance of short-forms

In developing a short-form of the original version, the aim is to reach a high degree of correspondence with the original measure (as indicated by "*external related" psychometric properties*) and to ensure a sufficient degree of psychometric performance of the short measure itself (as indicated by "*internal related" psychometric properties*).

Internal-related *psychometric performance* is indicated by conventional psychometric properties, including different forms of objectivity, reliability, validity, or sensitivity to change. Evaluation of externally-related psychometric performance of the short-form is based on the objectives of conceptual representation, statistical association, and psychometric comparability. Economy and efficiency of a short-form can be seen as an additional goal of short-form development, linking external and internal related psychometric properties.

Representation refers to the degree to which the short-form measure represents the conceptual model (including main content areas) or dimensional structure of the original version. *Association* between short-form measure and long-form measure usually is displayed statistically by correlation between respective scores of both measures (e.g. total scores). *Comparability* of both measures regarding psychometric performance refers to core psychometric aspects as well as to scoring issues, composite score performance, and score interpretation. *Economy* is defined by the reduction in test length and thus time saved.

3 The European DISABKIDS project

3.1 Background of the project

The *European DISABKIDS project* aimed at developing a modular system for the assessment of subjective health status, health care needs, and HRQOL of children and adolescents with different chronic conditions, adopted for use in cross-cultural context (Bullinger et al., 2002-a, 2002-b). This project focussed on cross-cultural development and validation by implementing a simultaneous approach on test construction (Bullinger et al., 1996; see chapter 1).

The methodological steps applied by the European DISABKIDS Group were largely identical with the ones of its sister project KIDSCREEN, a cross-cultural epidemiological project, which aimed at developing a survey instrument on HRQOL of children and adolescents in the general population (Ravens-Sieberer et al., 2001; The KIDSCREEN Group, 2007). Strategies implemented for ensuring cross-cultural comparability in particular have been reported in detail elsewhere (Schmidt et al., 2006-a).

The modular approach of the DISAKIDS project includes diagnostic facilities regarding different levels of assessment (generic/health-related vs. disease specific), different age groups (4-7 vs. 8-16 years), and different respondents (self- vs. proxy-report). Within the DISABKIDS conceptual framework with its modular tool set, various disease specific quality of life assessments (DSM; Baars et al., 2005), a “Child Health Care Questionnaire on Satisfaction, Utilization, and Needs” (CHC-SUN; Schmidt et al., 2007), a “Coping with Chronic Disease Inventory” (CODI; Petersen, 2004; Petersen et al., 2006), as well as a “Clinical and Parental Severity Assessment” (The DISABKIDS Group, in prep.) have been constructed.

However, it was also intended to develop a short-form of the core multidimensional DCGM-37 measure (Petersen et al., 2005; Schmidt et al., 2006-a; Simeoni et al., 2007), a generic questionnaire assessing HRQOL in children and adolescents (8-16 years) with different chronic conditions. Constructing such a short-form is referred to as the core rationale of this thesis.

3.2 DISABKIDS chronic generic module (DCGM-37)

The original "long-form" measure of the „*DISABKIDS chronic generic module*" (DCGM-37; Petersen et al., 2005; Schmidt et al., 2006-a; Simeoni, 2007) is an age group adopted, generic/-condition independent as well as multidimensional measurement, assessing general subjective HRQOL of children and adolescents (4 to 16 years old) with chronic conditions (e.g. asthma, diabetes, epilepsy, etc.).

Table 2: Domains and facets indicating the conceptual model of the DISABKIDS chronic generic module (DCGM) for the generic assessment of health-related quality of life in children and adolescents with chronic conditions

Construct	Domain	Facet	Concept/content
Health-related Quality of life	*Mental*	*Independence*	Autonomy, living without impairments caused by condition
		Emotion	Emotional worries, concerns, anger, problems; negative feelings and thoughts
	Social	*Social Inclusion*	Acceptance of others, positive social relationships
		Social Exclusion	Stigma, feeling left out
	Physical	*Physical Limtation*	Functional limitations, perceived health status
		Medication/ Treatment	(Emotional) impact of taking medication, receiving injections, taking insulin, applying cortisone, etc.

This module is a questionnaire comprising 37 items (statements), which have to be answered on a five-point Likert-type scale (ranging from 1 = "never" to 5 = "always"). With respect to the structure and measurement model – referring to the terminology employed by structural equation modelling (SEM) – these 37 items belong to 6 facets (with 6 items each, except for the "emotion" facet, which comprises 7 items). Moreover, within a more heuristic conceptual model, out of all 6 facets 3 pairs of 2 facets each are assigned to one higher ordered "domain" (see Table 2): (I) "Psychological" domain: (i) "Independence", (ii) "Emotions"; (II) "Social" domain: (iii) „Social exclusion", (iv) „Social inclusion"; (III) "Physical" domain: (v) „Physical limitations", and (vi) „Treatment/medication" (see Table 2). Items assigned to the "Treatment/Medication" facet are only applicable if the respondent receives treatment or takes medication (so-called "applicable" items). Finally, all items are also assigned to one general factor (HRQOL) or one summary score, respectively (31, resp. 37 items included). Item wordings and item-scale-assignments of all items of the final DISABKIDS chronic generic module are displayed in Table 3. Score computation algorithm uses equal-interval scoring of response categories (Likert-type scale scoring).

Table 3: Item wordings and item-scale assignments for the final version of the DISABKIDS chronic generic measure (DCGM-37)

Item No.	Facet (n_{items})	Abbr.	Item wording
01.	Independence (6 items)	IND	Are you confident about your future?
02.			Do you enjoy your life?
03.			Are you able to do everything you want even though you have your condition?
04.			Do you feel like everyone else even though you have your condition?
05.			Are you free to lead the life you want even though you have your condition?
06.			Are you able to do things without your parents?
07.	Limitation (7 items)	LIM	Are you able to run and move as you like?
08.			Do you feel tired because of your condition?‡
09.			Is your life ruled by your condition?‡
10.			Does it bother you that you have to explain to others what you can't do?‡
11.			Is it difficult to sleep because of your condition?‡
12.			Does your condition bother you when you play?‡
13.	Emotion (6 items)	EMO	Does your condition make you feel bad about yourself? ‡
14.			Are you unhappy because you have condition? ‡
15.			Do you worry about your condition? ‡
16.			Does your condition make you angry? ‡
17.			Do you have fears about the future because of your condition? ‡
18.			Does your condition get you down? ‡
19.			Does it bother you that your life has to be planned? ‡
20.	Exclusion (6 items)	EXC	Do you feel lonely because of your condition? ‡
21.			Do your teachers behave differently towards you than towards others? ‡
22.			Do you have problems concentrating at school because of your condition? ‡
23.			Do you feel that others have something against you? ‡
24.			Do you think that others stare at you? ‡
25.			Do you feel different from other children? ‡
26.	Inclusion (6 items)	INC	Do other kids understand about your condition?
27.			Do you go out with your friends?
28.			Are you able to play or do things with other children (e.g. sports)?
29.			Do you think that you can do most things as well as other children?
30.			Do your friends enjoy being with you?
31.			Do you find it easy to talk about your condition to other people?
32.	Treatment (6 items)	MED	*Does having to get help with medication from others bother you?* ‡ *
33.			*Is it annoying for you to have to remember your medication?* ‡ *
34.			*Are you worried about your medication?* ‡ *
35.			*Does taking medication bother you?* ‡ *
36.			*Do you hate taking your medicine?* ‡ *
37.			*Does taking medication disrupt everyday life?* ‡ *
Response choices			*"never" (1), "seldom" (2), "quite often" (3), "very often" (4), "always" (5)*

Notes: ‡ Reversed scored items. * "Applicable" items.

Psychometric performance of the final 37 item version was estimated using data from the DIS-ABKIDS field study sample (N = 1.606 participants; Schmidt et al., 2006-a; Simeoni et al., 2007). With respect to reliability in terms of internal consistency of the scales (α = .70 - .87) and the total score (α = .86) the DCGM-37 was found to be good. Stability in terms of test-retest reliability of the different scores (one-month interval) ranging from .71 to .83 (ICC, intra-class

correlation coefficient) and validity of the measure could also be demonstrated, e.g. with respect to its potential to discriminate between different clinical groups concerning severity of condition (Schmidt et al., 2006-a; Simeoni et al., 2007).

The DCGM-37 module is available in two *different versions*, namely as a self-report and as a proxy-report measure; each version can be administered in the six languages of the DISAB-KIDS project participants (Dutch, English, French, German, Greek, and Swedish). Moreover, versions in Italian, Portuguese, and Spanish language are in progress, with translations into more languages underway (The European DISABKIDS Group, 2006). Translations into further languages are planned. In addition, the DCGM-37 is available as computer-administered version.

Although use of the chronic-generic module is limited to children aged between 8 and 16 years, an additional condition generic module assessing HRQOL in the youngest (4 to 7 years) is available within the DISABKIDS tool set, denoted to as “DISABKIDS smiley version” (Chaplin et al., 2008; The European DISABKIDS Group, 2006). Despite its development and preferred use in the population of the “youngest”, it could be also applied in “older” age groups ranging from 8 to 16 years.

Part II

4 Methods and materials

4.1 Objective

The *objective* of the analyses presented here is to develop an economic DISABKIDS short-form measure, based on the original long-form measure, the DISABKIDS condition generic measure (DCGM-37; Petersen et al., 2005; Schmidt et al., 2006-a; Simeoni et al., 2007). This short-form should be strongly related in terms of conceptual representation, statistical association, and comparability regarding psychometric performance (see chapter 2). Thus, this short-form should reproduce the main content areas of the DCGM-37 as closely as possible, using a selection of the original item pool without changes in wording and response options. As compared to the long-form measure, it should provide an economic option for assessing HRQOL in children and adolescents with different chronic health conditions within cross-cultural context. Furthermore, it should be applicable in in clinical studies and surveys in particular.

Criteria were derived regarding item selection procedure to be choosen, namely:

- Items should not be modified in any way and no new items should be included, respectively.
- Every facet of the multidimensional construct should be represented by the short-form measure equivalent as compared to the original measurement approach.

4.2 Outline of general analytic strategy

For the intention to derive a short-form (index) of the DCGM-37, some special criteria for short-form development should be applied (see also chapter 2):

- *Concepts:* Based on expert consensus, it was decided to construct a short-form in very close correspondence to the original measure (DCGM-37; Petersen et al., 2005; Schmidt et al., 2006-a; Simeoni et al., 2007). “Close correspondence” refers to the issues of conceptual representation, statistical association, psychometric comparability, and economy and efficiency. The selected item pool should perform well regarding psychometric properties and should be applicable as an independent short-form assessment.

- *Methods:* Short-form development was to be composed of two different steps: selection of items and a-priori validation of the short-form. Item selection was performed not only using psychometric indicators of test construction, but by combining a rationale approach (first step) and a psychometric approach (second step). Final decision concerning the item pool was made based on expert consensus. A-priori validation of the extracted short-form item pool was to be provided including estimations of descriptive and psychometric properties on item and composite scale level.

4.2.1 Rationale of short-form construction

As outlined before, there are special requirements in short-form development, as opposed to test construction in general. Apart from conventional "internal-related" criteria of psychometric performance, within short-form development "external-related" performance is also of strong relevance. In addition, *economy* as related to test performance is a fundamental criterion to short-form development in particular – not just a secondary criterion as in conventional test construction in terms of psychometric test theory (Lienert & Ratz, 1994). All different kinds of criteria have to be taken into consideration for our purposes. These criteria are as follows: (I) *representation*: the short-form should represent the original version of the measure as closely as possible (with respect to main content); (II) *association*: the short-form should be associated with its original version as highly as possible; (III) *comparability*: the short form should be comparable to the original version with respect to basic psychometric performance (e.g. different types of validity); (IV) *economy*: the short-form should be more economic as compared to the original version of the measure; (V) *usability*: the short-form should be easy to use – this refers to its practicability and feasibility; (VI) *psychometric performance*: the short-form should be judged sufficiently satifying with respect to fundamental psychometric performance criteria (e.g. reliability); (VII) *dimensionality*: dimensions of the short-form should be homogenous in terms of internal consistency and uni-dimensional in terms of Rasch scalability; factor analysis should confirm the structure and measurement model hypothesized.

In summary, the short-form should be reflecting indicators of "internal" and "external" related psychometric performance.

4.2.2 Rationale of item selection and item requirements

The rationale outlined above dealt with the general methodological conception of the procedure, guidelines, and requirements for item selection. However, the more technical aspects of these criteria need to be formulated. In other words, the rational stated above needs to be translated into specific criteria. These criteria are derived from the general methodological conception, in accordance with the purposes a short-form needs to fulfill, and defined by expert consensus within our working group: (I) From every facet (dimension) of the original measure the *same number of items* should be selected (as item numbers are nearly the same for all facets of the original measure, with the exception of the facet "emotion"). (II) Items should *represent their respective facet (dimension)* best, thus items selected should display high corrected item-total-correlation (for item within the respective dimension score). (III) Items should cover the whole or at least the main *content of their respective facet (dimension)*.

4.2.3 Item selection methods

In general, item selection methods can be divided with respect to various criteria, such as the number of items evaluated at once. Here, two basic classes of methods (already well-known from general statistics) can be distinguished:

- *Univariate selection*: Conventional univariate methods refer to item selection by different types of single item characteristics. All commonly used methods provide several indicators of representation to identify so called "marker" items, e.g. coefficients of item-total correlation (corrected for overlap).
- *Multivariate selection*: This class of methods refer to item selection by evaluating psychometric performance of items with respect to a given item pool. Although not commonly used, there are various types of multivariate methods for item selection that maximize psychometric performance of an item pool, e.g. homogeneity (Yousfi, 2003).

Application of various methods of item selection was to be applied, including univariate as well as multivariate methods. Initial analyses should offer analysis perspectives for the item pool at hand.

4.2.4 Item selection procedure

In a *first step,* an expert consensus was made about how to shorten the original measure to at least 50%. Although in their literature review Coste et al. (1997) reported total reduction rates in a range from 27% to 86% (median: 61%), the reviewed studies were not limited to short-form development in particular. Next, it was decided to select items in a "structurally representative" way. This means every domain and every facet of the original measure (DCGM-37) should be represented by the short-form measure equivalently with respect to the relative number of items

per facet. Consequently, the structural model of the DCGM-37 was referred to as a framework for developing the measurement model of the short-form.

As a consequence of these selection decisions, all item selection strategies leading to or resulting in a structural reorganisation ("decomposing" or "recomposing"; Funke, 2003) that was different from the original DISABKIDS chronic-generic measurement approach could not be taken into consideration. Therefore, other strategies usually applied in short-form development were not applicable. This especially concerned structurally explorative methods (e.g. explorative factor analysis, cluster centroid analysis, multidimensional scaling). The approach choosen here for short-form construction was explicitly adapted to the DISABKIDS chronic-generic measurement approach (not the other way round). Thus, there are several restrictions of item selection to ensure equivalent representation of all main content areas. Those conceptual restrictions had to be translated into specific conditions of item selection as applied to the respective (short) test construction procedures.

Although the short-form of the DCGM-37 was to include the main content domains of the original measure, it should not been constructed just with respect to univariate item characteristics exclusively. This is to say that only adding up best performing or most representing single items of each facet or developing a composite measure of more or less "independent" single items (so called "multi-item scales") does not suffice. Therefore, additional multivariate strategies of item selection were used, as those methods count on item performance relative to the performance of other items, thus also considering suppression effects between items (Yousfi, 2003).

Consequently, different possible solutions are outlined with respect to the need for psychometric performance of the resulting item pool: (I) identification of a strictly unidimensional measure, extracting one main factor from the item pool; (II) retaining factorial structure by simply shortening the item number per facet; (III) retaining higher-order structure on the domain-level by constructing three facets, each comprising items from both facets assigned to each domain of the original version; (IV) a combination of most represenive items from each facet ("marker" or "indicator" items); (V) identification of an item pool with equivalent numbers of items per facet by maximizing association (with the original measure in terms of correlation coefficient)); (VI) identification of an item pool with equivalent numbers of items per facet by maximizing homogeneity in terms internal consistency of the respective item pool.

Finally, different possibilities of the short-form's potential structure model were examined, taking into account the criteria necessary, psychometric properties provided by the original measure, as well as results from an initial screening of the inter-item correlaton matrix of the item pool.

With respect to the amount of concordance with the structure and measurement model or the conceptual model, respectively, item selection of the short-form measure can be based on one of the following three approaches:

- "Little sibling approach": taking the *original measurement and structure model as a basis* (original measurement approach as sine qua non)
- "Near relative approach": taking the *conceptual model for granted*, despite of the measurement model (e.g. item-facet assignment etc.)
- "Remote relative approach": taking the *item pool for granted*, despite of any model provided by the original measure. This approach leads to the creation of new models (based on rational considerations or empirical results).

4.2.5 A-priori estimations of descriptive properties and psychometric performance on item level and scale level (a priori validation)

A-priori validation of the measure included the estimation of various indicators of psychometric performance on basic item- and aggregated score level, i.e. basic descriptive statistics (localization parameters, distribution patterns), factor analysis (EFA, CFA), psychometric analysis in terms of classic test theory regarding reliability and different forms of validity as well as an investigation of Rasch scalability.

4.2.5.1 Basic descriptive and psychometric analyses

Items and composite scales were evaluated regarding missing data rate and basic descriptive distribution indicators, such as mean and standard deviation, skewness, curtosis, and floor and ceiling effects (defined by a critical value of 20% per lowest or highest response choice category). Additionally, psychometric performance on item level was investigated calculating item internal reliability (item-total correlation, part-whole corrected for overlap) and squared multiple correlation coefficient. The inter-item correlation matrix was used for detecting basic association structures within the item pool. In order to investigate dimensionality various kinds of factor analysis were performed, including different types of exploratory factor analysis as well as confirmatory factor analysis. In addition, although strong unidimensionality of the item pool was not expected for reasons of the initial item pool's heterogeneity, Rasch analysis was used to specifically cover possible misfit on item level. The partial credit model (Rost, 2004), an extension of the original Rasch model to ordinal variables, was applied to the data, using Q-index statistics and threshold ordering estimation for detecting item misfit. Moreover, item bias in terms of differential item functioning (DIF) was analyzed. As Zumbo points out (2007, p. 227), "DIF methods allow one to judge whether items (and ultimately the test they constitute) are functioning in the same manner in various groups ... In broad terms, this is a matter of measurement invariance; that is, is the test performing in the same manner for each group?"

4.2.5.2 Testing reliability

Homogeneity of the measure was identified by Cronbach's alpha coefficient α (Cronbach, 1951). Concordance between self-report and proxy report assessment was investigated using Pearson correlation coefficients (r) and intraclass correlation coefficients (ICC). Stability (approximative one-month interval) was defined as no change in child's general health as evident by the same response of the child to the question pertaining to self-perceived general health at both points of assessment (test and retest). Pearson correlation coefficient (r) and intraclass correlation coefficient (ICC) were also used for indicating estimated amount of stability (test-retest reliability).

Discriminatory power of a test, referring to the measurement's ability to distinguish between individuals, was assessed with the generalized delta coefficient δ_G (Hankins, 2007), which has been derived from the original delta coefficient δ from Ferguson (1949). As coefficient δ was computed for tests with dichotomous items, coefficient δ_G was derived for the more general class of questionnaires including several response choice options (polytomous items). The generalized delta coefficient was calculated separately for different subsamples of the data, regarding various test languages (6 subgroups; German and Austrian sub-samples merged).

4.2.5.3 Exploring validity

Content validity: Checking for content validity in terms of comparability with the original measure was done by expert consensus. In order to quanitify content validity in terms of content equivalence between short-form assessment and long-form assessment, ICF linkage codings of both measures were used. In absence of a gold standard, the idea behind this strategy was to quantify content comparability by an external criterion to which the respective measurement is related, instead of taking content validity for granted provided similar representation of all QOL facets (internal criterion).

Construct validity: Construct validity was ensured by applying the strucure model of the original-version DCGM-37 to the short-form measure. Thus, the standard for achieving sufficient construct validity of the short-form measure is to ensure comparability with the original version with respect to structure model, especially regarding the QOL dimensions included and the relative amount to which each QOL dimension is represented in both questionnaires.

Convergent validity: Validity estimations were performed, calculating Pearson correlation coefficients for association between the composite DCGM-12 score and scores of associated indicators of subjective health and other multidimensional HRQOL measures.

Divergent validity: In order to investigate divergent validity, correlations with constructs assumed to be not or negatively associated with HRQOL were calculated, e.g. emotional symptoms, conduct problems, or difficulties regarding peers.

Discriminant validity: One-way analyses of variance were performed to examine discriminant validity with respect to different socio-demographic (age group, gender), socio-economic (family wealth, parental education), and clinical (clinical global impression, severity of condition) variables known to be sensitive to differences in HRQOL. In addition, a-priori standardized effect sizes for differences of two independent means were estimated using Cohen's d (Cohen, 1992). Cohen's conventions concerning critical values for meaningful differences, namely small (.20), medium (.50), and large (.80) effects were used for interpretation, providing a standardized indication for the total amount of discriminant power. According to research literature on HRQOL and expert consensus from clinical professionals participating in the European DISABKIDS Group, expected direction and amount of mean score differences in HRQOL were defined (see Table 4): younger children (8-12 years) compared with older children (13-16 years) should show higher HRQOL scores (small effect size); boys compared with girls should show higher HRQOL scores (small effect size); children from families with high parental educational status compared with other children should show higher HRQOL scores (small to medium effect size); children from families with high economic status compared with other children should show higher HRQOL scores (small to medium effect size); children clinically rated as "normally developed" compared with children with delayed development should show higher HRQOL scores (medium to large effect size); children rated as "mild" or "moderate" regarding the clinical severity of their respective condition compared with children with "severe" conditions should show higher HRQOL scores (medium to large effect size).

4.2.5.4 Estimating clinical usefulness

Estimation of Cohen's d also allows for the evaluation of relative discriminant power of DCGM-12 score as compared to the CHQ-KINDL index score (Ellert et al., 2001). As the CHQ-KINDL index is a generic HRQOL measure adopted for use in healthy and ill children and adolescents, the measure provides a meaningful reference for evaluating the usefulness of a newly developed short-form by comparing effect sizes concerning discriminating between relevant groups. With respect to clicinal usefulness, both measures should discriminate sufficiently with respect to their respective target groups. Thus, relative differences in effect sizes for discrimination between various specific "known-goups" are of most interest, despite of the total degree of effect sizes.

Table 4: **Hypotheses about expected HRQOL differences according to various variables and hypotheses about expected effect size differences between assessments (DCGM-12 vs. CHQ-KINDL-Index) according to various variables (see above)**

Variable	*QOL differences between groups*	*ES differences between assessments*
Socio-demographic Variables		
• **Gender**	$QOL_{Boys} > QOL_{Girls}$	$ES_{DCGM-12} \leq ES_{CHQ-KINDL}$
• **Age group**	$QOL_{Younger\ (8-12)} > QOL_{Older\ (13-16)}$	$ES_{DCGM-12} \leq ES_{CHQ-KINDL}$
Socio-economic Variables		
• **Family Wealth (FAS)**	$QOL_{High} > QOL_{Medium} > QOL_{Low}$	$ES_{DCGM-12} \geq ES_{CHQ-KINDL}$
• **Parental Education Status**	$QOL_{High} > QOL_{Medium} > QOL_{Low}$	$ES_{DCGM-12} \geq ES_{CHQ-KINDL}$
Clinical Variables		
• **Clinical Global Impression**	$QOL_{Mild} > QOL_{Moderate} > QOL_{Severe}$	$ES_{DCGM-12} > ES_{CHQ-KINDL}$

Notes: QOL = Quality of life; ES = effect size.

4.2.5.5 Checking comparability

The same criteria regarding usefulness outlined above also allowed for the evaluation of relative discriminant power of DCGM-12 scores as compared to the original version of the DISABKIDS chronic generic module. Thus, this could be of use for covering some issues of comparability between both measures. Checking comparability regarding psychometric performance also included estimations of localization parameters, different types of reliability, and discriminatory power (utilization).

4.2.5.6 Overview

Table 5 provides an overview on various criteria and specific indicators used for estimating psychometric performance of the newly developed measure.

Table 5: Criteria, indicators and methods used for a-priori validation

Criterion	Description	Indicator
Basic descriptive and psychometric characteristics on item level		
• *Localization indicators*	Localization characteristics of items	M, Md
• *Distribution indicators*	Distribution pattern of items	SD, Var, Min, Max, Range
• *Response patterns*	Category frequencies, Response effects	fr, floor/ceiling
Internal structure of measurement		
• *Item-inter correlations*	Observing associations	r
• *Exploratory facor analysis*	Exploring dimensionality	PCA & PAF
• *Confirmatory factor analysis*	Confirming dimensional structure	ST / ST-SM (MLE)
Scaling analysis		
• *Basic scalabilty*	Checking converngence	Relibility analysis (CTT)
• *Rasch analysis*	Scalability	Partial Credit Model, DIF
• *Discriminatory Power*	Utitizaition of scale range	δ_G
Reliability		
• *Homogeneity*	Internal consistency	Cronbach's alpha
• *Split-half*	Assocaition of paired test-halfs	r, ICC
• *Test-retest*	Stabiity (Consistency, Specifity)	r, ICC
• *Self-Proxy-Aggreement*	Concordance between versions	r, ICC
Validity		
• *Construct validity*	Representation of structure model	Face validity
• *Criterion validity*	Association with original score	r
• *Content validity*	Representation of main content	Frequency analysis of items and ICF linkages
Relational performance		
• *Representation*	Main content areas of original	"Marker" items
• *Association*	Replication of total score	r, r^2
• *Comparability*	Content, Application, Performance	Comparative analysis
Further issues		
• *Economy*	Ratio precision lost/time saved, Test lowerance	Comparative analysis
• *Usefulness*	Uniqueness of measurement	Comparative evaluation

Notes: M = mean; Md = median; SD = standard deviation; Var = variance, Min = minimum; Max = maximum; fr = frequency; Floor/Ceiling = floor effect/ceiling effect; r = correlation coefficient; PCA = principal component analysis; PAF = principal axis analysis; ST = single-trait-model; ST-SM = single-trait-single-method-model; MLE = maximum-liklihood-estimation; CTT = classic (psychometric) test theory; DIF = differential item functioning; δ_G = generalized delta coefficient; ICC = intra-class correlation coefficient; r^2 = determination coefficient.

4.3 Software

Multivariate item selection was performed using SPSS© version 12 and 14 (for reasons of varying analyses occasions). SPSS© syntax has been used for programming various types of multivariate item selection (conditioned maximization of homogeneity and association). Also conventional item and scale statistics in terms of item-scaling analysis was performed using the SPSS© software package. Item fit statistics according to the (ordinal) partial credit model were computed using the demonstration version of the WINMIRA© program, version 2001 (von Davier, 1996). Analyses of differential item functioning (DIF) were realized using a SPSS syntax file provided by Zumbo (1999). Generalized delta coefficients ("Delta calculator" V.1.0; Hankins, 2007) were estimated using an EXCEL©-spreadsheet provided by Michael Hankins. A different EXCEL©-spreadsheet was used for computing effect-sizes (Cohen's d; "Effect size determination program"; Wilson, 2001). AMOS© 5 and AMOS© 16 evaluation version (for reasons of varying analyses occasions) were used for investigating confirmatory factor analysis (CFA). All further standard statistics were computed using SPSS©.

4.4 Data

4.4.1 Data samples

Data for empirical analyses and applications were collected within the scope of two studies within the above international project on constructing measurements assessing HRQOL in children and adolescents with chronic conditions (DISABKIDS). (I) The sample from the *cross-sectional DISABKDIS pilot study* (n = 404) and (II) the sample from the *longitudinal implemented DISABKIDS field study* (N = 1.603) were used for developing a generic, age specific, and multidimensional short-form (index) for the assessment of condition generic quality of life in children and adolescents with chronic diseases. The samples were used in particular for item extraction and short-form *a-priori* validation purposes, as the short-form of the DISABKIDS chronic generic module was not applied in a stud yet.

- The *DISABKIDS pilot study sample* involved 404 children and adolescents at the age range of 8 to 16 years. The responses of the proxy version (included in the questionnaire for parents) were available for 367 of all cases (> 90%). Further descriptions of both socio-demographic and clinical characteristics of the sample are given in Table 6.

Table 6: Selected socio-demographic and clinical characteristics of both DISABKIDS samples (DISABKIDS pilot study sample, N = 404; DISABKIDS field study sample, N = 1.153) *

Characteristic **		***Sample I*** ***DISABKIDS pilot study sample***		***Sample II*** ***DISABKIDS field study sample***	
		n	*%*	*n*	*%*
Child (test)		*404*	*100.00*	*1.153*	*100.00*
Age	*Years [M / SD]*	[12.45 / 2.64]		[12.23 / 2.76]	
Age group	*8 – 12 years*	*223*	*55.2*	*592*	*51.3*
	13 - 16 years	*181*	*44.8*	*525*	*45.5*
	(Missing)			*(36)*	*(3.1)*
Gender	*Female*	*185*	*45.8*	*547*	*47.4*
	Male	*208*	*51.2*	*590*	*51.2*
	(Missing)	*11*	*2.7*	*(16)*	*(1.4)*
Country	*Germany*	*113*	*28.0*	*279*	*24.2*
	Netherlands	*67*	*16.6*	*286*	*24.8*
	United Kingdom	*27*	*6.7*	*122*	*10.6*
	France	*96*	*23.8*	*74*	*6.4*
	Greece	*49*	*12.1*	*77*	*6.7*
	Sweden	*30*	*7.4*	*207*	*18.0*
	Austria	*22*	*5.4*	*108*	*9.4*
Main diagnosis	*Asthma*	*164*	*40.6*	*405*	*35.1*
	Arthritis	*54*	*13.4*	*150*	*13.0*
	Dermatitis	*40*	*9.9*	*66*	*5.7*
	Diabetes	*46*	*11.4*	*207*	*18.0*
	Cerebral Palsy	*20*	*5.0*	*91*	*7.9*
	Cystic Fibrosis	*26*	*6.4*	*43*	*3.7*
	Epilepsy	*54*	*13.4*	*191*	*16.6*
Adult proxy (test)					
Age	*Years [M / SD]*	[41.69 / 5.69]		[41.55 / 6.29]	
Respondent	*Mother*	*300*	*74.3*	*889*	*77.1*
	Father	*44*	*10.9*	*151*	*13.1*
	Other	*1*	*.2*	*13*	*1.1*
	(Missing)	*(59)*	*(14.6)*	*(100)*	*(8.7)*
Marital status	*Married*	*287*	*71.0*	*668*	*57.9*
Proxy rating on child					
Childs general health	*Range 1-5 [M / SD] **	[2.94 / 0.79]		[2.78 / 0.94]	
Childs development	*Above normal / Normal*	*243*	*60.1*	*870*	*76.5*
	Slow / Delayed	*68*	*16.8*	*164*	*14.2*
	(Missing)	*(93)*	*(23.0)*	*(118)*	*(10.2)*
Child's problems (range: 1-5)	*Physical [M / SD]*	[1.98 / 1.14]		[1.88 / 1.07]	
	Emotional/Behavioural [M/SD]	[1.77 / 0.91]		[1.88 / 1.08]	
	Social [M / SD]	[1.62 / 0.98]		[1.59 / 0.98]	
Clinicians rating ***					
Severity of condition	*Mild*	-	-	*373*	*32.4*
	Moderate	-	-	*299*	*25.9*
	Severe	-	-	*81*	*7.0*
	(Missing)	-	-	*(400)*	*(34.7)*

Notes: * Data representing the older age group (8-16 years) of the DISABKIDS samples only; younger age group (4-7 years) was excluded from analysis for both samples, because item pool application of the DISABKIDS chronic generic measure was limited to children and adolescents from 8 to 16 years. ** Data referring to frequencies and percent unless indicated by squared brackets. Absolute frequencies vary as a function of the amount of missing data for each variable. ******* Clinicians ratings were available only for the DISABKIDS field study sample.

- The *DISABKIDS field study sample* included 1.153 children and adolescents with different chronic conditions, aged between 8 and 16 years. The original total field study sample additionally also included 453 children at the age of 4 to 7 years. Questionnaires for parents were available in more than 92% of all cases. Retest data for the children were available for a sample of n = 464 (about 40% of all included cases), and retest data of the parents for a sample of n = 509 (44%). Clinicians' assessment was filled out for n = 909 (79%) cases. Socio-demographic and clinical characteristics of the sample are shown in Table 6 for the respective older age groups only (8 - 16 years).

Samples were assessed by different collaborating centres of the DISABKIDS project in seven European countries: Austria (Vienna), France (Marseille), Germany (Hamburg & Luebeck), Greece (Thesaloniki), Netherlands (Leiden), Sweden (Lund), and the United Kingdom (Edinburgh). Children and adolescents in the age-groups of 4-7, 8-12, and 13-16 years with different chronic conditions (asthma, cerebral palsy, diabetes mellitus, epilepsy, juvenile rheumatic arthritis, dermatitis, and cystic fibrosis) participated in the DISABKIDS project. Parents or other proxies were also included.

4.4.2 Data preparation

Data were ordered on first level according to respondents. On a second level, data were ordered by time. In consequence the data were ordered as follows accoring to the the respective source: child-test, child-retest, proxy-test, proxy-retest, clinician's rating. Columns represent cases (children/adolescents). Note that the sample is not fully matched regarding the country * condition matrix (see Table 7).

Since completeness of data cases varied casewise regarding availability of all main data sources (first and second level of data sampling), analyses were performed using filters for age at completion (8-16 years) and for sub-sample (i.e. country) split if required.. Sum or mean scores were computed if missing data rate of the respective measure reached less than 20%. Missing data were imputed by respective mean scores. The inclusion of (listwise) complete cases for computations is indicated separately, if applied.

Table 7: Cross table of the DISABKIDS field study according to conditions and countries (DISABKIDS field study sample, N = 1.153)

Country	Germany	Nether-lands	United Kingdom	France	Greece	Sweden	Austria
Disease	*n*	*n*	*n*	*n*	*n*	*n*	*n*
Asthma	42	133	50	38	38	75	30
Arthritis	88	62	-	-	-	-	-
Cerebral palsy	37	-	27	-	10	-	17
Cystic fibrosis	-	-	-	-	29	-	14
Diabetes	18	58	-	-	-	84	47
Epilepsy	44	33	45	21	-	48	-
Dermatitis	50	-	-	15	-	-	-

4.5 Instruments

Table 8 gives an overview of the methods employed in the field study. Some of the measures included are described in more detail within the following section.

4.5.1 DISABKIDS modules

The DISABKIDS condition related modules – including condition generic, condition specific, and smiley modules - were assessed for the child's test and retest as well as for the proxy's test. In addition, retest of children included the DISABKIDS Coping module (CODI; Petersen, 2004; Petersen et al., 2006). The take home questionnaire for the parents (proxies) included a DISABKIDS module assessing health care needs (CHC-SUN; Schmidt et al., 2007). Clinicians's data also included a module on severity assessment.

The *DISABKIDS condition generic module* (DCGM) assesses HRQOL in children and adolescents with different chronic health conditions from the perspective of children and adolescents, as well as their parents. The DISABKIDS condition generic instrument aims at being applicable in different national and cultural contexts. The interim version of the module consists of 56 Likert-scaled items that can be assigned to 6 dimensions: "Independence", "Limitation", "Inclusion", "Exclusion", "Emotion", and "Medication". Pilot test results showed satisfactory internal consistency of these scales, with alpha coefficients ranging from .79 ("Inclusion") to .90 ("Emotion"). The sub-scales can be combined to produce a total score for HRQOL. Intercorrelations between both "social" facets ("inclusion" and "exclusion") was high (r = .73) while the intercorrelations between the other scales were moderate (r = .20 - .70). The scales were developed to discriminate between chronic conditions with a more severe impact, such as cerebral palsy, and conditions that do not have such an impact on children's daily life. Both a self-report as well as a proxy version of the DISABKIDS instrument has been developed. The proxy-version is equivalent to the self-report version; item validity, composite score validity, and structural validity have been demonstrated (The European DISABKIDS Group, 2006).

A generic *DISABKIDS smiley version* ("TAKE 6"; Chaplin et al., 2008) has been developed and tested for very young children (age group 4-7 years), which is also applicable to older children (age group 8-16 years). The term "smiley" refers to the graphical distribution of the Likert-like rating scale, using different pictograms ("smiling faces"), each representing an emotional status one could have, ordered from "very sad" to "very happy". The smiley item wordings are formulated taking into account the age of the target respondent. The final version of the *smiley module* contains 6 items. Development process and psychometric properties of the measure are reported elsewhere (Chaplin et al., 2008).

Table 8: Overview of questionnaires and contents included in the DISABKIDS field study

Module	***Content***
(I) Children	**4 to 7 years**
• *Questionnaires for Children (4 to 7 years) -* ***T01***	• General questions • DISABKIDS smiley items
• ***Retest f****or Children (4 to 7 years) -* ***T02***	• General questions • DISABKIDS smiley questions
(II) Children	**8 to 16 years**
• *Questionnaire for Children (8 to 16 years) -* ***T01***	• Sociodemographic items • Anchor items • General health perception • DISABKIDS condition generic module • DISABKIDS smiley items • DISABKIDS disease specific modules • KIDSCREEN anchor items • Validation instruments (HRQOL), e.g. KINDL (different age versions: Kid-Kindl 8-12; Kiddo-Kindl 13-16)
• ***Retest*** *for Children & Adoles-cents (8 to 16 years) -* ***T02***	• Sociodemographic items • General health perception • Change in treatment • DISABKIDS condition generic module • DISABKIDS smiley items • DISABKIDS disease specific modules • DISABKIDS coping module
(III) Parent (Proxy)	**4 to 7 years (children)**
• *Parent questionnaire (4 to 7 years) -* ***T01***	• Sociodemographic items • Anchor items • General health perception • DISABKIDS smiley items • General clinical ratings • Disease specific parent clinical variables • Health status • Validation instruments (HRQOL), e.g. KINDL (different age versions: Kid-Kindl 8-12; Kiddo-Kindl 13-16) • Additional questions about the family
• ***Take Home*** *Parents (4 to 7 years) –* ***T02***	• Sociodemographic items • General health perception • Changes in treatment • Health care needs (Part 1: health services & Part 2: quality of services) • SDQ • SF-12: General (physical and mental) health status
(IV) Parent (Proxy)	**8 to 16 years (children)**
• *Parent questionnaire* • *(8 to 16 years) –* ***T01***	• Sociodemographic items • Anchor items • General health perception • DISABKIDS condition generic module • DISABKIDS disease specific modules • General clinical ratings • Disease specific parent clinical variables • Health status • Validation instruments (HRQOL), e.g. KINDL • Additional questions about the family

(Table continued)

• ***Take Home*** *Parents (4 to 7 years, 8 to 16 years) –* ***T02***	• Sociodemographic items • General health perception • Changes in treatment • Health care needs (Part 1: health services & Part 2: quality of services) • SDQ • SF-12: General (physical and mental) health status
(V) Physician	
• *Physicians*	• Medical documentation for clinical variables for all children • Disease specific clinical variables (7 modules)

DISABKIDS disease specific modules are core additional tools of the DISABKIDS modular approach, aiming to assess the more specific impact and burdens related to the diminishement in the quality of life resulting from living with a chronic health condition. Those modules were developed for all groups of chronic health conditions included in the DISABKIDS project design (i.e. asthma, arthritis, dermatitis, diabetes, cerebral palsy, cystic fibrosis, and epilepsy). Construction and validation of the modules has been reported elsewhere (Baars et al., 2005).

4.5.2 Clinical assessment

For *clinical assessment*, pediatricians assessed the primary ICD-10 diagnosis, as well as comorbid conditions. Children with severe mental disabilities were excluded from the study. Furthermore, clinicians were asked to diagnose the emotional, cognitive, social, and behavioural impairments as well as mental retardation on a categorical level in addition to disease specific clinical severity measures. These severity measures were developed within the European DISABKIDS project staff in order to differentiate between mild, moderate, and severe cases. These measures were derived from a list of disease specific clinical variables that were assessed in each condition. They were furthermore based on criterias and guidelines set forth by international research groups, such as the SCORAD for atopic dermatitis (European-Task-Force-on-Atopic-Dermatits, 1993).

4.5.3 Selected measurements for a-priori validation purposes

As noted above, for a-priori validation purposes, the *DISABKIDS field study* included various measurements assessing subjective health-status and HRQOL, reported by child or proxy. (More detailed information regarding the DISABKIDS field study assessment methodology in general and included questionnaires in particular is provided by the Manual of the European DISABKIDS Group, 2006).

Child's self-reported HRQOL was investigated using various already validated multidimensional measures which differing depending on the country of assessment: In Austria, Germany, and Greece the KINDL questionnaire (Ravens-Sieberer & Bullinger, 1998), in the Netherlands and in

Sweden the DUX-25 questionnaire (Koopman et al., 1998), in the United Kingdom the PedsQL questionnaire (Chan et al., 2005; Varni et al., 2003), and in France the VSP-A questionnaire (Simeoni et al., 2000; Sapin et al., 2005-b) were administered. Furthermore, the CHQ-KINDL index (Ellert et al., 2001), a short dynamic subjective health-status assessment derived from items originally assigned to the Child Health Questionnaire (CHQ: Landgraf et al., 1998) and the KINDL (Ravens-Sieberer & Bullinger, 1998) was applied. It assessed childrens' subjective health within all respective countries, ensuring a cross-cultural assessment perspective. Children and adolescents were also directly asked to report their self perceived health in general ("Child's General Health Perception"). In addition, non-health related assessments were included, amongst others basic socio-demographic information and socio-economic indicators ("Family Affluence Scale", FAS: Boyce et al., 2006; Currie et al., 1997).

For *proxy assessment* of the childrens' subjective health status and HRQOL, proxy-versions of the respective questionnaires used in the child assessment were administered (DUX-25, KINDL, PedsQL, VSP-A, CHQ-KINDL-Index; see above). Further measures related to childrens' health status included within the proxy assessment were two subscales of the *"Functional Status" questionnaire* (FS-II-R: Stein & Jessop, 1990), measuring "General Health" and "Interpersonal Functioning", as well as the *"Strengths and Difficulties" questionnaire* (SDQ; Goodmann, 1997). Parents were also assessed regarding basic socio-demographic and socio-economic variables.

Data obtained from *health professionals,* clinicians' general ratings of the *childrens' developmental status* ("normal" vs. "delayed/impaired"), and *severity of condition* ("mild"/"moderate", "severe") were included in data analyses.

5 Short-form construction and a-priori validation

5.1 Initial analysis

5.1.1 General introduction

The *objective* of the chapter is to document the development of an economic measure, based on the original multidimensional DISABKIDS condition generic measure (DCGM-37; see previous chapter). The short-form of the DCGM-37 should be strongly related to the respective original version in terms of conceptual representation, statistical association, and comparability regarding psychometric performance. Thus, the short-form should reproduce the main content areas of the original measure as closely as possible, using a selection of the original item pool without changes regarding wording and scaling. As compared to the long-form measure, it should provide an economic option for assessing HRQOL while applying the same measurement approach in the respective target population.

5.1.2 Item performance equivalence check across data samples

The data base for the following analyses is provided by two samples of children and adolescents with chronic health conditions (DISABKIDS pilot and field study sample), both including the 37 items of the final DCGM-37 version. However, both samples differ concerning some aspects of item application. (I) With respect to the *embeddedness* of items, several issues need to be mentioned here: (i) The pilot study sample includes more items within the DISABKIDS condition generic item pool. (ii) Compared to the pilot study, item sequence was changed in field study. (iii) Moreover, as items of the DCGM-37 are presented within item blocks referring to their respective item-scale assignment, some items are presented with different “block primers” (introductory words) in the pilot compared to the field study. (II) Regarding *item presentation*, the pilot study version provides an additional sixth “non-applicable” response choice option, indicating whether or not an item seems applicable.

5.1.2.1 Comparing pilot study and field study data sample

To account for variations in item application procedure, the potential occurance of substantial differential item performance across both data samples needs to be investigated. To examine item equivalence between both samples, difference values for selected item characteristics were computed: item mean scores and two coefficients of corrected item-total correlations for each item. Analyses were performed using t-test statistics for average mean score differences and tests of significance for differences in correlation coefficients between both samples (Bortz, 2004). In addition to sample size biased proof of significance of differences, critical values to

interpret absoulte degree of differences were included. Thus, referring to *a-priori* fixed critical values, more than 5% of a possible difference value range was interpreted as being a meaningful indicator for considerablel differences in item performance, 10% was interpreted as a large difference, and 20% as very a very large difference. Applying these criteria to respective item characteristics, critical values for absolute average *mean* difference scores (range = 4) are $\Delta_M >$ .20 (5%), $\Delta_M >$.40 (10%), and $\Delta_M >$ .80 (20%). Accordingly, critical values for absolute average difference scores for *correlation coefficients* (range = |1|) are $\Delta_{IIC} >$ .05 (5%), $\Delta_{IIC} >$.10 (10%), and $\Delta_{IIC} >$ .20 (20%). It should be noted that these critical values were defined a-priori and by the author.

Average mean differences reached more than .20 (5%) for 1/4 of the item pool (9 items), with 6 items reaching average mean difference values of more than .40 (10%). Most of these items belong to the "treatment/medication" subscale of the measure. With respect to a critical value of .05 (5%) nearly half of the items (18) displayed average differences for item total-correlation coefficients on total-score level. About 10 items displayed differences of more than .10 (10%). At the facet-level, with respect to a critical value of .05 (5%) more than 2/3 of all items (25) reached a higher differences and 15 items (about 40% of the item pool) displayed differences of about >.10 (10%). Results are presented in more detail in Table 9.

Due to extensive violation of critical value for differences, it was decided to split the field study sample. Equivalence checks of items diplayed sufficiently better values. Average mean score differences reached critical values of Δ **>** .20 (5%) for just 1 item. Although item total-correlation coefficients with respect to the total score reached values higher than critical value of about .05 for 13 items, just 1 item displayed differences > .10. Differences in item total-correlation coefficients at the facet level were obtained for 11 items (with a critical value of about .05) and just 3 items reached differences of more than .10. Average difference values of item performance within the DISABKIDS pilot study sample versus the DISABKIDS field study sample were above the critical value of 5% of the possible range for all three indicators.

Table 9: **Comparing differences between selected item characteristics of the 37 items of the final DCGM-37 using data from two sets of paired samples (DISABKIDS pilot study sample, N = 404; DISABKIDS field study sample, N = 1.153)**

Item No.	*Facet (n_{items})*	*t-test*			*DISABKIDS pilot sample (N = 406) vs. DISABKIDS field sample (N = 1.153)*	
DCGM-37		*mean difference*	*T*	*p*	Δ (r_{it} *total)*	Δ (r_{it} *facet)*
					(Δ x 100 = %)	(Δ x 100 = %)
01.	**Independence** (6 items)	- .064	- .988	.323	.01	**.12** ++
02.		.146	2.920	**.004**	**.11** ++	.04
03.		- .033	- .446	.655	.03	**.11** ++
04.		.039	.534	.594	**.07** +	.00
05.		**- .262** +	- 3.358	**.001**	**.36** +++	**.18** ++
06.		- .080	- 1.177	.240	**.26** +++	**.14** ++
07.	**Limitation** (6 items)	- .043	- .634	.526	**.07** +	**.08** +
08.		**.256** +	3.576	**.000**	.02	**.07** +
09.		.115	1.544	.123	.05	**.07** +
10.		- .143	- 1.631	.104	**.07** +	**.13** ++
11.		.094	1.553	.121	.05	.03
12.		.005	.070	.944	.03	.01
13.	**Emotion** (7 items)	- .027	- .437	.662	.04	**.06** +
14.		.088	1.246	.213	**.08** +	**.07** +
15.		- .061	- .891	.373	.04	.03
16.		- .026	- .360	.719	.04	.04
17.		- .139	- 2.161	**.031**	.01	.03
18.		- .221	- 3.185	**.002**	.05	.04
19.		- .073	- .874	.382	**.18** ++	**.12** ++
20.	**Exclusion** (6 items)	- .110	- 2.041	**.042**	.02	**.10** +
21.		- .110	- 1.718	.086	**.18** ++	**.06** +
22.		.099	1.424	.155	.05	.05
23.		- .129	- 2.466	**.014**	.05	.04
24.		- .057	- 1.012	.312	.01	**.08** +
25.		.012	.182	.855	.04	**.11** ++
26.	**Inclusion** (6 items)	- .121	- 1.550	.121	**.23** +++	**.11** ++
27.		**- .483** ++	- 5.984	**.000**	**.07** +	**.13** ++
28.		.157	2.877	**.004**	**.15** ++	**.14** ++
29.		.001	.019	.985	**.06** +	**.13** ++
30.		- .121	- 2.628	**.009**	.01	**.12** ++
31.		**- .230** +	- 2.557	**.011**	**.11** ++	.04
32.	**Medication** (6 items)	1.092	12.482	**.000**	.03	**.11** ++
33.		- .843	- 9.667	**.000**	**.07** +	.00
34.		.929	10.569	**.000**	**.36** +++	**.18** ++
35.		- .614	- 8.268	**.000**	**.26** +++	**.14** ++
36.		- .078	- .827	.408	**.07** +	**.08** +
37.		.528	5.890	**.000**	.02	**.07** +
Average Δ		**| .206 |**			**.091**	**.083**
		(>5%/.20)			**(>5%/.05)**	**(>5%/.05)**

Notes: T-test: T = T value; p = p-value; Δ r_{it} = total differences between coefficients of item-total correlations (corrected for overlap). Marked in **bold**: p-values < .05 (5%); Δ ICC > .05 (5%).

Table 10: **Comparing differences between selected descriptive item characteristics of the 37 items of the final DCGM-37 using data from two sub-samples (splitted DISABKIDS field study sample, N = 1.153, n_{I} = 575, n_{II} = 578)**

Item No.	***Facet (n_{items})***	***t-test***			***DISABKIDS field sub-sample I (n_{I} = 575) vs. DISABKIDS field sub-sample II (n_{II} = 578)***	
DCGM-37		*averge difference*	*T*	*p*	Δ (r_{it} *total)*	Δ (r_{it} *facet)*
01.	**Independence** (6 items)	.026	.421	.674	.00	.04
02.		.006	.104	.917	.00	**.09**
03.		.054	.816	.415	.01	.05
04.		.090	1.267	.206	**.06**	.05
05.		**.116**	1.752	.080	**.07**	**.07**
06.		.046	.759	.448	**.08**	**.06**
07.	**Limitation** (6 items)	.025	.381	.703	**.06**	.01
08.		- .017	- .240	.811	**.08**	**.06**
09.		.046	.638	.523	.01	.00
10.		- .070	- .874	.382	.05	.01
11.		.055	.871	.384	.03	.03
12.		.097	1.499	.134	.02	**.07**
13.	**Emotion** (7 items)	- .035	- .570	.569	.04	.05
14.		.011	.160	.873	.03	.05
15.		.026	.385	.700	.01	.02
16.		.004	.060	.952	.02	.02
17.		.030	.479	.632	.03	.02
18.		.068	1.123	.262	.01	.02
19.		.113	1.445	.149	.01	.00
20.	**Exclusion** (6 items)	.049	.966	.334	**.07**	.03
21.		- .020	- .333	.739	**.16**	**.23**
22.		- .033	- .492	.623	**.06**	**.11**
23.		.004	.076	.939	.03	.03
24.		.026	.468	.640	.03	.04
25.		.058	.904	.366	.03	.00
26.	**Inclusion** (6 items)	- .051	- .692	.489	**.07**	.03
27.		.086	1.079	.281	.00	.05
28.		- .003	- .054	.957	.04	**.07**
29.		.036	.568	.570	.03	.00
30.		.014	.316	.751	**.08**	**.07**
31.		**.159**	1.950	.051	.03	.01
32.	**Medication** (6 items)	.041	.585	.559	**.10**	.02
33.		- .032	- .348	.728	.03	.02
34.		.038	.580	.562	.04	.02
35.		.088	.972	.331	.00	.05
36.		.074	.786	.432	**.08**	**.06**
37.		**.289**	**4.004**	**.000**	**.10**	**.13**
Δ (average)		**|.034|**			**.043**	**.046**
		(<5%/.10)			**(<5%/.05)**	**(<5%/.05)**

Notes: T-test: T = T value; p = p-value; Δ r_{it} = total differences between coefficients of item-total correlations (corrected for overlap). Marked in **bold**: p-values < .05 (5%); Δ ICC > .05 (5%).

Comparisons of item performance between both subsamples of the DISABKIDS field study sample show that average differences of all three indicators are less than 5% with respect to the possible range of differences. When re-analyzing differences for alpha coefficients of all included facets between both paired samples, differences reached values ≥.05 (5%) for 5 of the 6 scales (Δ_{α} = 05 - .12), as compared to the other paired samples with just 1 item reaching a difference score above the critical value (.06).

Overall, these initial analyses of the pilot study data suggest limited usefulness of these data for developing the short-form of the DCGM-37, as average mean values and item-total correlations (strongly) differ between DISABKIDS pilot study sample and DISABKIDS field study sample for some items (see Table 9).

5.1.2.2 Selecting and comparing two sub-samples of field study data sample

According to the two steps of the analytic study design recommended in the literature on short-form development (Coste et al., 1997; Smith et al., 2000), data analyses should be performed based on two different samples. As a result of the item performance equivalence check across samples reported above, both samples needed for short-form development were derived from the longitudinal implemented DISABKIDS field study by splitting the data set into two sub-samples. The first sub-sample served as a data base for item extraction and "exploratory analysis" *("extraction sample";* n = 575). The second sub-sample of the DISABKIDS field study was used for a-priori validation purposes and "confirmatory analysis" in terms of cross-validation (*"a-priori validation sample";* n = 578), since the short-form of the DISAKIDS chonic generic module was not applied in an independent study.

Table 11: Selected socio-demographic and clinical characteristics of both sub-samples from the splitted DISABKIDS field study sample (N = 1.153) *

Characteristic **		***"Extraction sample"*** ***(DISABKIDS field study sub-sample I: $n_I = n_{ES}$)***		***"A-priori validation sample"*** ***(DISABKIDS field study sub-sample I: $n_{II} = n_{VS}$)***	
		n_{ES}	*%*	*n_{VS}*	*%*
		575	*100.00*	*578*	*100.00*
Child (test)					
Age	*Years [M / SD]*	[12.17 / 2.76]		[12.30 / 2.77]	
Age group	*8 – 12 years*	*278*	*48.4*	*275*	*47.6*
	13 - 16 years	*225*	*39.1*	*233*	*40.3*
	(Out of range + Missing)	*(36+36)*	*(12.5)*	*(30/40)*	*(12.1)*
Gender	*Female*	*272*	47.3	*275*	47.6
	Male	*293*	51.0	*297*	51.4
	(Missing)	*(10)*	*1.7*	*(6)*	1.0
Country	*Germany*	139	24.2	140	24.2
	Netherlands	143	24.9	143	24.7
	United Kingdom	61	10.6	61	10.6
	France	37	6.4	37	6.4
	Greece	38	6.6	39	6.7
	Sweden	103	17.9	104	18.0
	Austria	54	9.4	54	9.3
Main diagnosis	*Asthma*	203	35.3	203	35.1
	Arthritis	75	13.0	75	13.0
	Dermatitis	33	5.7	32	5.5
	Diabetes	104	18.1	103	17.8
	Cerebral Palsy	44	7.7	47	8.1
	Cystic Fibrosis	21	3.7	22	3.8
	Epilepsy	95	16.5	96	16.6
Adult proxy (test)					
Age	*Years [M / SD]*	[41.70 / 5.74]		[41.33 / 6.92]	
Respondent ***	*Mother*	394	*68.5*	383	*66.3*
	Father	63	*11.0*	71	*12.3*
	Other	5	*0.9*	5	*0.9*
	(Missing)	(113)	*(19.6)*	(119)	*(20.6)*
Partnership	*Live together with a partner*	417	*72.5*	426	*73.7*
Proxy rating on child					
Childs general health	*Range 1-5 [M / SD] **	[3.23 / 0.90]		[3.22 / 0.98]	
Childs development	*Above normal / Normal*	*441*	*76.7*	*439*	*76.0*
	Delayed	*82*	*14.3*	*82*	*14.2*
	(Missing)	*(52)*	*9.0*	*(57)*	*9.9*
Child's problems (range: 1-5)	*Physical [M / SD]*	[1.88 / 1.06]		[1.88 / 1.09]	
	Emotional [M/SD]	[1.92 / 1.11]		[1.84 / 1.04]	
	Social [M / SD]	[1.61 / 0.99]		[1.58 / 0.97]	
Clinicians rating					
Severity of condition	*Mild*	188	32.7	185	32.0
	Moderate	147	25.6	152	26.3
	Severe	40	7.0	41	7.1
	(Missing)	200	34.8	200	34.6

Notes: * Data representing the older age group (8-16 years) of the DISABKIDS samples only; younger age group (4-7 years) was excluded from analysis for both samples, because item pool application of the DISABKIDS chronic generic measure was limited to children and adolescents from 8 to 16 years. ** Data referring to frequencies and percent unless indicated by squared brackets. Absolute frequencies vary as a function of the amount of missing data for each variable. *** Sum of percent value reaches 100.1% resulting from rounding of single percent rates.

5.1.3 Initial analyses and evaluation of outlined strategies

5.1.3.1 Identification of a unidimensional measure

Applying principal component analysis (PCA) with the restriction to extract just one factor allows for evaluating items with respect to the ranking of factor loadings. Item rankings according to their factor loadings are displayed in Table 12. As can be seen in Table 12, items assigned to the "Emotion" facet are ranked highest (a ≥ .63). On the contrary, items assigned to the "Exclusion" and "Medication" facets ranked low (a ≤ .51). Thus, the item pool extracted with respect to items with highest loadings on a general factor (Eigenwert = 10.89, explained variance = 29.44) does not represent all main content areas of HRQOL of children and adolescents with chronic health conditions in terms of facets of the multidimensional original measure (DCGM-37). This result is stable if corrected for overrepresentation of items assigned to the "Emotion" facet in the item pool. Thus, applying principal component analysis for extracting just one factor violates the need for equivalent representation of all facets.

Table 12: Rankings for all DCGM-37 items according to factor loadings within principal component analysis restricted to one factor (sub-sample I from the DISABKIDS field study sample, n = 575)

IND			**LIM**			**EMO**			**EXCL**			**INCL**			**MED**		
No	*a*	*rank*	*No*	*a*	*rank*	*No*	*a*	*rank*	*No*	*a*	*rank*	*No*	*a*	*rank*	*No*	*a*	*Rank*
01.	.410	**31.**	07.	.468	**26.**	13.	.728	**02.**	20.	.635	**09.**	26.	.435	**29.**	32.	.361	**34.**
02.	.524	**18.**	08.	.496	**22.**	14.	.715	**03.**	21.	.329	**36.**	27.	.283	**37.**	33.	.460	**27.**
03.	.591	**12.**	09.	.584	**13.**	15.	.639	**08.**	22.	.534	**17.**	28.	.481	**23.**	34.	.439	**28.**
04.	.701	**04.**	10.	.543	**16.**	16.	.677	**07.**	23.	.478	**24.**	29.	.509	**21.**	35.	.512	**19.**
05.	.619	**11.**	11.	.512	**19.**	17.	.634	**10.**	24.	.557	**15.**	30.	.350	**35.**	36.	.469	**25.**
06.	.384	**33.**	12.	.571	**14.**	18.	.739	**01.**	25.	.680	**06.**	31.	.428	**30.**	37.	.406	**32.**
-	-	-	-	-	-	19.	.700	**05.**	-	-	-	-	-	-	-	-	-

Notes: Principle component analyses (PCA) of the DCGM-37 item pool restricted to extraction of one factor (Eigenwert = 10.89, explained variance = 29.44). No = item number according to the DCGM-37. a = factor loading. Rank = item rank according to factor loading.

5.1.3.2 Retain factorial structure by shortening item number per facet

Despite of other criteria of relevance, even if item selection is just driven by stepwise omitting with the lowest item-scale correlation for each facet to retain or improve homogeneity *("stepwise ommitting" approach*), some facets do not reach sufficient homogeneity in terms of Cronbach's alpha coefficient when compared to a critical value of $\alpha > .70$ ("Limitation" facet: $\alpha = .65$; "Inclusion" facet: $\alpha = .66$). As recommendations for meaningful cut-off scores for acceptable values of internal consistency in terms of Cronbach's alpha coefficient strongly vary in the respective literature (e.g. $\alpha > .50$ [Lienert & Ratz, 1994] to $\alpha > .80$ [Bühner, 2006]), we assume $\alpha > .70$ to be a meaningful a-priori value for indicating the lowest level of sufficient homogeneity. It should be noted here that this value is not suitable for individual diagnostic purposes. In addition, despite

of its total value coefficent, alpha indicates nothing more than a “lower bound” of reliability estimation, provided that non-correlated residuals can be assumed (Steyer & Eid, 1993).

Table 13: Cronbach’s alpha coefficients for all 6 facets of the DCGM-37 according to different numbers of items included (sub-sample I from the DISABKIDS field study sample, n = 575)

	IND	**LIM**	**EMO***	**EXCL**	**INCL**	**MED**
Items	α	α	α	α	α	α
	Stepwise lowerance (omitting item with lowest item-total correlation coeficient)					
6	.758	.721	.864	.716	.689	.782
5	.747	.712	.849	.736	.666	.775
4	.750	.695	.828	.751	.669	.780
3	.766	.648	.794	.706	.663	.793

Notes: * Initial coefficient (7 items) = .875. α = Cronbach’s alpha coefficient.

To account for possible suppression effects in item pools (Yousfi, 2003), due to biasing item selection within “stepwise omitting” approach, permutations of all possible combinations for 3 out of 6 items per facet (except for 3 out of 7 items of the “emotion” facet) were computed (“permutation approach”). The interpretation of results from the “stepwise omitting approach” seems to be confirmed by the results from this “permutation approach” in general, although coefficents of final 3-item solutions for some facets vary to some degree.

In general, with respect to our results, applying these strategies does not seem warranted given the desired ratio of precision and meaningful reduction of test length.

5.1.3.3 Retain conceptual higher-order structure on domain-level

To retain higher-order structure by constructing three facets, comprising items from both facets assigned to each domain of the original measure does not seem feasable for at least three reasons:

- *Limited comparability*: Basically, there are no guidelines for scoring or reference scores at the domain-level for the original version (DCGM-37). Thus, comparability with respect to mode of administration and scoring issues is lacking.
- *Practical limitations:* A problem in applying this approach to respondents’ scores lies within the constituion of the “Physical” domain, which comprises items derived from the “Limitation” as well as from the “Medication” facet. As “Medication” items are only applicable in case of the respective child getting medication, a potentialially high rate of non-applicable items is to be expected. Pilot and field study results of the DISABKIDS project indicate an average non-aplicable rate of items assigned to the “Medication” facet of approximative 15%. With respect

to equivalent representation of all domains and facets in the short-form measure, a facet constructed at the domain level is thought to be comprised by 50% of items derived from the original "Limitation" facet and 50% of items derived from the original "Medication" facet. Thus, for all children not getting any medication, this approach could not be applied in a meaningful manner without bringing about an impairment in the critical value of an acceptable amount of missing data, as for all these cases it will definitely reach 50%.

- *Impaired robustness of domains*: The conceptual model at the domain level is not supported by empirical data, as indicated by various parameters and coefficients (e.g. intercorrelation between manifest facet scores). Except for the "Inclusion" facet, other facets display highest correlation coefficients with the "Emotion" facet score. Thus, except for the "Independence" facet score, other facet scores display higher levels of assocaitions in terms of correlation coefficients with other facets than their respective paired facet score referring to domain level. At least for the "Emotion" facet score this holds to be true, as the "Emotion" score is correlated stronger with the "Exclusion" facet score than with its "paired" facet "Independence". Thus, intercorrelations between facet scores of the original version display discrepancies between the conceptual model and empirical data with respect to facet-domain assignment. A more in-depth view on the basic internal structure is provided by inspection of the intercorrelation matrix at the item level.

Table 14: **Coefficients for intercorrelations (Pearson's r) between facet scores (TMS) of the self-report version DISABKIDS chronic generic module (sub-sample I from the DISABKIDS field study sample, n = 575**)

Domain	*Facet*	IND	EMO	EXCL	INCL	LIM	MED
Mental	**IND**	*[1]*	**.638**	.576	.565	.621	.345
	EMO	.638	*[3]*	**.696**	.459	.692	.538
Social	**EXCL**	.576	**.696**	*[4]*	.517	**.623**	.398
	INCL	**.565**	.459	.517	*[2]*	.478	.319
Physical	**LIM**	.621	**.692**	.623	.478	*[5]*	.405
	MED	.345	**.538**	.398	.319	.405	*[2]*

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". Highest correlation coefficients for each of the analysed facet scores are marked in **bold**. Coefficients on grey marked sqared table fields indicating intercorrelations between assigned paired facet scores, according to their respective common domain. Numbers in brackets on grey marked sqared table fields indicating rank of these correlation coefficients for each facet as compared to correlation coefficients with all other facet scores.

5.1.3.4 Combination of "marker" items from each facet

Another approach is based on selecting "marker" items per facet. "Marker items (sometimes also referred to as "indicator" items) are items with the highest ranked coefficients for item-total correlation concerning each facet. The main advantage of this approach is that it accounts for including the most representative items of each facet. The main disadvantage lies in the problem of heterogeneity, as divergent validity between "marker items" from different facets of a multidimensional measure should be and is usually higher than between different items from one facet. The marker item approach could pose a problem for the construction of a consistent item pool. Computations for marker item solutions with respect to three initial test length versions suggest the 12-item version to be most efficient. For our purposes, efficiency is definied by a meaningful balance between test length and psychometric performance (see Table 15).

Table 15: Cronbach's alpha and (squared) correlation coefficient for all 3 test version lengths of the DCGM short-form measure at different levels of measuremeht approach (sub-sample I from the DISABKIDS field study sample, n = 575)

Level	***facet***			***domain***			***total***		
Length	α	*r*	R^2	α	*r*	R^2	α	*r*	R^2
6	.650	.900	.808	.667	.896	.803	.716	.882	.778
12	.823	.948	.895	.817	.942	.888	.832	.943	.886
18	.817	.963	.924	.811	.965	.930	.877	.977	.952

Notes: α = Cronbach's Alpha; $r_{sf,ov}$ = Association between respective DCGM total score/item pool (6, 12, 18 items) and DCGM-37 total score (Pearson correlation coefficient, not corrected for overlap); $R^2_{sf,ov}$ = determination coefficient for prediction of DCGM-37 average score by respective DCGM total score/item pool (6, 12, 18 items).

5.1.4 Summary of results from initial analysis and decision on item selection strategy

As none of the furthermentioned procedures served to be fully sufficient with respect to the criteria defined for a short-form development, expert consensus within the DISABKIDS Group decided on choosing a combined approach of two further multivariate options:

Identification of an item pool with equivalent number of items per facet
by maximizing association

As improving (statistical) association with the original measure (DCGM-37) is one of the main objectives of this short-form development, we investigated association coefficients by applying a multivariate approach to the item pool for all three versions of initially possible test length (6,12,18 items) by using regression analysis. Determination coefficient served as indicator of association between item pool and total score of the original measure.

Identification of an item pool with equivalent number of items per facet by maximizing homogeneity

As apparent by the previous analyses, short-form construction would not be meaningful if solely focused on maximizing economic reproduction of the original measure. The process of item selection was guided by the psychometric performance of the measure itself. Thus, ensuring homogeneity is a main purpose within this short-form development, in parallel to aggregating a multidimensional item pool into a consistent measure. Again, applying a multivariate approach seems to be most fruitfull for our purposes.

The procedure choosen was applied to a total of 12 items, i.e. 2 items for each facet of the original measure represented within the to-be-constructed short-form. As compared to other options this 12-item version alternative was assumed to provide best efficiency in terms of providing an optimal balance of economic and informational criteria. This decision was based on results from initial analyses in terms of multivariate item selection strategies for the proposed three general test length versions (6, 12, 18 items).

- The first multivariate approach constitutes a restricted total permutation for *maximizing homogeneity*, under specifyied conditions for the three versions: (1) 6-item version: select 1 item from each facet; (2) 12-item version: select 2 items from each facet; (3) 18-item version: select 3 items from each facet.
- The second multivariate approach aimed at *maximizing association* using regression analyses, under specified conditions for the three versions: (1) 6-item version: select 1 item from each facet; (2) 12-item version: select 2 items from each facet; (3) 18-item version: select 3 items from each facet.

Results of apllying both approaches to the given data sample provide two points of general interest:

- First, the best performing item pools for both approaches do not converge. Yousfi (2003) ran simulation studies investigating variation of psychometric performance (reliability, validity) as a function of different multivariate methods of item selection (test construction) and observed effects similar to ours.
- Second, the (best) 6-item versions reached satisfactory homogeneity values but insufficient association with the total score of the original version. Comparing 12-item versions and 18-item versions, slight increases of homogeneity (for both) and association (for the 18-item version) does not justify an increase in test length by 50% as this would substantially reduce the test's economy and efficiency. These results confirm the results from the initial analyses using marker items for estimation of homogeneity and association (see above).

As none of the outlined strategies proved fully sufficient for our purposes – namely to ensure representation, association, comparability, and homogeneity by simultaneously ensuring economy

in terms of substantial loss of test length – conditions for applying this strategy to the respective item pool had to be redefined. For the selection of items which best representing each facet of the original assessment, inclusion criteria were defined with respect to different possible short-form lengths. Specific criteria aimed at including items with highest item-total correlation coefficients on facet-level. Inspection of item-total correlation coefficients shows that "marker" items share more variance with the item pool than other items from the respective facet. However, this trend is inconsistent across facets, as e.g. evident for the "marker" item from the "Inclusion" facet. Thus, the following criteria were defined:

- Test reduction should be at least 50%. Thus, with respect to the criterion of symmetric representation, a maximum of 3 items is to be extracted.
- Three test lengths are optional: (1) 6-item version: select 1 item from each facet; (2) 12-item version: select 2 items from each facet; (3) 18-item version: select 3 items from each facet.
- To account for the representation of a facet by its most representive items, it was decided to add the following selection criteria as further conditions to the general approach. (1) 6-item version: One of the first two most representive items (as indicated by item-total correlations) must be included. (2) 12-item version: Both items derived from each facet should range amongst the first four of the most representative items, but inclusion of the first "marker" item of each respective facet is mandatory. (3) 18-item version: All three items derived from each facet should range amongst the first five of the most representing items, but inclusion of both highest ranked items of each respective facet according to item-total correlation coefficients is mandatory.

To ensure sufficient representation, association, comparability, and homogeneity, the following principles were applied:

- Ensuring a lower bound of sufficient *comparability* by restricting potencially included items to stem from the original item pool without any modification, as well as using the same scoring procedure.
- Ensuring a lower bound of sufficient *representation* by specifying conditions for item selection using univariate indicators (corrected item-total correlations at the facet level).
- Ensuring a lower bound of sufficient *association* using multivariate item selection methods directed to maximize common variance (multiple regression modelling).
- Ensuring a lower bound of sufficient *homogeneity* using multivariate item selection methods directed to maximize internal consistency.

At the univariate item level, initial coefficients for corrected item-total correlations of selected items should rank among the first three of each respective factor. This ensures that most representative items for each facet are selected. The first three ranks are allowed to vary, as not all

items ranked to be the first two per facet are at the same time also ranked as first two per total score. Fixing items at the top range would not have made sense with respect to the prerequisites of running permutations. As average coefficients of item-total correlations strongly vary between facets, critical values in terms of total numbers are less suitable. As a minimum requirement, however, absolute values of item-total-correlation coefficients of about .40 (minimum) are needed.

As the DCGM-37 measure is multidimensional, the core challenge of this investigation is to construct a short measure which displays sufficient homogeneity and at the same time represents the variety of content by including the marker item of various facets. Overall, balancing of internal and external related performance criteria is the task. According to that, univariate and multivariate methods were used for extracting a final item pool of the short-form.

With respect to a-priori estimated psychometric performance indicators, the following critical values for sufficient short-form performance were formulated:

- At the “univariate” item level, “final” corrected item-total correlations should reach a value of at least .40. According to (expert consensus from) other projects (e.g. EUROHIS; Power, 2003) items with smaller values can not be assumed to be representing the underlying construct sufficiently,.
- At the “multivariate” composite score level, the item pool should have a satisfactory internal consistency ($\alpha > .80$) to ensure a minimum of required homogeneity, even if items were originally assigned to different facets of the HRQOL.
- At the “multivariate” composite score level, the extracted set of items should explain about 90% of DCGM-37 total score variance, in line with other studies of this type (e.g. VSP-A/S: Sapin et al., 2005-a; SF-12/36: Ware et al., 1996),.

There are numerous critical issues regarding each of these criteria; the most important are shortly denoted in the following section:

- Selecting items with respect to their item-total correlation is commonly used in test construction, test refinement, and short-form development (Coste et al., 1997). Nevertheless, this procedure is neither derived nor justified by psychometric test theory (Yousfi, 2007).
- Ensuring or improving homogeneity is frequently applied in test refinement, even though it poses several problems and potential pitfalls. For instance, when maximizing homogeneity without reflecting item content in terms of validity, homogeneity or consistency tends to generate redundancy. In addition, internal consistency is a weak indicator of measurement reliability (Steyer & Eid, 1993).
- Computing association between long-form and short-form scores not corrected for common error is psychometrically insufficient. While Coste et al. (1997) argued in line with this

criticism, Smith et al. (2000) recommend its use despite of these problems, due to pragmatic reasons as there truly is no meaningful alternative at hand.

Therefore further criteria guided our approach: compute different strategies on item selection simultaneously and compare the resulting preffered item pool solutions; final decision on item pool solution should be based on expert consensus according to empirical and rationale issues.

The initial item pool for item derivation came from the sub-sample (age group 8-16 years) derived from the DISABKIDS field study sample (n = 575). This item pool includes all 37 Likert-scaled items from the final DISABKIDS condition generic module. Items have five response choices with shared response choice wordings. Multivariate methods of item selection were applied (conditioned total permutation for maximizing item pool homogeneity and iterative application of regression models for calculating prediction of the DCGM-37 total score). Coefficient were estimated for all possible item combinations restricted to a total of 12 items, and for each given 2 items per original facet. Consequently, a number of 25 items were deleted in the process, i.e. more than two-thirds (67.6%) of the original item pool. The final item pool was selected after expert consensus about best performing possible solutions with respect to the criteria mentioned above. The set of items of the children's self-report version served as a template for the proxy-version.

Table 16: **Pattern of corrected item-total correlations of the DCGM-37 item pool on global score level, domain level, and facet level (sub-sample I from the DISABKIDS field study sample, n = 575**)

	r_{it}	rank	r_{it}	rank	r_{it}	rank	r_{if}	rank
Level	**37**		**total**		**domain**		**facet**	
	403							
01.	.379	33.	.379	05.	.356	13	.398	06.
02.	.485	18.	.485	04.	.634	3	.402	05.
03.	.557	12.	.557	03.	.612	7	.584	02.
04.	**.658**	**05.**	**.658**	**01.**	.689	**1**	**.580**	**03.**
05.	**.582**	**10.**	**.582**	**02.**	.471	**11**	**.609**	**01.**
06.	.348	34.	.348	06.	.415	12	.420	04.
IND	**.502**	**18.67**	**.502**				*(.758)*	*5204.9*
07.	.424	28.	.424	06.	.291	12	.460	04.
08.	.457	22.	.457	04.	.460	6	.464	03.
09.	**.523**	**14.**	**.523**	**02.**	.448	**8**	**.524**	**01.**
10.	.517	15.	.517	03.	.403	10	.375	06.
11.	.457	23.	.457	04.	.312	11	.401	05.
12.	**.525**	**13.**	**.525**	**01.**	.480	**5**	**.520**	**02.**
PHY	**.484**	**19.17**	**.484**				*(.721)*	*537*
13.	.689	01.	.689	01.	.606	8	.649	05.
14.	**.666**	**04.**	**.666**	**04.**	.611	**6**	**.697**	**02.**
15.	.576	11.	.576	07.	.604	9	.654	04.
16.	.633	07.	.633	05.	.622	5	.661	03.
17.	.585	08.	.585	06.	.560	10	.617	06.
18.	**.683**	**02.**	**.683**	**02.**	.668	**2**	**.708**	**01.**
19.	.669	03.	.669	03.	.627	4	.617	06.
EMO	**.643**	**5.14**	**.643**				*(.875)*	*533*
20.	**.582**	**09.**	**.582**	**02.**	.587	**1**	**.549**	**02.**
21.	.293	36.	.293	06.	.437	8	.269	06.
22.	.488	17.	.488	04.	.436	9	.382	05.
23.	.429	27.	.429	05.	.505	6	.507	03.
24.	.511	16.	.511	03.	.447	7	.490	04.
25.	**.634**	**06.**	**.634**	**01.**	.572	**2**	**.559**	**01.**
EXCL	**.490**	**18.50**	**.490**				*(.716)*	*533*
26.	.414	29.	.414	03.	.433	10	.379	06.
27.	.262	37.	.262	06.	.380	11	.399	05.
28.	.444	24.	.444	02.	.529	4	.461	03.
29.	**.482**	**19.**	**.482**	**01.**	.516	**5**	**.485**	**02.**
30.	**.324**	**35.**	**.324**	**05.**	.559	**3**	**.487**	**01.**
31.	.409	30.	.409	04.	.379	12	.406	04.
INCL	**.389**	**29.00**	**.389**				*(.689)*	*526*
32.	.431	26.	.431	04.	.409	9	.410	06.
33.	.433	25.	.433	03.	.550	4	.600	03.
34.	.481	20.	.481	01.	.457	7	.424	05.
35.	**.468**	**21.**	**.468**	**02.**	.610	**1**	**.681**	**01.**
36.	**.406**	**31.**	**.406**	**05.**	.592	**2**	**.620**	**02.**
37.	.393	32.	.393	06.	.553	3	.464	04.
MED	**.435**	**25.83**	**.435**				*(.782)*	*470*

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". r_{it} = Item-total correlation (corrected for overlap)

5.2 Test construction – Item selection

5.2.1 Defining the initial item pool

The finite item pool was selected using the second sub-sample of the DISABKIDS field study sample (n = 578). Although the DISABKIDS pilot study sample was originally selected for a-priori validation, initial analysis indicated insufficient usefulness of these data for that purpose, as outlined before (see previous section).

5.2.2 Description of the initial item pool

The initial item pool for the construction of the DISABKIDS index included 37 Likert-scaled items with five response choices. Data were available for a total of 578 persons (no listwise complete cases). After adjusting missing data rate for items derived from the "medication" facet for systematic overestimation, missing values were moderate (< 5.00 %) for all chronic generic DISABKIDS items. Mean values for most items ranged between 3.5 and 4.5 (SD = .74 to 1.47), 2 items reached a mean score below 3.5 (M = 3.38, SD = 1.19/ M = 3.38, SD = 1.41), whereas 3 items, all belonging to the "Social" domain, reached mean scores of more than 4.5 (M_{max} = 4.57). One item was strongly skewed with – 2.24; curtosis value of the respective item reached 4.84.

5.2.3 Effects of non-balanced item pool

Estimated parameters can be expected to be biased as a result of non-balanced item-facet assignment in terms of number of items selected per facet (instead of resulting from the "true" correlations between items). In order to estimate the impact of variance on the internal structure of the measurement model resulting from imbalanced overrepresentation of 7 items assigned to the "Emotion" facet as compared to all other facets represented within the item pool by 6 items, item-total correlations were computed for every possible total balanced case with respect to item number per facet (see appendix for details). Thus, 7 item pool versions with a total number of 36 items were computed. Each version was constituted by excluding another item from the "Emotion" subscale (see Table 17). Difference scores for item-total correlations and average item-total-correlations per facet as well as ranges for all item-total correlation ranks for every item within each inspected item pool were computed, also. These indicators allow for health professionals were included a-priori estimation of non-balanced item-facet assignments effects as compared to a balanced design. Thus, the main objective is an estimation of equalization effects resulting from balancing the item pool.

Table 17: **Pattern of corrected item-total correlations of the DCGM-37 and "DCGM-36" item pool on global score level (sub-sample I from the DISABKIDS field study sample, n = 575)**

	DCGM-37		"DCGM-36" * (7 versions)			Range
	37		IIC	ICC		
	403	rank	Range	Δ (min, max)	M	rank(s)
01.	.379	33.	.375-.380	.005		32. (33.)
02.	.485	18.	.481-.487	.006		16.-19. (18.)
03.	.557	12.	.558-.560	.002		11. (12.)
04.	.658	05.	.653-.659	.006		04.-05. (05.)
05.	.582	10.	.579-.586	.007		07.-09. (10.)
06.	.348	34.	.345-.353	.008	.006	33. (34.)
IND	**M = .502 (2)**	**M = 19.**	***.500-.503***	***.003***	**M = *.502 (2)***	**M = *18.***
07.	.424	28.	.425-.430	.005		25.-17.
08.	.457	22.	.452-.461	.009		21.-22.
09.	.523	14.	.518-.524	.006		13.
10.	.517	15.	.513-.517	.004		14.-15.
11.	.457	23.	.449-.459	.010		21.-23.
12.	.525	13.	.520-.526	.006	.007	12.
LIM	**M = .484 (4)**	**M = 19.**	***.482-.484***	***.002***	**M = *.483 (4)***	**M = *18.***
13.	.689	01.	.683-.690	.007		01. (01.)
14.	.666	04.	.656-.663	.007		03.-05. (04.)
15.	.576	11.	.563-.573	.010		10. (11.)
16.	.633	07.	.626-.633	.007		05.-07. (07.)
17.	.585	08.	.578-.585	.007		07.-10. (08.)
18.	.683	02.	.672-.679	.007		01.-02. (02.)
19.	.669	03.	.664-.671	.007	.007	02.-03. (03.)
EMO	**M = .643 (1)**	**M = 05.**	***.632-.649***	***.017***	**M = *.640 (1)***	**M = 06.**
20.	.582	09.	.575-.583	.008		08.-09.
21.	.293	36.	.289-.297	.008		35.
22.	.488	17.	.477-.490	.013		16.-19.
23.	.429	27.	.424-.433	.009		25.-27.
24.	.511	16.	.506-.516	.010		14.-15.
25.	.634	06.	.631-.634	.003	.009	05.-06.
EXCL	**M = .490 (3)**	**M = 19.**	***.486-.489***	***.003***	**M = *.488 (3)***	**M = 18.**
26.	.414	29.	.411-.418	.007		28.-29.
27.	.262	37.	.254-.271	.017		36.
28.	.444	24.	.443-.454	.011		21.-23.
29.	.482	19.	.484-.491	.007		16.-18.
30.	.324	35.	.324-.331	.007		34.
31.	.409	30.	.406-.416	.010	.010	28.-30.
INCL	**M = .389 (6)**	**M = 29.**	***.390-.394***	***.004***	**M = *.392 (6)***	**M = 28.**
32.	.431	26.	.422-.436	.014		24.-27. (26.)
33.	.433	25.	.432-.435	.003		24.-27. (25.)
34.	.481	20.	.474-.484	.010		18.-19. (20.)
35.	.468	21.	.467-.471	.004		20. (21.)
36.	.406	31.	.400-.412	.012		29.-30. (31.)
37.	.393	32.	.389-.398	.009	.009	31. (32.)
MED	**M = .435 (5)**	**M = 26.**	***.435-.438***	***.003***	**M = *.436 (5)***	**M = 26.**

Notes: * "DCGM-36" refers to the item pool of the DCGM-37, excluding 1 item of the "emotion" facet (7 versions). *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". M = mean; ICC = Intraclass correlation coefficient (one-way, absolute agreement); Δ (min, max) d= Difference between lowest and highest ranked coefficients.

Effects of balancing item numbers per facet over the whole item pool resulted in equal effects with respect to the degree of item-total correlations, the differences in item-total correlations, and the ranks of item-total correlations. Thus, also after balancing the “Emotion” facet remained the “marker” facet of the item pool. Irrespective of which specific item was excluded for each 36-item “version”, all remaining items of the “Emotion” facet ranked at least within the top 11 items, with 5 out of them being consistently amongst the top 7, and 4 of them always being amongst the top 5 (see Table 17).

5.2.4 Item selection

Computing various item selection strategies resulted in different preffered item pool solutions recommended for constructing the short-form. The final item pool was extracted amongst other possible solutions, fulfilling the criteria mentioned above at the composite scale level. Nevertheless, no possible item pool could be identified without at least one item violating the prerequisite of an item-total correlation not being smaller than .40. Thus, an item pool with few violations of this criterion was selected for an a-priori estimation of psychometric performance. Final decision on item pool extracted for the short-form version was done by expert-consensus. In the final round, if items displayed weak perfomance despite of their relevance, in case of doubts it was decided to keep these items within the pool. This is in line with recommendations in the respective literature (e.g. Bühner, 2006; for a more methodologically rational see Yousfi, 2007).

5.3 Test performance – A-priori estimations

A-priori estimations of descriptive and psychometric properties at the item and scale level – A *note*: Analyses reported are based on combined data, including both sub-samples. Results for different issues of test performance are each reported separately for the self-report and proxy-report version. As calculations were not performed based on an independent validation sample, all descriptive and psychometric properties are *a-priori estimations* of the short-form’s performance (Coste et al., 1997; Smith et al., 2000).

5.3.1 Analysis on item level

Initial analyses started on basic item level, and considered single performance of each item regarding parameters of locality and distribution, response patterns, missing data rate etc.

5.3.1.1 Basic descriptive analysis on item level I: Response choice frequencies

Self-report version (see Table 18): Response choice frequencies indicated skewed item distributions in general. Non-ordered response choice frequencies according to a linear/smoothed distribution pattern of frequencies appeared for the two items derived from the "medication/treatment" facet indicating "collapsing" distributions (see also Figure 1). Except for item 11 ("Does taking medication bother you?") and item 12 ("Do you hate taking your medicine?"), both derived from the "treatment/medication" facet with a missing data rate of about nearly 15%, missing value rate was satisfyingly low (< 5.00%) for the remaining 10 items. Nevertheless, missing data rates for item 11 and 12 are systematically overestimated, due to inclusion of respondents without any medical treatment; adjusted missing data rates amount to 0.3% (n = 2) for item 11 and 0.5% (n = 3) for item 12.

Table 18: A-priori estimations for selected descriptive item characteristics I of the DISABKIDS chronic generic short-form measure (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

	DCGM-37		N_{vc}	**Missing**	**1**	**2**	**3**	**4**	**5**
					never	*seldom*	*quite often*	*very often*	*always*
	Facet	*No.*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*
01	IND	04	563	15 (2.6)	6.1	6.7	15.4	25.6	43.6
02	IND	05	563	15 (2.6)	**4.2**	6.6	16.4	26.8	43.4
03	EMO ‡	14	564	14 (2.4)	**4.2**	6.9	17.5	23.9	45.2
04	EMO‡	18	560	18 (3.1)	**2.2**	**3.8**	14.7	26.1	50.0
05	EXCL ‡	20	559	19 (3.3)	**1.7**	**1.7**	8.5	12.1	**72.7**
06	EXCL ‡	25	562	16 (2.8)	**4.0**	**4.0**	17.0	22.3	50.0
07	INCL	29	563	15 (2.6)	**3.1**	5.7	15.1	26.8	46.7
08	INCL	30	560	18 (3.1)	**0.5**	**2.1**	7.1	24.4	**62.8**
09	LIM ‡	09	562	16 (2.8)	5.5	9.2	18.7	22.1	41.7
10	LIM ‡	12	563	15 (2.6)	**2.8**	6.6	24.6	23.2	40.3
11	MED ‡	35	493	***85 (14.7)*****	12.3	8.1	18.5	12.1	34.3
12	MED ‡	36	492	***86 (14.9)*****	15.6	5.2	15.9	15.4	33.0

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: missing data rate > 5%; response rates > 50% / < 5%. N_{VC}: valid cases; ‡ Reversed scored items (recoded before analysis). * Missing data rates are systematically overestimated, as they include the non-applicable rates of respondents without any medical treatment; adjusted missing data rates reaching n = 2 (0.3%) for item 11/35 and n = 3 (0.5%) for item 12/36.

Proxy-report version (see Table 19): Also for the proxy-version, similar non-ordered response choice frequencies according to a linear/smoothed distribution pattern of frequencies appeared for the two items derived from the "medication" subscale, indicating "collapsing" distributions (see also Figure 2). Again, items 11 ("Does taking medication bother your child?") and 12 ("Does your child hate taking his/her medicine?") displayed high missing data rates of nearly 15%. As for the self-report version, these missing values rates were systematically overestimated; here, adjusted missing data rates amount to 0.6% (n = 3) for both items. As compared to the self-report version, all other items displayed slightly higher, but still acceptable missing data rates (5.5% - 6.9%).

Table 19: A-priori estimations for selected descriptive item characteristics I of the DISABKIDS chronic generic short-form measure (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)

DCGM-37		***N vc***	**Missing**	**1**	**2**	**3**	**4**	**5**
				never	*seldom*	*quite often*	*very often*	*always*
Facet	*No.*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*
01 IND	04	499	**33 (6.2)**	**2.6**	6.8	15.1	37.5	32.8
02 IND	05	493	**39 (7.3)**	**2.8**	**4.9**	15.8	38.9	31.1
03 EMO ‡	14	502	**30 (5.6)**	**2.8**	7.0	24.2	27.4	32.6
04 EMO‡	18	500	**32 (6.0)**	**1.3**	**4.0**	20.2	30.4	38.5
05 EXCL ‡	20	498	**34 (6.4)**	**1.1**	**4.0**	9.2	21.5	**57.9**
06 EXCL ‡	25	497	**35 (6.6)**	**1.3**	6.4	17.5	29.1	40.4
07 INCL	29	498	**34 (6.4)**	**1.5**	5.8	10.6	35.7	40.9
08 INCL	30	501	**31 (5.8)**	**.6**	**1.3**	10.0	30.9	**51.3**
09 LIM ‡	09	502	**30 (5.6)**	5.5	9.8	24.0	28.5	27.0
10 LIM ‡	12	500	**32 (6.0)**	**1.5**	5.3	25.7	29.4	32.8
11 MED ‡	35	448	***84 (15.8)*****	9.2	8.9	23.4	20.8	23.6
12 MED ‡	36	448	***84 (15.8)*****	8.9	7.5	18.7	22.3	28.7

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: missing data rate > 5%; response rates > 50% / < 5%. N_{VC}: valid cases; ‡ Reversed scored items (recoded before analysis). * Missing data rates are systematically overestimated, as they include the non-applicable rates of respondents without any medical treatment; adjusted missing data rates reaching n = 3 (0.6%) for item 11/35 and n = 3 (0.6%) for item 12/36.

5.3.1.2 Basic descriptive analysis on item level II: Key indicators

Self-report version: Mean values of items ranged from 3.53 to 4.52 (possible range: 1 - 5). Skewness values ranged from │.53│ to │2.24│; two items displayed curtosis values higher than│.2│ and one item was strongly skewed (>│2│). Floor effects (> 20 %) were not detected for any of the items, whereas all 12 items displayed substantial ceiling effects (> 20 %), with response frequencies strongly above 50% regarding the highest category ("5") for 2 out of 12 items. Both these items were derived from the "social" domain but differ with respect to facet-assignment and thus also to scoring direction. Item 5 (DCGM-37: item 20) display a ceiling effect of 72.7 (curtosis = 4.84, skewness: - 2.24), item 8 (DCGM-37: item 30) of 62.8 (curtosis: 3.14, skewness: - 1.75).

Table 20: A-priori estimations for selected descriptive item characteristics II of the DISABKIDS chronic generic short-form measure (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

DCGM-37		***M***	***(SD)***	***Floor („1")***	***Ceiling („5")***	***Curtosis***	**Skewness**
No.	*Facet*	*No.*	*(1-5)*	*(1-5)*	*%*	*%*		
01	IND	4	3.96	1.201	6.1	43.6	.135	- 1.037
02	IND	5	4.01	1.127	4.2	43.4	.265	- 1.031
03	EMO ‡	14	4.01	1.145	4.2	45.2	.112	- 0.997
04	EMO‡	18	4.22	0.994	2.2	50.0	1.082	- 1.254
05	EXCL ‡	20	4.57	0.862	1.7	**72.7**	**4.835**	**- 2.244**
06	EXCL ‡	25	4.14	1.097	4.0	50.0	.740	- 1.204
07	INCL	29	4.11	1.070	3.1	46.7	.608	- 1.141
08	INCL	30	4.52	0.768	0.5	**62.8**	**3.143**	- 1.753
09	LIM ‡	9	3.88	1.224	5.5	41.7	-.369	- 0.824
10	LIM ‡	12	3.94	1.091	2.8	40.3	-.299	- 0.725
11	MED ‡	35	3.56	1.452	12.3	34.3	-1.075	- 0.531
12	MED ‡	36	3.53	1.500	15.6	33.0	-1.080	- 0.584

Notes: Facet abbreviations are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Ceiling or floor effect over 50%; |curtosis| > 2; |skewness| > 2; N_{VC}: valid cases; M: mean; SD: standard deviation. ‡ Reversed scored items (recoded before analysis).

Proxy-report version: Mean values of items ranged from 3.47 to 4.47 (possible range: 1 - 5). Skewness values ranged from │.43│ to │1.41│. Floor effects (> 20%) were not detected for any of these items. The highest response rates for the first category of the answering scale pertains to items derived from the "medication" facet, with nearly 10% for both items (8.9% and 9.2%). Ceiling effects (> 20 %) were observed for all 12 items. The same items as for the self-report version displayed strong "ceiling effects": recoded item 5 (DCGM-37: 20) with a value 57.9% (curtosis: 1.04, skewness: - 1.36) and and item 8 (DCGM-37: 30) reaching a value of 51.3% (curtosis: 1.95, skewness: - 1.41).

Table 21: A-priori estimations for selected descriptive item characteristics II of the DISABKIDS chronic generic short-form measure (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)

DCGM-37			***M***	***(SD)***	***Floor („1“)***	***Ceiling („5“)***	**Curtosis**	**Skewness**
No.	*Facet*	*No.*	*(1-5)*		*n (%)*	*n (%)*		
01	IND	04	4.04	.993	2.6	32.8	.613	- 1.015
02	IND	05	3.98	1.028	2.8	31.1	.420	- .971
03	EMO ‡	14	3.82	1.069	2.8	32.6	- .146	- .674
04	EMO‡	18	3.91	1.024	1.3	38.5	- .489	- .578
05	EXCL ‡	20	4.34	.955	1.1	**57.9**	1.043	- 1.363
06	EXCL ‡	25	4.01	.981	1.3	40.4	.074	- .765
07	INCL	29	4.18	.936	1.5	40.9	.750	- 1.117
08	INCL	30	4.47	.729	.6	**51.3**	1.948	- 1.413
09	LIM ‡	09	3.65	1.164	5.5	27.0	- .468	- .584
10	LIM ‡	12	3.88	1.044	1.5	32.8	- .276	- .630
11	MED ‡	35	3.47	1.277	9.2	23.6	- .771	- .433
12	MED ‡	36	3.60	1.284	8.9	28.7	- .595	- .632

Notes: Facet abbreviations are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Ceiling or floor effect over 50%. N_{VC} = valid cases; M = mean; SD = standard deviation. ‡ Reversed scored items (recoded before analysis).

Overall, basic descriptive analysis at the item level indicated two items with more than 50% ceiling effect both in the self- and the proxy-report version, with one item (item 20) additionally being strongly skewed (│2.24│). None of the 12 items included displays floor effects according to a critical value of more than 20%, but all items reached response rates above 20% for the highest response choice option, indicating substantial ceiling effects. Both items assigned to the "medication" facet displayed non-ordered response choice frequencies according to a linear/smoothed distribution pattern of frequencies. Distributions for both items are "collapsing" for both, self- and proxy-report version.

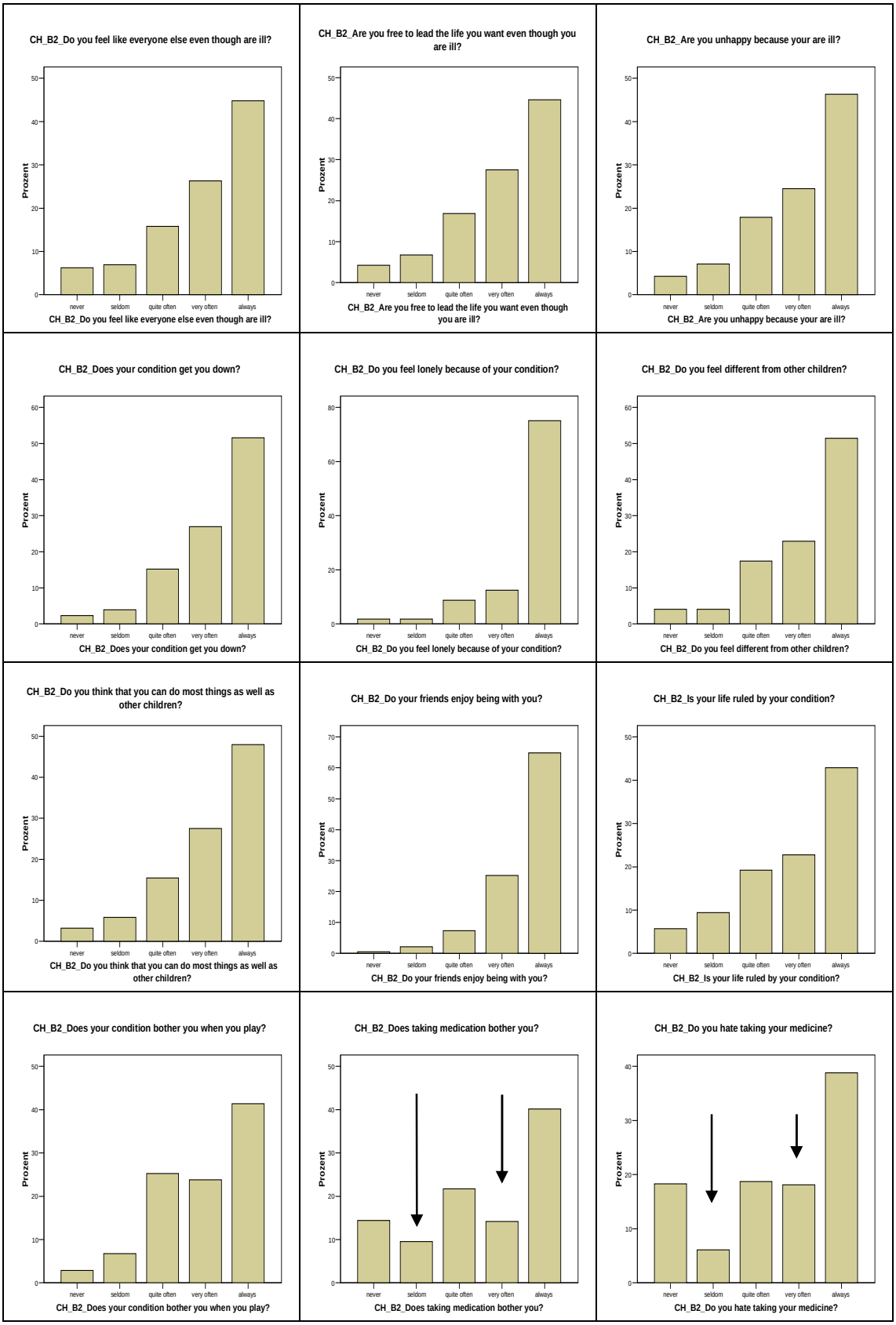


Figure 1: **A-priori estimations of item response choice frequency distributions of the DISABKIDS chronic generic short-form measure (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)**

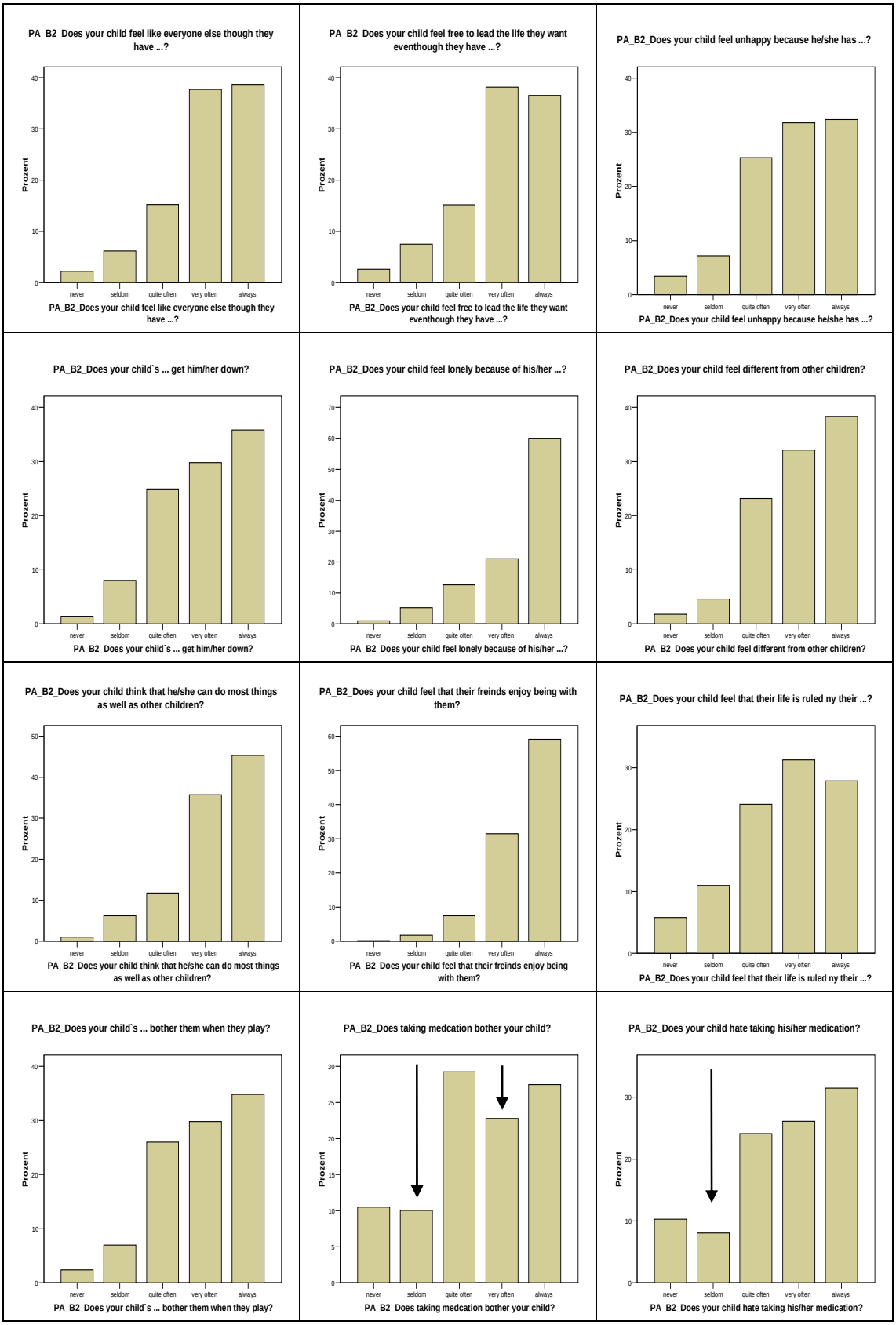


Figure 2: **A-priori estimations of item response choice frequency distributions of the DISABKIDS chronic generic short-form measure (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)**

5.3.2 Analyses on inter-item level

Self-report version: With respect to the conceptual model of the DISABKIDS measurement approach, the inter-item correlation matrix shows that discriminant validity of items was high at the facet level, but absent at the domain level. At the facet level, correlation coefficients (Pearson r) for paired items from each facet rank highest, exept for items derived from both "social" facets.

Table 22: A-priori estimations for coefficients of item-inter correlations (Pearson's r) of the DISABKIDS chronic generic short-form measure item pool (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

(rank)		**IND**	**IND**	**EMO***	**EMO***	**EXC***	**EXC***	**INC**	**INC**	**LIM***	**LIM***	**MED***	**MED***
01	**IND**	*(1)*	**.594**	.393	.416	.403	*.502*	*.473*	.306	.390	.300	.152	.199
02	**IND**	**.594**	*(1)*	.414	.454	.408	*.492*	*.499*	.287	.385	.426	.226	.208
03 *	**EMO** ‡	.393	.414	*(1)*	**.491**	.388	*.449*	.288	.193	.312	.344	.392	.326
04 *	**EMO**‡	.416	.454	.491	*(2)*	***.495***	*.460*	.330	.208	.408	.368	.304	.298
05	**EXCL** ‡	*.403*	*.408*	*.388*	***.495***	*(2)*	.484	.399	.362	.345	.333	.198	.239
06	**EXCL** ‡	***.502***	*.492*	*.449*	*.460*	.484	***(3)***	.402	.303	*.419*	*.419*	.339	.309
07*	**INCL**	***.473***	*.499*	.288	.330	.399	.402	***(5)***	.340	.292	.368	.063	.095
08 *	**INCL**	.306	.287	.193	.208	**.362**	.303	.340	*(2)*	.156	.204	.063	.095
09*	**LIM** ‡	*.390*	*.385*	*.312*	*.408*	*.345*	*.419*	*.292*	.156	*(1)*	**.434**	.198	.189
10 *	**LIM** ‡	*.300*	*.426*	*.344*	*.368*	*.333*	*.419*	*.368*	.204	**.434**	*(1)*	.256	.234
11 *	**MED** ‡	.152	.226	*.392*	*.304*	.198	*.339*	.114	.063	.198	.256	*(1)*	**.704**
12 *	**MED** ‡	.199	.208	*.326*	*.298*	*.239*	*.309*	.169	.095	.189	.234	**.704**	*(1)*

Notes: *Facet abbreviations* according to DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Line-by-line highest ranked coefficient for inter-item correlation of each respective item included. In **parenthesis**: Rank of coefficient for correlation (Pearson's r) between items derived from the same facet of the DCGM-37.

Thus, discriminant validity is ensured for 8 out of 12 items at the facet level by means of manifest inter-item corelations. With respect to the domain level, discriminant validity of items is strongly impaired, as all items display higher correlation coefficients for at least one item not originally assigned to the respective domain (compared to correlation coefficients of items assigned to a shared domain, despite the item having originally been assigned to the same facet).

Proxy-report version: For the proxy-report versison, the inter-item correlation matrix also shows that discriminant validity of items is high – even higher than for the self-report version – at the facet level, but again absent at the domain level. At the facet level, correlation coefficients (Pearson r) for paired items from each facet rank highest, except for both items derived from the physical "limitation" facet.

Table 23: **A-priori estimations for coefficients of item-inter correlations (Pearsons r) of the DISABKIDS chronic generic short-form measure item pool (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578**)

(rank)		**IND**	**IND**	**EMO***	**EMO***	**EXC***	**EXC***	**INC**	**INC**	**LIM***	**LIM***	**MED***	**MED***
01	IND	*(1)*	**.665**	.538	.510	*.541*	*.589*	*.539*	.431	.474	.441	.230	.234
02	IND	**.665**	*(1)*	.466	.455	*.521*	*.520*	*.559*	.352	*.514*	.437	.217	.232
03 *	EMO ‡	.538	.466	*(1)*	**.633**	*.541*	*.555*	.415	.276	.518	.503	.455	.452
04 *	EMO‡	.510	.455	**.633**	*(1)*	*.580*	*.561*	.383	.300	*.555*	.458	.319	.336
05	EXCL ‡	*.541*	*.521*	*.541*	*.580*	*(1)*	**.598**	.507	.424	.482	.422	.232	.253
06	EXCL ‡	*.589*	*.520*	*.555*	*.561*	**.598**	*(1)*	.500	.407	.481	*.514*	.368	.358
07*	INCL	*.539*	***.559***	.415	.383	.507	.500	*(2)*	.541	.314	.444	.123	.158
08 *	INCL	*.431*	.352	.276	.300	.424	.407	**.541**	*(1)*	.196	.286	.183	.238
09*	LIM ‡	*.474*	*.514*	*.518*	***.555***	*.482*	*.481*	*.314*	.196	***(7)***	.403	.290	.271
10 *	LIM ‡	*.441*	*.437*	*.503*	*.458*	*.422*	***.514***	*.444*	.286	.403	***(8)***	.311	.329
11 *	MED ‡	.230	.217	*.455*	*.319*	.232	*.368*	.123	.183	.290	.311	*(1)*	**.793**
12 *	MED ‡	.234	.232	*.452*	*.336*	.253	*.358*	.158	.238	.271	.329	**.793**	*(1)*

Notes: *Facet abbreviations* according to DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Line-by-line highest ranked coefficient for inter-item correlation of each respective item included. In **parenthesis**: Rank of coefficient for correlation (Pearson's r) between items derived from the same facet of the DCGM-37.

Correlation coefficient for the first item from the "inclusion" facet with its paired item just ranks on second place – highest coefficient is displayed with item 2 from the original "independence" subscale. Differences between total values of coefficients are small (Δ_r =.18). Thus, discriminant validity of items is given for 9 out of 12 items at the facet level for the manifest inter-item corelations. With respect to the domain level, discriminant validity of items is impaired, as all items displayed higher correlation coefficients with at least one item not originally assigned to the respective domain (compared to correlation coefficients of items assigned to a shared domain, despite the item having originally been assigned to the same facet). Correlation coefficients for items originally assigned to the same domain but to different facets do rank rather low for selected items. *Overall* (and despite of item validity issues with respect to the original measurement model), adjusting for correlational heterachical structure with respect to item-facet and item-domain assignment can be judged as a meanigful procedure for our purposes.

5.3.3 Analyses of structure and measurement model

5.3.3.1 Exploratory factor anaylsis

Two different approaches of exploratory factor analysis were applied: (I) principal component analysis (PCA) with oblique Varimax rotation, resulting in uncorrelated factors; and (II) principle axis analysis with oblimin Promax rotation. The latter is the preferred method in the context of this investigation because it results in correlated factors. As our model is specified to indentify a measure with one dimension, a multidimensional structure of the item pool (according to the Eigenvalue criterion of a critical value > 1.0) would be violating this assumption.

Both procedures resulted in a two-factor solution for the 12 items and a one-factor solution for 10 items. Both "applicable" items derived from the "medication" facet are assigned exclusively to the second factor within the two factor solution. Consequently, excluding both items resulted in a one factor model, irrespective of the method used.

Table 24: Results of explorative factor analysis for final DCGM-12 item pool (self-report) using two different methods (PCA with Varimax rotation and PAF with Promax rotation) (sub-sample II from the DISABKIDS field study sample, n = 578)

		12-item solution				**10-item solution**	
Method		*PCA (Varimax)*		*PAF (Promax)*		*PCA (Varimax)*	*PAF (Promax)*
		Factor I	*Factor II*	*Factor I*	*Factor II*	*Factor I*	*Factor I*
01	IND	**.753**	.077	**.755**	-.119	**.721**	**.682**
02	IND	**.760**	.145	**.768**	-.070	**.762**	**.733**
03	EMO	**.508**	.477	**.486**	.247	**.647**	**.594**
04	EMO	**.603**	.375	**.591**	.142	**.698**	**.652**
05	EXCL	**.680**	.201	**.656**	.009	**.689**	**.641**
06	EXCL	**.654**	.351	**.639**	.131	**.749**	**.716**
07	INCL	**.697**	.015	**.654**	-.113	**.659**	**.607**
08	INCL	**.554**	-.070	**.471**	-.102	**.477**	**.418**
09	LIM	**.517**	.275	**.492**	.077	**.611**	**.554**
10	LIM	**.496**	.301	**.461**	.113	**.616**	**.559**
11	MED	.060	**.898**	-.098	**.963**	-	**-**
12	MED	.080	**.863**	.013	**.740**	-	**-**
Eigen value		4.753	1.500			4.455	
Variance (%)		39.609	12.501	*(dito)*		44.549	*(dito)*
Total variance (%)			52.110			44.549	

Notes: Facet abbreviations are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". Values for PCA with 12-item model are provided for the rotated factor solution; values for PAF with 12-item model are provided for the structure matrix; PCA = principle component analysis; PAF = principle axis analysis.

5.3.3.2 Confirmatory factor analysis

After exploring the internal structure of the item pool, the measurement model of the DISABKIDS 12 item short-form was empirically investigated using confirmatory factor analysis (CFA). We use Maximum-likelihood parameter estimation for testing the model. Despite of impaired normal distribution of items, this method can be applied as it is assumed to be robust even if the data violates the assumption of normal distribution. Two models were included in confirmatory testing: A single-trait-model and a single-trait-model with a method factor. Within the last model, apart from a latent variable defining the model as strictly unidimensional (all items are expected to be explained by the same underlying variable), a method factor is specified that captures variance stemming from different poling of items.

Self-report version: For the one-factor model of the short-form with 12 items, the model did not fit the data sufficiently (RMR = .15; GFI = .86, CFI = .79, RMSEA = .13). Estimating the model fit by excluding both "applicable" items (stemming from the "treatment" facet), fit statistics improved notably and were acceptable, although not sufficient (RMR = .05; GFI = .94, CFI = .92, RMSEA = .08). The same item solution was also tested with one specified method factor (single-trait-single-method model). Overall model fit for both item pool solutions again improved notably, although the 12 item solution again did not fit the data (RMR = .14; GFI = .89, CFI = .83, RMSEA = .12). The 10-item single-trait-model with a method factor displayed good fit (RMR = .04; GFI = .96, CFI = .96, RMSEA = .07) as compared to the other solutions and the overall model fit improved notably. Table 25 provides a detailed overview on the results for all models. On the item level, one item (item 10) reach a standardized regression weight lower than .40.

Proxy-report version: Models were also tested for the proxy version, resulting in similar results. Model fit for the single-factor model with 12 items was poor (RMR = .13; GFI = .80, CFI = .76, RMSEA = .16), and overall fit statistic improved significantly when the model was estimated for the 10 item version, thus discarding both "applicable" items" from the "medication/treatment" facet (RMR = .05; GFI = .89, CFI = .90, RMSEA = .11). Testing both models by specifying an additional method-factor for reversed scored items, fit statistics for both models indicated even better performance of the 12 item version (RMR = .12; GFI = .84, CFI = .81, RMSEA = .15) as well as of the 10-item version (RMR = .04; GFI = .94, CFI = .95, RMSEA = .09). An overview of the fit statistics is given in Table 26, with values for regression coefficients on item level also being provided.

Table 25: **Selected indicators of fit statistics and standardized regression weights of confirmatory factor analysis (one-factor model) for the DISABKIDS 12 [10] item short-form measure (self-report version (sub-sample II from the DISABKIDS field study sample, n = 578**)

Fit statistics	**CFA (ST-model)**	*(QOL→Item)*		**CFA (ST-Model with MF)**	*(QOL→Item)*	
Item version	*12 [10]*		*12 [10]*	*12 [10]*		*12 [10]*
Parameter	24 [20]	*01.*	.674 [.691]	28 [24]	*01.*	.593 [.618]
Discrepancy (chi-square)	457.355 [165.851]	*02.*	.713 [.737]	372.651 [107.787]	*02.*	.638 [.672]
df	54 [35]	*03.*	.628 [.596]	50 [31]	*03.*	.649 [.616]
p	.000 [.000]	*04.*	.672 [.648]	.000 [.000]	*04.*	.693 [.678]
Discrepancy (chi-square)/ df	8.470 [4.739]	*05.*	.650 [.633]	7.453 [3.477]	*05.*	.653 [.654]
RMR	.150 [.052]	*06.*	.711 [.712]	.136 [.040]	*06.*	.717 [.725]
GFI	.864 [.940]	*07.*	.575 [.610]	.892 [.961]	*07.*	.497 [.544]
NFI	.767 [.905]	*08.*	.405 [.419]	.810 [.939]	*08.*	.358 [.391]
TLI	.740 [.902]	*09.*	.537 [.554]	.776 [.935]	*09.*	.551 [.572]
CFI	.788 [.923]	*10.*	.521 [.555]	.830 [.955]	*10.*	.532 [.569]
RMSEA	.126 [.084]	*11.*	*.414 [-] **	.119 [.068]	*11.*	*.467 [-] **
RMSEA (CI - 95 %)	(.116 - .137) / [(.071 - .094)]	*12.*	*.413 [-] **	(.106 - .128) / [(.054 - .082)]	*12.*	*.458 [-] **
p	.000 [.001]			.000 [.015]	*(MF→Item)*	
AIC	480.079 [205.851]			428.651 [155.787]	*01.*	.481 [.444]
BCC	506.721 [206.694]			430.244 [156.799]	*02.*	.464 [.401]
BIC	605.072 [291.459]			544.987 [258.517]	*07.*	.387 [.341]
CAIC	629.072 [311.459]			572.987 [282.517]	*08.*	.208 [.148]

Notes: Maximum-likelihood parameter estimation. ST-model = single-trait model; SF-model with MF = single-trait -method model with method factor. CFA = confirmatory factor analysis. df = degrees of freedom; p = p-value for chi-square statistics; RMR = root mean square residual; NFI = normed fit index; GFI = genral fit index; TLI = Tucker-Lewis fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation. AIC = Akaike Information Criteria, BCC = Browne-Cudeck criterion, BIC = Bayes Information Criterion, CAIC = Consisten Akaike Information Criterion. * Items 11 and 12 of the DISABKIDS 12 item short-form are “applicable” items.

Table 26: **Selected indicators of fit statistics and standardized regression weights of confirmatory factor analysis (one-factor model) for the DISABKIDS 12 [10] item short-form measure (proxy-report version) (sub-sample II from the DISABKIDS field study sample, n = 578**)

Fit statistics	**CFA (ST-model)**	*(QOL→Item)*		**CFA (ST-Model with MF)**	*(QOL→Item)*	
Item version	*12 [10]*		*12 [10]*	*12 [10]*		*12 [10]*
Parameter	24 [20]	*01.*	.738 [.759]	28 [24]	*01.*	.674 [.712]
Discrepancy (chi-square)	657.860 [252.055]	*02.*	.708 [.721]	536.659 [154.620]	*02.*	.641 [.669]
df	54 [35]	*03.*	.754 [.737]	50 [31]	*03.*	.788 [.765]
p	.000 [.000]	*04.*	.748 [.736]	.000 [.000]	*04.*	.776 [.771]
Discrepancy (chi-square)/ df	12.183 [7.202]	*05.*	.731 [.757]	10.733 [4.988]	*05.*	.723 [.757]
RMR	.126 [.052]	*06.*	.749 [.761]	.116 [.040]	*06.*	.742 [.761]
GFI	.801 [.894]	*07.*	.610 [.657]	.840 [.935]	*07.*	.527 [.583]
NFI	.744 [.891]	*08.*	.476 [.600]	.791 [.933]	*08.*	.408 [.429]
TLI	.705 [.877]	*09.*	.652 [.641]	.744 [.921]	*09.*	.674 [.671]
CFI	.759 [.904]	*10.*	.603 [.611]	.806 [.946]	*10.*	.601 [.613]
RMSEA	.162 [.114]	*11.*	*.417 [-] **	.151 [.092]	*11.*	*.468 [-] **
RMSEA (CI - 95 %)	(.151 - .173) / [(.101 - .1484)]	*12.*	***.441** [-] **	(.140 - .163) / [(.077 - .106)]	*12.*	***.475** [-] **
p	.000 [.000]			.000 [.015]	*(MF→Item)*	
AIC	705.860 [292.055]			592.659 [202.620]	*01.*	.441 [.315]
BCC	707.371 [293.002]			594.422 [203.756]	*02.*	.447 [.343]
BIC	803.223 [375.406]			706.249 [302.641]	*07.*	.481 [.510]
CAIC	827.223 [395.406]			734.249 [326.641]	*08.*	.362 [.428]

Note: Maximum-likelihood parameter estimation. ST-model = single-trait model; STSM-model = single-trait-single-method model. CFA = confirmatory factor analysis. df = degrees of freedom; p = p-value for chi-square statistics; RMR = root mean square residual; NFI = normed fit index; GFI = genral fit index; TLI = Tucker-Lewis fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation. AIC = Akaike Information Criteria, BCC = Browne-Cudeck criterion, BIC = Bayes Information Criterion, CAIC = Consisten Akaike Information Criterion. * Items 11 and 12 of the DISABKIDS 12 item short-form are “applicable” items.

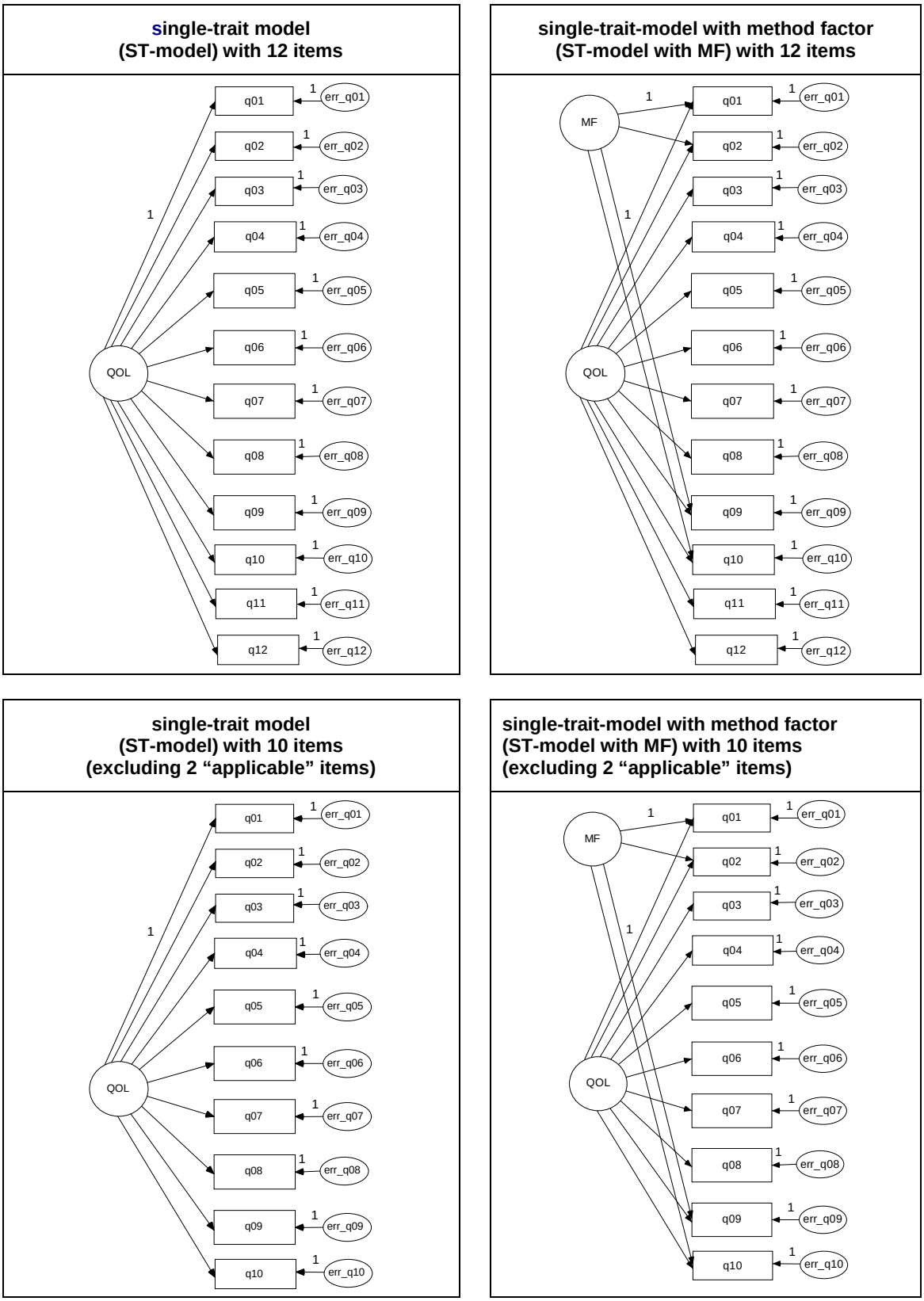


Figure 3: **Different structure and measurement models for confirmatory factor analysis of the DISABKIDS short-form measure**

Self-report version: Item internal consistencies as indicated by (part-whole) corrected item-total correlations reached acceptable values, with item 10 (.38) being an exception. Nevertheless, almost more than 90% of the items (12-item version: 91.7%; 10-item version: 90%) reached item-total correlations above .40. Exclusion of any of the items did not further increase internal consistency of the composite measure (α = .86).

Table 27: **A-priori estimations for selected descriptive item characteristics of the DISABKIDS chronic generic short-form measure (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578**)

DCGM-37			***R^2***		***IIC***		**α_{-i}**	
No.	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND‡	04	.479	[.467]	.572	[.624]	.834	[.839]
02	**IND**‡	**05**	.503	[.501]	.623	[.676]	.831	[.835]
03	EMO	14	.376	[.334]	.590	[.545]	.833	[.847]
04	**EMO**	**18**	.419	[.407]	.613	[.602]	.832	[.842]
05	EXCL	20	.412	[.390]	.582	[.590]	.836	[.844]
06	**EXCL**	**25**	.446	[.442]	.649	[.658]	.829	[.836]
07	INCL‡	29	.355	[.368]	.496	[.558]	.840	[.845]
08	INCL‡	30	.203	[.198]	**.343**	[**.380**]	.849	[.858]
09	LIM	09	.294	[.312]	.482	[.514]	.841	[.850]
10	LIM	12	.280	[.326]	.487	[.522]	.840	[.848]
11	MED	35	.541	-	.454	-	.846	-
12	MED	36	.512	-	.451	-	.847	-

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Marker items according to ICC ranking ("top three"); IIC < .40. IIC: item internal consistency. ‡ Reversed scored items (recoded before analysis).

Proxy-report version: Item internal consistencies reached satisfactory values for all items (r = .47-.72), with critical value of .40 being the basis of evaluation.

Table 28: **A-priori estimations for selected descriptive item characteristics of the DISABKIDS chronic generic short-form measure (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578**)

DCGM-37			R^2		*IIC*		α_{-I}	
No.	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND‡	04	.574	[.569]	.661	[.716]	.874	[.885]
02	IND‡	05	.562	[.547]	.637	[.684]	.875	[.887]
03	**EMO**	**14**	.575	[.532]	.727	[.696]	.870	[.886]
04	**EMO**	**18**	.557	[.550]	.685	[.697]	.873	[.886]
05	EXCL	20	.511	[.532]	.656	[.714]	.875	[.885]
06	**EXCL**	**25**	.516	[.527]	.703	[.720]	.872	[.885]
07	INCL‡	29	.481	[.496]	.546	[.624]	.880	[.891]
08	INCL‡	30	.360	[.340]	.443	[.465]	.885	[.900]
09	LIM	09	.455	[.448]	.591	[.598]	.878	[.894]
10	LIM	12	.350	[.362]	.570	[.581]	.879	[.894]
11	MED	35	.640	-	.471	-	.887	-
12	MED	36	.638	-	.489	-	.886	-

Notes: *Facet abbreviations* according to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: Marker items according to ICC ranking ("top three"); IIC < .40. IIC: item internal consistency. ‡ Reversed scored items (recoded before analysis).

5.3.3.3 Applying the Rasch model

Self-report-version: Significance tests (p-values) for the z[Q]-statistics displayed sufficient fit for applying the (ordinal) partial credit model to all 12 items, indicating no violation of the assumption of unidimensionality for the extended Rasch model. Item locations vary between – 0.72 (item 8/DCGM-37: 30) and 0.63 (item 12/DCGM-37: item 55). However, most items displayed disordered thresholds (not reported in detail).

Table 29: A-priori estimations for selected Rasch analytic item characteristics of the DISABKIDS chronic generic short-form measure (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

DCGM-37			***Q* ****		**z (Q)***		**p ***		**location ***		**thresholds ***	
No	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND	04	0.11	[0.09]	- 0.24	[- 0.22]	0.59	[0.59]	0.17 (09.)	[0.31 (09.)]	**+**	[**+**]
02	IND	05	0.10	[0.08]	- 0.68	[- 0.64]	0.75	[0.74]	0.05 (07.)	[0.18 (07.)]	**+**	[**+**]
03	EMO	14	0.10	[0.11]	- 0.46	[0.20]	0.68	[0.42]	0.09 (08.)	[0.19 (08.)]	**+**	[**+**]
04	EMO	18	0.11	[0.10]	- 0.72	[- 0.24]	0.76	[0.60]	- 0.20 (03.)	[- 0.11 (03.)]	**+**	[**+**]
05	EXCL	20	0.08	[0.07]	- 0.78	[- 0.54]	0.78	[0.71]	- 0.52 (02.)	[- 0.50 (02.)]	**++**	[**++**]
06	EXCL	25	0.09	[0.08]	- 0.90	[- 0.51]	0.82	[0.69]	- 0.02 (06.)	[0.09 (05.)]	**++**	[**+**]
07	INCL	29	0.15	[0.12]	0.10	[0.15]	0.76	[0.44]	- 0.14 (04.)	[0.01 (04.)]	**+**	[**+**]
08	INCL	30	**0.19**	**[0.18]**	0.40	[0.55]	0.35	[0.29]	- 0.72 (01.)	[- 0.67 (01.)]		
09	LIM	09	0.15	[0.12]	0.50	[0.67]	0.31	[0.25]	0.18 (10.)	[0.35 (10.)]		
10	LIM	12	**0.16**	[0.14]	0.37	[0.58]	0.35	[0.28]	- 0.04 (05.)	[0.15 (06.)]	**++**	[**+**]
11	MED	35	*0.14*	-	*1.05*	-	*0.15*	-	0.54 (11.)	**-**	**++**	-
12	MED	36	*0.15*	-	*1.33*	-	*0.09*	-	0.63 (12.)	**-**	**++**	-

Notes: *Facet abbreviations* to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: thresholds in case of non-ordering ("+" indicating frequency of collapsing categories). N_{VC} = valid cases; Q = Q-index; z (Q) = z-value of Q-index); Q > 0.15; p = p-value. ‡ Reversed scored items (recoded before analysis). * Item fit statistic assessed by the Q-index according to the (Ordinal) Partial Credit Model.

Additional anaylsis I – Exploring effects of recoded response choices: After merging both response choice options into one category by means of recoding, 5 items still displayed collapsing thresholds for the 12-item solution. For the 10-item version, the (recoded) data completely fitted the (ordinal) Rasch model with respect to significance tests (p-values) for the z[Q] and threshold ordering, except for item 5 (DCGM-37: 20) displaying one collapsed threshold.

Table 30: A-priori estimations for selected Rasch analytic item characteristics of the DISABKIDS chronic generic short-form measure after recoding of answering categories (DCGM-12 self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

	DCGM-37		***Q* ****		**z (Q)***		**p ***		**location ***		**thresholds ***	
No	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND	04	0.11	0.09	- 0.45	- 0.40	0.67	0.66	0.15 (09.)	0.30 (09.)		
02	IND	05	0.11	0.08	- 0.81	- 0.74	0.79	0.77	0.08 (06.)	0.21 (06.)		
03	EMO	14	0.11	0.11	- 0.61	0.21	0.73	0.42	0.13 (08.)	0.23 (07.)		
04	EMO	18	0.11	0.10	- 0.91	- 0.33	0.82	0.63	- 0.19 (03.)	- 0.12 (03.)		
05	EXCL	20	0.15	0.08	- 0.97	- 0.64	0.83	0.74	- 0.71 (01.)	- 0.70 (01.)	**++**	[**+**]
06	EXCL	25	**0.20**	0.08	- 1.18	- 0.69	0.88	0.76	- 0.12 (05.)	0.00 (04.)	**+**	
07	INCL	29	0.15	0.12	0.25	0.22	0.40	0.42	- 0.17 (04.)	0.02 (05.)		
08	INCL	30	**0.16**	**0.18**	0.75	0.87	0.23	0.19	- 0.69 (02.)	- 0.66 (02.)		
09	LIM	09	0.15	0.12	0.58	0.75	0.28	0.23	0.30 (10.)	0.45 (10.)		
10	LIM	12	**0.16**	0.14	0.54	0.75	0.29	0.23	0.11 (07.)	0.27 (08.)	**+**	
11	MED	35	*0.15*	-	*1.27*	-	*0.10*	-	0.54 (11.)	**-**	**+**	-
12	MED	36	***0.16***	-	*1.53*	-	*0.06*	-	0.56 (12.)	**-**	**+**	-

Notes: *Facet abbreviations* to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: thresholds in case of non-ordering ("+" indicating frequency of collapsing categories). N_{VC} = valid cases; Q = Q-index; z (Q) = z-value of Q-index); Q > 0.15; p = p-value. ‡ Reversed scored items (recoded before analysis). * Item fit statistic assessed by the Q-index according to the (Ordinal) Partial Credit Model.

Additional anaylsis II – Exploring differential response choice patterns with respect to age group: As non-ordered thresholds are caused by insufficient answering distribution patterns (high skewness, non-ordering of category frequencies etc.), it can be expected that the appearance of collapsed categories is related to differential response behaviour. With respect to the specific sample characteristics of the DISABKIDS study, the question raises if non-ordered thresholds are more likely to appear for children then for teenagers due to different stages of intellectual developement. Thus, an additional computation of the originally non-recoded data was computed separately for the younger age group (8-12 years) and the older age group (13-16 years) separately. Results did not indicate substantial differences in response patterns between both age groups (see appendix).

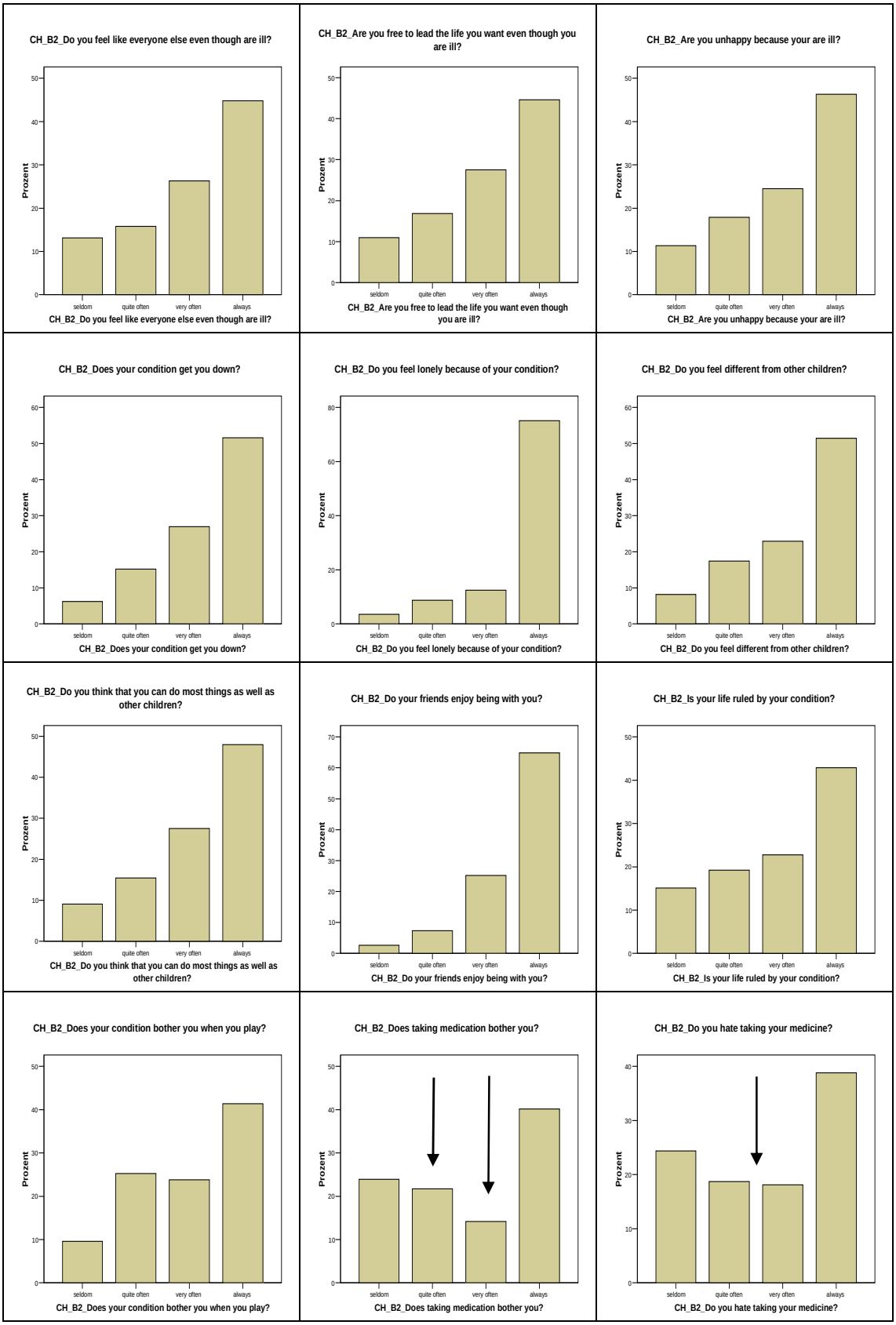


Figure 4: **A-priori estimations of item response choice frequency distributions of the DISABKIDS chronic generic short-form measure after recoding of answering categories (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)**

Proxy-report version: Item fit statistics for the extended Rasch model displayed insufficient fit for both items derived from the "medication facet" of the DCGM-37: Item 11 "Does taking medication bother your child?" and Item 12 "Does your child hate taking his/her medicine?" (Q_{11} = .16, $z[Q]_{11}$ = 2.22, $p[z(Q)]_{11}$ = 0.01; Q_{12} = .16, $z[Q]_{12}$ = 2.53, $p[z(Q)]_{12}$ = 0.01). Thus, inclusion of these two items violatedes the assumption of unidimensionality underlying the (ordinal) partial credit model. In addition, both items as well as item 8 displayed disordered thresholds (see table; not reported here in detail).

Table 31: **A-priori estimations for selected Rasch analytic item characteristics of the DISABKIDS chronic generic short-form measure (DCGM-12 proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578**)

DCGM-37			***Q** **		**z (Q)***		**p ***		**location ***		**thresholds ***	
No	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND	04	0.10	0.08	- 0.64	- 0.53	0.74	0.70	- 0.14 (04.)	0.08 (05.)		
02	IND	05	0.11	0.09	- 0.36	- 0.20	0.64	0.58	- 0.01 (07.)	0.17 (07.)		
03	EMO	14	0.08	0.09	- 1.05	- 0.22	0.85	0.59	0.26 (09.)	0.44 (09.)		
04	EMO	18	0.09	0.09	- 0.75	- 0.31	0.77	0.62	- 0.02 (06.)	0.09 (06.)		
05	EXCL	20	0.08	0.06	- 0.72	- 0.65	0.76	0.74	- 0.53 (02.)	- 0.45 (02.)		
06	EXCL	25	0.08	0.08	- 0.96	- 0.59	0.83	0.72	- 0.08 (05.)	0.04 (04.)	**+**	
07	INCL	29	0.14	0.11	0.21	0.18	0.42	0.73	- 0.48 (03.)	- 0.26 (03.)		
08	INCL	30	**0.17**	**0.16**	0.35	0.68	0.36	0.25	- 1.05 (01.)	- 1.10 (01.)		
09	LIM	09	0.13	0.12	0.45	0.90	0.33	0.18	0.57 (10.)	0.76 (10.)		
10	LIM	12	0.13	0.13	0.22	0.73	0.41	0.23	0.03 (08.)	0.23 (08.)		
11	MED	35	***0.16***	-	*1.70*	-	***0.04***	-	*0.79 (11.)*	**-**	**+**	**-**
12	MED	36	*0.15*	-	*1.52*	-	*0.06*	-	*0.76 (12.)*	**-**	**+**	**-**

Notes: *Facet abbreviations* to the DCGM-37 are as follows: IND = "Independence"; EMO = "Emotion"; EXCL = "Exclusion"; INCL = "Inclusion"; LIM = "Limitation"; MED = "Medication". In **bold** print: thresholds in case of non-ordering ("+" indicating frequency of collapsing categories); Q > 0.15; p (z[Q]) < 0.05. N_{VC} = valid cases; Q = Q-index; z (Q) = z-value of Q-index; p = p-value. ‡ Reversed scored items (recoded before analysis). * Item fit statistic assessed by the Q-index according to the (Ordinal) Partial Credit Model.

5.3.4 Differential item functioning

A main assumption of the Rasch model is the equivalent functioning of items across various subgroups. Thus, it should be shown that selected items do not display differential item functioning according to variables known to be sensitive to differential response behaviour. We decided to test for gender, age group, country, condition, and clinical severity (of condition). Zumbos's logistic regression approach (1999) for separately detecting uniform and nonuniform DIF was used to investigate both, direction (uniform vs. non-uniform) and amount of potential DIF. The analyses were conducted for the self-report-version exclusively.

Results indicate less profound differential item functioning than expected. Two items displayed "high" R^2 differences (according to effect size definition provided by Cohen, 1992). Thus, items performed well with respect to differential item functioning.

Table 32: A-priori estimations for differential item functioning of the DISABKIDS chronic generic short-form measure using a logistic regression approach (sub-sample II from the DISABKIDS field study sample, n = 578)

	DCGM-37		***Gender***		***Age group****		***Country *****		***Condition ******		***Severity *******	
			Δ R^2	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*	*Δ R^2*
No.	*Facet*	*No.*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*	*12*	*10*
01	IND‡	04	.002	[.004]	.003	[.003]	.015	[.010]	.004	[.002]	.004	[.003]
02	IND‡	05	.000	[.001]	.002	[.002]	.002	[.001]	.002	[.001]	.001	[.009]
03	EMO	14	.011	[.009]	.003	[.004]	.002	[.002]	.006	[.001]	.001	[.007]
04	EMO	18	.007	[.008]	.005	[.004]	.007	[.005]	.004	[.004]	.000	[.005]
05	EXCL	20	.001	[.004]	**.021**	[.012]	.008	[.008]	.001	[.003]	.010	[.005]
06	EXCL	25	.000	[.001]	.002	[.005]	.000	[.001]	.003	[.006]	.000	[.002]
07	INCL‡	29	.002	[.004]	.000	[.001]	.013	[.009]	.010	[.006]	.009	[.007]
08	INCL‡	30	.015	[.013]	.008	[.005]	**.096**	**[.088]**	.000	[.000]	**.020**	[.019]
09	LIM	09	.003	[.002]	.007	[.004]	.004	[.004]	.012	[.010]	.002	[.004]
10	LIM	12	.005	[.002]	.003	[.004]	.004	[.004]	**.117**	**[.090]**	**.023**	[.015]
11	MED	35	.005	-	.001	-	.009	-	.001	-	.018	-
12	MED	36	.010	-	.002	-	.002	-	.005	-	.016	-

Notes: Interpretation of R^2 differences according to effect sizes (Cohen, 1992): 0.20: small; 0.50: medium; 0.80: large. Meaningful absolute R^2 differences ($\Delta_d \geq .20$) are indicated in **bold**. $\Delta R^2 = R^2$ difference. * Age group I (8-12 years old) vs. age group II (13-16 years old). ** Including Austria, France, Germany, Greece, Netherlands, Sweden, United Kingdom. *** Including asthma, arthritis, cerebral palsy, cystic fibrosis, diabetis, epilepsy, and dermatitis. **** Clinician's judgement on a 3-point ordinal scale.

5.3.5 Analysis on composite score level

Scoring of composite measure: The score of the DCGM-12 multi-item composite measurement scale is based on equal-interval scoring of response categories using Likert-like scaling of items. Thus, the same ease-of-use scoring algorithm as for the DCGM-37 facet scores (and the total score, respectively) was used. All items were equally weighted for scoring.

5.3.5.1 Reliability

Cronbach's alpha as a measure of internal consistency reached satisfactory values, with α = .85 [.86] for the self-report version and α = .89 [.90] for the proxy-report version of the DCGM-12 [10]. Split-half reliability (Guttmann's coefficient) indicates sufficient consistency with values of .91 [.89] for the self-report version and .93 [.92] for the proxy-report version. Internal consistency of test halfs reached values of α = .71 and .72 for the self-report version (r = .83) and α = .79 and .77 for the proxy-report version (r = .87).

Concordance between self- and proxy-report short-form measures, estimated using different correlation coefficients, indicated satisfactory convergence (r = .64; ICC = .64). Respective coefficients were also used to estimate test-retest reliability (r = .84; ICC = .84), indicating high stability with respect to the critical value of > .60. Generalized delta coefficient indicated excellent discriminatory power of the composite scale for both versions ($\delta_{G(S)}$ = .98, $\delta_{G(P)}$ = .99). These results are presented in Table 33.

5.3.5.2 Comparability

According to a scale ranging from 0 to 100 (linear transformed from the raw score distribution), mean score differences (M_{Δ}) between DCGM-12 total score and total score of the DCGM-37 reached about half a point (M_{Δ} = 0.25) for the self-report versions; for the proxy-report version overall mean score difference was below one point (M_{Δ} = 0.72).

Table 33: A-priori estimations for selected descriptive and psychometric composite scale characteristics of the short-form of the DISABKIDS condition generic module (DCGM-12 self-report/proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)

Descritive and psychometric characteristics		DISABKIDS 12-item short-form	
		Self-report	**Proxy-report**
Descritive characteristics			
Transformed score (0-100)	*M (SD)*	76.22 (17.89)	73.79 (18.09)
Raw score	*M (SD)*	47.23 (8.65)	46.09 (9.10)
Utilization: *Min-Max / Range*	*observed* *(possible)*	12-60 / 48 (12-60 / 48)	13-60 / 47 (12-60 / 48)
Floor / Ceiling	*%*		
Skewness		- 0.80	- 0.74
Discriminatory Power	δ_G	.98	.99
Reliability			
Internal consistency	α	.85 [.86]	.89 [.90]
Split-half reliability	*Guttman's*	.91 [.89]	.93 [.92]
Internal consistencies of test-half's	α *(T1)*	.72 [.73]	.79 [.82]
	α *(T2)*	.71 [.75]	.77 [.79]
Correlation between test-half's	$r_{T1, T2}$	.83 [.80]	.87 [.86]
Retest-reliabilty	$r_{t1, t2}$	.84	(n.a.)
	$ICC_{t1, t2}$	.84	(n.a.)
Self-Proxy-Aggreement	$r_{sr, pr}$	.64	(dito)
	$ICC_{sr, pr}$	.64	(dito)
Association ("criterion" validity)			
Item pool	$r_{sf, ov}$	.95 [.94]	.96 [.95]
	$r^2_{sf, ov}$	.91 [.89]	.92 [.91]
Score	$r_{sf, ov}$	.93	.95
	$r^2_{sf, ov}$	.87	.90

Notes: M = mean; SD = standard deviation; α = Cronbach's Alpha; r_{tt} = Test-retest reliability (Pearson correlation coefficient); ICC_{tt} = Test-retest reliability (Intraclass correlation coefficient: one-way, absolute agreement); $r_{s,p}$ = child-proxy-concordance (Pearson correlation coefficient); $ICC_{s,p}$ = child-proxy-concordance (Intraclass correlation coefficient: one-way, absolute agreement); $r_{sf,ov}$ = Association between DCGM-12 total score/item pool and DCGM-37 total score (Pearson correlation coefficient, not corrected for overlap); $R^2_{sf,ov}$ = determination coefficient for prediction of DCGM-37 average score by DCGM-12 total score/item pool. δ_G = Generalized delta coefficient. na = Coefficient not applicable for proxy-report version, as retest data do not include the DCGM item pool for the proxy version.

5.3.5.3 Construct and content validity

Conceptually, *construct validity* can be assumed because the structure of the DCGM-37 has been preserved in the short-form's measurement model and all respective dimensions have been included by representing each of them with 2 items. According to the concordance between short-form total score and total score of the original measure, association is high in terms of correlation coefficient for both self-report ($r_{SR(SF,OV)}$ = .93) and proxy-report ($r_{PR(SF,OV)}$ = .95), not corrected for overlap. Consequently, the DCGM-12 total score explained 87% (self-report) and 93% (proxy-report) of the DCGM-37 total score variance.

Content validity: The DISABKIDS chronic generic proxy version with 37 items has just a slightly higher rate of concepts linked to the ICF (86.4%) as compared to the short-form measure (12 items) which has 78.4% concepts linked to the ICF. Moreover, content density (ratio of the number of meaningful concepts identified, divided by the number of items) for both DISABKIDS chronic generic measures has been reported elsewhere (Fava et al., 2009). Hereby, a value of "1" means that each item of the instrument refers to one concept, and higher values such as 1.5 indicate that on average each item of the module can be referred to 1.5 meaningful concepts. The content density ratio of the measures shows the same values for both condition generic modules (1.2), indicating that on average every item of these questionnaires contain a specific and direct reference to a meaningful health-related concept of the ICF.

Table 34: Frequencies and percentages showing how much ICF categories from a respective general ICF component ("categories") were addressed in the different DISABKIDS chronic generic modules

ICF	DISABKIDS chronic generic modules			
	DCGM-37		DCGM-12	
	n	*%*	*n*	*%*
Number of items (DISABKIDS measure)	37		12	
Number of identified meaningful concepts	44		14	
Content density (identified concepts per item)	1.2		1.2	
Total number of concepts not linked to the ICF	6	13.6%	3	21.4%
Total number of concepts linked to the ICF	38	86.4%	11	78.6%
Number of concepts linked to different ICF components	20	54.1%	6	50.0%
*Total frequency of identified concepts of a measure linked to different ICF categories (per component) **				
(B) "Body-functioning" ICF categories	8	21.6%	3	25.0%
(D) "Activities-and-participation" ICF categories	6	16.2%	1	8.3%
(E) "Environmental factors" ICF categories	6	16.2%	2	16.7%

Notes: * Denominators of percentages are represented by the total frequencies of identified concepts of a measure linked to all ICF categories and are reported in the penultimate row of the table.

5.3.5.4 Convergent validity

Convergent validity was investigated using different other measures of (I) subjective health status and associated constructs (e.g. functional status) in general, and (II) HRQOL in particular

(I) Subjective health status and associated constructs in general: The DCGM-12 score displays moderate correlations with the "General Health Perception" item ($r_{SR(MI,MII)}$ = .32; $r_{PR(MI,MII)}$ = .37) and the CHQ-KINDL-Index score ($r_{SR(MI,MII)}$ = .46; $r_{PR(MI,MII)}$ = .50). Similarly, with respect to the proxy-report, the DCGM-12 score also shows moderate correlations with both subscale scores of the FS-II-R assessing functional status with the factors "General Health" ($r_{PR(MI,MII)}$ = .42) and "Interpersonal Functioning" ($r_{PR(MI,MII)}$ = .36).

Table 35: A-priori correlation coefficients (Pearson r) between total score of the short-form of the DISABKIDS condition generic module (DCGM-12; self-report/proxy-report) and various constructs related to HRQOL (sub-sample II from the DISABKIDS field study sample, n = 578)

DISABKIDS 12-item short-form (DCGM-12)	**Self-report version**	**Proxy-report version**
	Pearson's r	*Pearson's r*
Proof of convergent validity		
	($n_{min-max}$ = 494 - 560)	*($n_{min-max}$ = 498 - 501)*
General Health Perception *(self/proxy)*	.32	.37
CHQ-KINDL-Test Index *(self/proxy)*	.46	.50
FS-II-R *(proxy-report only)* *		
• General Health	*(.27)*	.42
• Interpersonal Functioning	*(.24)*	.36

Notes: All correlations are significant with at least p ≤ .01. Interpretation of correlation coefficients (Pearson r): r < 0.30: low; r = 0.30 - 0.60: moderate; r > 0.60: high. * Coefficients for correlations between the DCGM-12 self-report score and FS-II-R scores are suited in parenthesis because scores for computation were only available for the proxy-report version.

(II) HRQOL in particular: With the exception of the score for the "Everyday Functioning" subscale of the KINDL measure ($r_{S,S}$ =.24, $r_{P,P}$ =.25), the DCGM-12 displays moderate to high correlations with nearly all subscale scores and sum scores of the respective HRQOL measures (DUX-25, KINDL, PedsQL, VSP-A: $r_{SR(MI,MII)}$ = .34-.68; $r_{PR(MI,MII)}$ = .39-.73) for both, the self- and the proxy report version.

Table 36: A-priori estimations for correlation coefficients (Pearson r) for the total score of the short-form of the DISABKIDS condition generic module (DCGM-12; self-report/proxy-report) with various subscale and sum scores of different other multidimensional HRQOL assessments (sub-sample II from the DISABKIDS field study sample, n = 578)

DISABKIDS 12-item short-form (DCGM-12)	**Self-report version**	**Proxy-report version**
	Pearson's r	*Pearson's r*
Association with other generic multidimensional health-related quality of life measures		
PedsQL *(United Kingdom)*	*($n_{min-max}$ = 56-59)*	*($n_{min-max}$ = 48-50)*
• Physical	**.73**	.56
• Emotional	**.61**	**.64**
• Social	**.74**	**.72**
• School	.53	.59
DUX *(Netherlands, Sweden)*	*($n_{min-max}$ = 238)*	*($n_{min-max}$ = 134)*
• Physical	**.52**	**.44**
• Emotional	**.51**	.35
• Social	.43	.31
• Home	.42	.35
• Sum	**.55**	**.42**
KINDL-R *(Austria, Germany, Greece)*	*($n_{min-max}$ = 191-223)*	*($n_{min-max}$ = 178-186)*
• Physical Well-Being	.36	.30
• Emotional Well-Being	**.43**	**.42**
• Self-Esteem	.29	.38
• Family	.28	.23
• Friends	**.49**	**.54**
• Everyday Functioning	.17	.19
• Total	**.42**	**.44**
• Disease Module	**.57**	**.60**

Notes: Interpretation of correlation coefficients (Pearson r): r < 0.30: low; r = 0.30 - 0.60: moderate; r > 0.60: high. Highest correlation coefficients for each of the analysed questionnaires are marked in **bold**.

5.3.5.6 Divergent validity

Divergent validity was assessed exploring association with different subscale scores and the "total difficulties" sum score of the SDQ questionnaire. As the SDQ assess childrens' "strenghts and difficulties" as seen from the perspective of a related proxy, all difficulty facet scores were expected to be negatively associated. Except for a positive correlation of the DCGM-12 score with the "Prosocial" subscale score ($r_{PR(MI,MII)}$ = .17), all other coefficients were indeed negatively correlated, reaching low to moderate values ($r_{PR(min\text{-}max)}$ = -.11 to - .38; see Table 37 for detailed information).

Table 37: A-priori correlation coefficients (Pearson r) between total score of the short-form of the DISABKIDS condition generic module (DCGM-12; self-report/proxy-report) and SDQ scores (sub-sample II from the DISABKIDS field study sample, n = 578)

DISABKIDS 12-item short-form (DCGM-12)	**Self-report version**	**Proxy-report version**
	Pearson's r	*Pearson's r*
Proof of divergent validity		
SDQ *(proxy-report only)* *	*($n_{min\text{-}max}$ = 259 - 260)*	*($n_{min\text{-}max}$ = 255 - 256)*
• Prosocial **	*(.11)*	.17
• Emotional Symptoms	***(- .33)***	**- .38**
• Conduct Problems	*(- .14)*	- .15
• Hyperactivity / Inattention	*(- .14)*	- .11
• Peer Problems	*(- .30)*	- .31
• Total Difficulties	***(- .31)***	**- .32**

Notes: All correlations are significant with at least p ≤ .01. Interpretation of correlation coefficients (Pearson r): r < 0.30: low; r = 0.30 - 0.60: moderate; r > 0.60: high. * Coefficients for correlations between DCGM-12 self-report score and SDQ scores are suited in parenthesis because scores for computation were only available for the proxy-report version. ** According to research, coefficients for correlation between DCGM-12 total score and the "Prosocial" facet score were assumed to be positively directed, although for this subscale of the SDQ measure exclusively. Nevertheless, it was decided to report it beside of the other dimensions of the SDQ measure for reasons of readability, even this indicator of association in particular could be replaced within the convergent validity section of the table above for (methodo)logical reasons. Highest correlation coefficients for the SDQ are marked in **bold**.

5.3.5.7 Known-groups validity and usefulness

Discriminant (known-groups) validity: On a statistically significance level, the self-report version of the DCGM-12 discriminates between differences in the impairment of HRQOL in children and adolescents with respect to core socio-demographic (age group, gender: low effect size), socio-economic (family wealth, parental education status: low to nearly moderate effect sizes) and clinical (severity of condition: moderate to high effect sizes) indicators (see Table 38). This holds also for the proxy-version of the short-form on a descriptive level of observation, but does not reach statistical significance for gender and both indicators of socio-economic status.

Table 38: A-priori estimated discriminate validity of the short-form of the DISABKIDS condition generic module (DCGM-12; self-report/proxy-report) according to different variables (gender, age group, family wealth, parental education, clinical global impression on severity of condition) (sub-sample II from the DISABKIDS field study sample, n = 578)

Proof of discriminant / known groups validity		**DISABKIDS 12-item short-form**	
		Self-report	**Proxy-report**
		M ± SD	*M ± SD*
Gender	• Boys	77.65 ± 17.67	74.61 ± 18.26
(H1: QOL $_{Boys}$ > QOL $_{Girls}$)	• Girls	74.65 ± 18.04	73.05 ± 17.63
df (F / p)	*1*	*(3.99 / < 0.05)*	*(0.94 / n.s.)*
Age group	• 8 - 12 years	76.83 ± 17.86	75.26 ± 17.34
(H1: QOL $_{Younger\ (8\text{-}12)}$ > QOL $_{Older\ (13\text{-}16)}$)	• 13 - 16 years	76.66 ± 16.93	72.62 ± 18.20
df (F / p)	*1*	*(0.12 / n.s.)*	*(2.45 / n.s.)*
Family Wealth (FAS) *	• High	79.22 ± 16.76	75.60 ± 17.22
(H1: QOL $_{High}$ > QOL $_{Medium}$ > QOL $_{Low}$)	• Medium	74.74 ± 17.98	72.41 ± 18.11
	• Low	71.29 ± 17.67	73.22 ± 19.25
df (F / p)	*2*	*(6.01 / < 0.01)*	*(1.76 / n.s.)*
Parental Education Status **	• High	80.39 ± 16.46	76.25 ± 17.82
(H1: QOL $_{High}$ > QOL $_{Medium}$ > QOL $_{Low}$)	• Medium	76.27 ± 18.66	73.08 ± 18.66
	• Low	76.00 ± 16.93	73.69 ± 18.19
df (F / p)	*2*	*(2.43/ n.s.)*	*(1.16 / n.s.)*
Clinical Global Impression ***	• Mild	81.14 ± 16.42	77.79 ± 16.89
(H1: QOL $_{Mild}$ >QOL $_{Moderate}$ >QOL $_{Severe}$)	• Moderate	75.60 ± 15.64	73.79 ± 16.36
	• Severe	65.75 ± 17.58	60.89 ± 20.10
df (F / p)	*2*	*(16.31 / < 0.001)*	*(14.69 / < 0.001)*

Notes: Analyses were performed using ANOVA tests. M: mean; SD: standard deviation; p: significance level.* Composite score of the "Family Affluence Scale" (FAS) was calculated according to Boyce et al. (2006) using a 3-point ordinal scale with following scoring algorithm: "low" affluence = 0-2; "medium" affluence = 3-5; "high" affluence = 6-9. ** Parental education status (PES) was coded as follows: "low" PES = no/less/pre-/primary education, first/second stage of basic education; "medium" PES = upper and post secondary education (non-tertiary education); "high" PES = different stages of tertiary education. *** Clinician's judgement on a 3-point ordinal scale.

Efffect sizes: In comparison, the CHQ-KINDL index reaches higher effect sizes for discrimination of groups with respect to socio-demographic variables, especially for different age groups. Conversely, effect sizes for the DCGM-12 score are higher for discrimination of groups with respect to socio-economic variables, and in particular regarding clinical variables. Effect sizes differ substantially for most variables, except for gender.

Table 39: **A-priori estimated effect sizes (Cohen's d) of the short-form of the DISABKIDS condition generic module (DCGM-12) and the CHQ-KINDL Index (self-report/proxy-report) according to different variables (gender, age group, family wealth, parental education, clinical global impression on severity of condition) (sub-sample II from the DISABKIDS field study sample, n = 578)**

	Self-report versions			***Proxy-report versions***		
	DCGM-12	***CHQ-KINDL Index***		***DCGM-12***	***CHQ-KINDL Index***	
	d	*d*	Δ_d	*d*	*d*	Δ_d
Socio-demographic Variables						
Gender						
$QOL_{Boys} > QOL_{Girls}$	.17	.20	- .03	.09	.05	+ .04
Age group						
$QOL_{Younger\ (8\text{-}12)} > QOL_{Older\ (13\text{-}16)}$	.00	.26	**- .26**	.15	.23	- .08
Socio-economic Variables						
Family Wealth (FAS) *						
$QOL_{High} > QOL_{Medium}$	.26	.15	+ .11	.18	.17	+ .01
$QOL_{High} > QOL_{Low}$	.47	.23	**+ .24**	.22	.22	+/- .00
Parental Education Status **						
$QOL_{High} > QOL_{Medium}$	.23	.08	+ .15	.17	.12	+ .05
$QOL_{High} > QOL_{Low}$	.26	(.01)	**+ .27**	.10	(.10)	**+ .20**
Clinical Variables						
Clinical Global Impression ***						
$QOL_{Mild} > QOL_{Moderate}$	.35	.09	**+ .26**	.24	.01	**+ .23**
$QOL_{Mild} > QOL_{Severe}$	**.93**	.26	**+ .67**	**.97**	.11	**+ .86**

Notes: Interpretation of effect sizes (Cohen's d; Cohen, 1992): 0.20: small; 0.50: medium; 0.80: large. Medium or high effect sizes (d ≥ .50) as well as meaningful absolute effect size differences ($\Delta_d \geq .20$) are indicated in **bold**. d = effect size. Δ_d = effect size difference. * Composite score of the "Family Affluence Scale" (FAS) was calculated according to Boyce et al. (2006) using a 3-point ordinal scale with following scoring algorithm: "low" affluence = 0-2; "medium" affluence = 3-5; "high" affluence = 6-9. ** Parental education status (PES) was coded as follows: "low" PES = no/less/pre-/primary education, first/second stage of basic education; "medium" PES = upper and post secondary education (non-tertiary education); "high" PES = different stages of tertiary education. *** Clinician's judgement on a 3-point ordinal scale.

Psychometric performance of the short-form vs. long-form measure: Table 39 provides an overview on discriminative potential of the DCGM-12 score as compared to the DCGM-37 score. No meaningful absolute effect size differences could be identified, referring to the critical value of $\Delta_d \geq .20$ as lower boundary for small effect size for meaningful differences. Thus, substantial comparability can be assumed with respect to known-groups validity.

Table 40: **A-priori estimated effect sizes (Cohen's d) of the short-form (DCGM-12) and the long-form (DCGM-37) of the DISABKIDS condition generic module (self-report/ proxy-report)) according to different variables (gender, age group, family wealth, parental education, clinical global impression on severity of condition) (sub-sample II from the DISABKIDS field study sample, n = 578)**

	Self-report versions			***Proxy-report versions***		
	DCGM-12	***DCGM-37***		***DCGM-12***	***DCGM-37***	
	d	*d*	Δ_d	*d*	*d*	Δ_d
Socio-demographic Variables						
Gender						
$QOL_{Boys} > QOL_{Girls}$	.17	.14	+ .03	.09	.01	+ .08
Age group						
$QOL_{Younger\ (8\text{-}12)} > QOL_{Older\ (13\text{-}16)}$	.00	.09	- .09	.15	.10	+ .05
Socio-economic Variables						
Family Wealth (FAS) *						
$QOL_{High} > QOL_{Medium}$	.26	.27	- .01	.18	.21	- .03
$QOL_{High} > QOL_{Low}$	.47	.47	.00	.22	.23	- .01
Parental Education Status **						
$QOL_{High} > QOL_{Medium}$	.23	.23	.00	.17	.16	+ .01
$QOL_{High} > QOL_{Low}$	.26	.37	- .13	.10	.15	- .05
Clinical Variables						
Clinical Global Impression ***						
$QOL_{Mild} > QOL_{Moderate}$	.35	.38	- .03	.24	.30	- .06
$QOL_{Mild} > QOL_{Severe}$	**.93**	**.93**	.00	**.97**	**.88**	+ .09

Notes: Interpretation of effect sizes (Cohen's d; Cohen, 1992): 0.20: small; 0.50: medium; 0.80: large. Medium or high effect sizes (d ≥ .50) are indicated in **bold**. d = effect size. Δ_d = effect size difference. * Composite score of the "Family Affluence Scale" (FAS) was calculated according to Boyce et al. (2006) using a 3-point ordinal scale with following scoring algorithm: "low" affluence = 0-2; "medium" affluence = 3-5; "high" affluence = 6-9. ** Parental education status (PES) was coded as follows: "low" PES = no/less/pre-/primary education, first/second stage of basic education; "medium" PES = upper and post secondary education (non-tertiary education); "high" PES = different stages of tertiary education. *** Clinician's judgement on a 3-point ordinal scale.

Table 41: A-priori estimations for selected descriptive and psychometric composite scale characteristics of the short-form of the DISABKIDS condition generic module (DCGM-12 self-report/proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)

Descritive and psychometric characteristics		Self-report	
		DCGM-12	**DCGM-37**
Descritive characteristics			
Transformed raw score (0 – 100)	*M (SD)*	76.22 (17.89)	76.47 (15.07)
Mean raw score (1 – 5)	*M (SD)*		
Utilization: *Min-Max / Range*	*observed* *(possible)*	12-60 / 48 (12-60 / 48)	
Floor / Ceiling	*%*	- 0.80	- 0.78
Skewness			
Discriminatory Power	δ_G	.98	
Reliability			
Internal consitency	α	.85 [.86]	
Split-half reliability	*Guttman's*	.91 [.89]	
Internal consistencies of test-half's	α *(T1)*	.72 [.73]	
	α *(T2)*	.71 [.75]	
Correlation between test-half's	$r_{T1, T2}$	.83 [.80]	
Retest-reliabilty	$r_{t1, t2}$	.84	.89
	$ICC_{t1, t2}$	.84	.89
Self-Proxy-Aggreement	$r_{sr, pr}$	.64	.66
	$ICC_{sr, pr}$	.64	.66
Association ("criterion" validity)			
Item pool	$r_{sf, ov}$	.95 [.94]	
	$r^2_{sf, ov}$	.91 [.89]	
Score	$r_{sf, ov}$	.93	*(dito)*
	$r^2_{sf, ov}$	.87	*(dito)*

Notes: M = mean; SD = standard deviation; α = Cronbach's Alpha; r_{tt} = Test-retest reliability (Pearson correlation coefficient); ICC_{tt} = Test-retest reliability (Intraclass correlation coefficient: one-way, absolute agreement); $r_{s,p}$ = child-proxy-concordance (Pearson correlation coefficient); $ICC_{s,p}$ = child-proxy-concordance (Intraclass correlation coefficient: one-way, absolute agreement); $r_{sf,ov}$ = Association between DCGM-12 total score/item pool and DCGM-37 total score (Pearson correlation coefficient, not corrected for overlap); $R^2_{sf,ov}$ = determination coefficient for prediction of DCGM-37 average score by DCGM-12 total score/item pool. δ_G = Generalized delta coefficient. na = Coefficient not applicable for proxy-report version, as retest data do not include the DCGM item pool for the proxy version.

5.3.5.8 Performance across groups

Investigation of psychometric performance at the level of relevant sub-samples or groups (country-specific and condition specific sub-samples) was investigated using selected general indicators. The following analyses focus on discriminatory power of the short-form score as well as on reliability of the composite scale and at the level of single-items.

Discriminatory power in relevant sub-samples: Generalized delta coefficients reached values around $\delta_G \geq .95$, consistently indicating sufficient discriminatory power of the composite scale for both versions in all sub-samples for different languages ($\delta_{G\ (min\text{-}max)} = .95\text{-}.99$). For sub-samples of different chronic health conditions (main diagnosis) generalized delta coefficients reached values around $\delta_G \geq .95$. Discriminatory power for all specific subgroups referring to a country (language) x chronic health condition (main diagnosis) cross-table have not been computed, as due to small sample sizes this test would have not been useful.

Table 42: A-priori estimations for discriminatory power (generalized delta coefficient δ_G) of the short-form of the DISABKIDS condition generic module (DCGM-12; self-report/proxy-report) by language and chronic health condition (sub-sample II from the DISABKIDS field study sample, n = 578)

DISABKIDS 12-item short-form (DCGM-12)	Self-report version	Proxy-report version
	δ_G*	δ_G*
Language (country subsample/s included)(*)	*($n_{min\text{-}max}$ = 36-192)*	*($n_{min\text{-}max}$ = 67-304)*
• *Dutch* (Netherlands)	.97	.98
• *English* (United Kingdom)	.98	.98
• *French* (France)	.96	.95
• *German* (*) (Austria, Germany)	.98	.98
• *Greek* (Greece)	.98	.97
• *Swedish* (Sweden)	.98	.98
Chronic health condition (main diagnosis)	*($n_{min\text{-}max}$ = 36-192)*	*($n_{min\text{-}max}$ = 67-304)*
• *Asthma*	.97	.98
• *Arthritis*	.98	.98
• *Dermatitis*	.96	.95
• *Diabetes*	.98	.98
• *Cerebral Palsy*	.98	.97
• *Cystic Fibrosis*	.98	.97
• *Epilepsy*	.98	.98

Notes: δ_G = Generalized delta coefficient (Hankins, 2007).

Basic psychometric performance in relevant sub-samples: To account for small sample sizes, indicators less sensitive to lowered test power were used, focussing on reliability. At the composite score level indicators included Cronbach's alpha coefficient (internal consistency), split-half coefficient's (reliability of test-halfs), as well as intra-class correlation coefficients (ICC's) to indicate test-retest reliability and self-proxy aggreement. At the item-level coefficients for item-total correlations were investigated (see appendix).

5.3.5.9 *Relation to DISABKIDS modules*

As the DISABKIDS modular approach includes a variety of diagnostic tools in addition to chronic generic measures, relationships between the short-form version of the chronic-generic module and further DISABKIDS measures were investigated. Analyses focus on correlations with facet scores of the disease specific modules.

Table 43: A-priori estimated effect sizes (Cohen's d) of the short-form (DCGM-12) and the long-form (DCGM-37) of the DISABKIDS condition generic module (self-report/ proxy-report)) according to different variables (gender, age group, family wealth, parental education, clinical global impression on severity of condition) (sub-sample II from the DISABKIDS field study sample, n = 578)

Version		***Self-report versions***			***Proxy-report versions***		
		DCGM-12		***DCGM-37***	***DCGM-12***		***DCGM-37***
		r / r^2	:	r / r^2	r / r^2	:	r / r^2
Condition generic modules		*($n_{min-max}$ = 440-505)*			*($n_{min-max}$ = 401-448)*		
Smiley module	*Total score*	.565 / .319	<	.578 / .334	-		-
DCGM-37	*Independence*	**.751 / .564**	>	**.777 / .604**	**.822 / .676**	<	**.846 / .716**
	Emotion	**.807 / .651**	<	**.794 / .630**	**.836 / .699**	<	**.898 / .806**
	Exclusion	**.742 / .551**	>	**.804 / .646**	**.784 / 615**	<	**.844 / 712**
	Inclusion	.611 / .373	<	**.781 / .610**	.694/ .482	<	**.837/ .701**
	Limitation	**.774 / .599**	<	**.812 / .659**	**.798 / .636**	<	**.819 / .671**
	Medication	.691 / .477	>	.588 / .346	.646 / .410	>	.560 / .314
Disease specific modules		*($n_{min-max}$ = 36-192)*			*($n_{min-max}$ = 36-192)*		
Arthtritis	*Impact*	.701 / .491	>	.689 / .474	**.734 / .539**	<	**.769 / .591**
	Understanding	.150 / .225	<	.226 / .051	.115 / .013	<	.172 / .030
Asthma	*Impact*	.635 / .403	>	.573 / .328	.672 / .452	<	.676 / 457
	Worry	.632 / .399	<	.637 / .406	.423 / .179	>	.421 / .018
Cerebral Palsy	*Impact*	.592 / .350	<	.617 / .381	.522 / .272	<	.602 / .362
	Communication	.592 / .350	>	.588 / .346	.387 / .150	<	.445 / .198
Cystic Fibrosis	*Impact*	.214 / .046	<	.221 / .049	.333 / .111	<	.426 / .181
	Treatment	**.748 / .559**	<	**.771 / .594**	**.899 / .808**	>	**.872 / .760**
Diabetes	*Impact*	.700 / .490	<	**.730 / .533**	.589 / .347	<	.593 / .352
	Treatment	.572 / .327	<	.579 / .335	**.742 / .548**	<	**.707 / .500**
Epilepsy	*Impact*	.645 / .416	<	.690 / .476	.638 / .407	>	.515 / .265
	Social	**.717 / .514**	<	**.722 / .523**	**.806 / .650**	>	**.790 / .624**
Dermatitis	*Impact*	.632 / .399	>	.530 / .281	.288 / 083	>	-.235 / .055
	Stigma	.544 / .300	<	.608 / .370	**.759 / .576**	>	.560 / .314

Notes: Common variance between scores of more than 50% ($r^2 \geq .50$) is indicated in **bold**. R = correlation coefficient; r^2 = determination coefficient.

5.4 Test characteristics – Summary sheet

The following section gives a summary of the preliminary results of the DISABKIDS 12 item measures' (DCGM-12) psychometric performance and points out further issues of relevance regarding applying the measure:

- *Description:* The DCGM-12 is a short-form developed from the DCGM-37. It provides the possibility of assessing HRQOL in children and adolescents with different chronic health conditions in a more economic way. This short-form was derived using the conceptual background of the DCGM-37.
- *Purpose:* The DCGM-12 aims to assess HRQOL of children and adolescents (8-16 years) with chronic health conditions (e.g. asthma, diabetes, epilepsy, etc.). Thus, the short-form measures facets of HRQOL that pertain to specific circumstances of being young and in particular of having a severe chronic condition ("CRQOL"/"CGQOL"). The DCGM-12 measure should be applicable in different national and cultural contexts.
- *Versions:* Two versions of the short-form of the DCGM-12 are available: as a self-report measure and a proxy-report measure. Additionally, a *computer-assisted version* of the *paper-and-pencil version* of the DISABKIDS chronic generic module is also available.
- The DCGM-12 measure consists of 12 Likert-scaled items selected from the DCGM-37, which were originally assigned to three domains ("mental", "social", "physical") and six facets ("independence", "emotion", "exclusion", "inclusion", "limitation", "medication"), respectively. The items of each domain and facet are combined to produce a global score of the short-form measure. Every facet and every domain of the original measure (DCGM-37) is represented with the same frequency of items. The inclusion of all domains was based on the conceptual model of three higher-order dimensions, following an age-appropriate and condition generic diagnostic approach to CGQOL assessment. Thus, the DCGM-12 was developed in a strictly representative ("symmetric") way, according to the structure model and the measurement model of the DCGM-37.
- *Population:* The DCGM-12 measure can be administered to children and adolescents from 8 to 16 years (with the 5 item DISABKIDS smiley measure, an additional HRQOL measure for children aged 4 to 7 years is also provided within the DISABKIDS modular system).
- *Scoring:* The DCGM-12 consists of 12 items which are scored on a 5-point Likert-like scale ranging from "1" to "5" with fixed response choise wordings. The *time frame* for responses refers to the last 2 weeks. Scores can be calculated for a general HRQOL score by simply adding all 10, respectively 12 items provided applicability of the 2 items (item 11 and item 12) assigned to the original DCGM-37 facet "treatment/ medication".

- *Time required* for filling in the DCGM-12 is approximately 2 to 5 minutes for the *self-report-version*, depending on intellectual and biological age of the child or adolescent and the severity and impact of hes or her condition. Time required for the *proxy-version* of the DCGM-12 is approximately 1 to 3 minutes.
- *Development:* In a first pragmatic step, it was decided to use and retain the conceptual structure of the DCGM-37 as a framework for the short-form development. This approach lead to some restrictions in item selection, especially in terms of selecting items in a so-called "structural representative" way. Thus, every domain and every facet of the original measure (DCGM-37) is represented in the short-form measure by an equivalent relative number of items per facet (with the facet "emotion" being the sole exception). Short-form construction was performed by different kinds of item selection criteria and various item selection strategies, including multivariate methods of test construction.
- *Reliability: Internal consistency* values (Cronbach`s Alpha) reach .85 (SR) and .89 (PR), *split-half reliability* values reach .90 (SR) and .94 (PR), respectively.
- *Validity:* Validity of the short-term measure can assumed, as it is indicated by its association with various age group specific measure of HRQOL (DUX, KINDL-R, PEDISQOL, VSP-A; *convergent validity*) and by its discrimination of groups, for example concerning different types of chronic health condtions or different levels of severity status (*discriminatory validity*).

Table 44: Items included in the short-form of the DISABKIDS chronic generic measure (DCGM-12)

Domains included	$\sum$ items	Facets included	$\sum$ items	Items	Possible range of raw score (Min, Max)
Mental	4	Independence	2	1+2	
		Emotion	2	3+4	
Social	4	Inclusion	2	5+6	One composite score
		Exclusion	2	7+8	48 (12-60)
Physical	4	Limitation	2	9+10	
		Medication*	2*	11+12*	

Notes: * "Applicable" items. $\sum$ = sum; Min = minimum; Max = maximum.

- *Cross-cultural applicability:* Cross cultural applicability is present, as the DCGM-12 was developed implementing a simultanous approach to test construction and a-priori validated using DIF anaylsis which did not detect language or country-specific biases (i.e., results show acceptable amount of DIF).

- *Languages:* The DCGM-12 measure is available in a variety of languages, including all language versions applied within the original project of the European DISABKIDS Group: Dutch, English, French, German, Greek, and Swedish. Translations into Italian and Spanish are in progress.
- *Weaknesses:* Initial results indicate a preferred use of a composite score based on 10 items (i.e., excluding the "applicable" items derived from the "medication/treatment" facet). Furthermore, as the DCGM-12 measure is a newly developed QOL assessment future research is needed to establish the robustness of the findings reported here.
- *Administration:* The scores achieved on the DCGM-12 measure represent a global screening assessment of childrens' or adolescents' HRQOL from the respondent's point of view (child/proxy). Results concerning the DISABKIDS field study (N = 1.152; see above) should be used as a preliminary reference for children and adolescents with chronic conditions until data from independent samples are available for the DCGM-12.
- *Suggested usage:* The DCGM-12 should be used whenever a short version is needed for economic reasons, e.g. in survey studies on HRQOL in children with chronic conditions, in survey studies on using child populations that include a filter item asking for a chronic condition, or in studies where different modules (generic measures, chronic generic measure, condition specific measure, smileys) need to be combined.
- *Origin:* The DISABKIDS chronic generic measure DCGM-37 (original long version; Schmidt et al., 2006-a; Simeoni et al., 2007).
- *(Cautionary note:* All validation data are a-priori estimations, thus they have so far not been independently tested. Results reported are based on data from the DISABKIDS field trial sample.)

5.5 DISABKIDS chronic generic measures: Condition generic quality of life (CGQOL) in specific age groups (children/adolescents)

This section is intended to provide a short overview of the general framework of the DISABKIDS chronic generic modules, as an addition to the examples given in chapter 1 of this thesis. It locates the DISABKIDS chronic generic measurement approach within the general framework of QOL research.

General information: The *European DISABKIDS project* was funded by the European Union (EU) and aimed at developping a modular system for assessment of subjective health status, health care needs, and QOL of children and adolescents with different chronic conditions, adapted for usage in a cross-cultural context (Bullinger et al., 2002-a, 2002-b). This project focussed on cross-cultural development and validation of the scale, implementing a *simultaneous approach* for test construction (Bullinger et al., 1996). The methodological steps applied by the DISABKIDS Group were shared by the related KIDSCREEN Group (2007). The modular DISABKIDS approach included the development of diagnostic facilities regarding health care services and different levels of QOL assessment (condition generic QOL and condition specific QOL), different age groups (4-7 vs. 8-16 years), and different perspectives (self- vs. proxy-report).

Original measure (DCGM-37): The „*DISABKIDS chronic generic module*“ (*DCGM-37*; The DISABKIDS Group Europe, 2006; Petersen et al., 2005; Schmidt et al., 2006-a; Simeoni et al., 2007) is an age-appropriate, condition generic as well as multidimensional measurement tool that assesses general subjective condition generic quality of life (CGQOL) of children and adolescents (8 to 16 years old) with chronic conditions (e.g. asthma, arthritis, diabetes, etc.). This module is a questionnaire comprising 37 items belonging to 6 facets. Each facet is comprised by 6 items each, with exception of the “emotions” facet, which includes 7 items. The measure results inone summary score. Within the conceptual model of the DISABKIDS Group, 2 out of the 6 facets are assigned to 1 out of 3 higher-ordered “domains”: (I) *“Psychological”* domain: (i) “Independence”, (ii) “Emotions”; (II) *“Social”* domain: (iii) „Social exclusion“, (iv) „Social inclusion“; (III) *“Physical”* domain: (v) „Physical limitations“, and (vi) „Treatment/ medication“. Items assigned to the “Treatment/Medication” facet are only applicable if the respondent receives treatment or takes medication. The score computation algorithm uses equal-interval scoring of response categories (Likert-type scale scoring). Psychometric performance of the DCGM-37 has been shown with respect to reliability in terms of internal consistency of the scales (α = .70-.87). The DCGM-37 module is available in all languages of the DISABKIDS project participants, with translations into more languages underway.

Short-form measure (DCGM-12): The short-form of the “*DISABKIDS chronic generic module*“ (*DCGM-12*; Muehlan et al., in prep.) is an age group adapted and condition generic measurement, assessing general subjective CGQOL of children and adolescents (8 to 16 years old) with

chronic conditions (e.g. asthma, arthritis, diabetes, etc.). This module is a questionnaire comprising 12 items (statements), which have to be answered with respect to their degree of appropriateness on a five-point Likert-like scale with a fixed response choice format. These 12 items were derived from the DCGM-37 by using a mixed approach, combining expert-based and statistical procedures. Originally it was aimed to design the DCGM-12 allowing for computing sum scores for different domains of CGQOL outlined in the DISABKIDS conceptual model (physical, mental, and social QOL domain). Finally, it is recommended to simply compute a composite score including all items. Psychometric performance of the 12 item measure is considered to be very good with respect to reliability in terms of internal consistency of the scale ($\alpha = .86$).

Index (not available): An index of the DCGM-37 is not available beside of the DCGM-12 yet. However, for the time being the DCGM-12 short-form measure can be referred to as an index as is also produces only one composite score. A question for future research to address is whether or not an additional index or medium-length short-form should be developed along the lines of the DISABKIDS (chronic-generic) conceptual model beside of the DCGM-12.

Additional notes: Within the modular tool set of the DISABKIDS conceptual framework, a variety of measurements have been developed to accompany the core generic questionnaires assessing QOL in children and adolescents with different chronic conditions (DCGM-37, DCGM-12). In addition, seven *disease specific quality of life assessments* (*"DSM"-measures*; asthma, arthritis, cerebral palsy, cystic fibrosis, dermatitis, diabetes, epilepsy; Baars et al., 2005), as well as a *child health care questionnaire on satisfaction, utilization, and needs* (*CHC-SUN*; Schmidt et al., 2008) are also available. Finally, a "smiley" measure with 5 items for CGQOL assessment in the very yound age groups (4-7 years) are exist (The DISABKIDS Group Europe, 2006; Chaplin et al., 2008).

6 Discussion

In order to retain sufficient information and gain efficiency, the particular challenge in developing a short-form of the DCGM-37 was to balance the need for a high degree of correspondence with the original measure ("external related" performance), with a sufficient degree on psychometric performance of the short measure itself regarding psychometric properties ("internal related" performance). It should be noted once more, that the above analyses were primary directed to construct a measure by means of retaining a strong representation of the structure and measurement model of the original scale.

6.1 Performance of the measure

6.1.1 "External related" performance

Evaluation of *external related performance* of the short-form focuses on objectives related to the original measure, such as conceptual representation, statistical association, psychometric comparability, and economy or efficiency, respectively.

- Regarding *representation*, the DISABKIDS short-form measure has been developed according to a rationale strategy, oriented towards the structural representation within the original measurement approach. Thus, every domain and every facet of the original measure (DCGM-37) is represented by the short-form measure; this also holds for the respective relative number of items per facet (except for facet "emotion"). The inclusion of all domains and facets was based on the conceptual model of the DIABKIDS chronic generic module, which includes six facets, three higher-order dimensions, and one global score (HRQOL). A lack of fit is caused by the conceptual assumption of linking the medication facet with "physical limitation" to comprise a "physical" domain. A question remaining is whether the facet is indeed an integral part of HRQOL or not.
- *Association* in terms of a correlation between the DCGM-12 total score and the DCGM-37 total score is high. Accordingly, the DCGM-12 total score and the item pool predict the DCGM-37 total score variance to a high degree. Correlation coefficients between the DCGM-12 score and various DCGM-37 facet scores range between r = .61 and r = .81, thus proportion of common variance ranges between 37% and 66%, with 33% of items per facet retained within the short-form measure.

- Meaningful *comparability* of scores with respect to scoring algorithm, score performance, and score interpretation is high: The scoring algorithm is the same, mean scores do not differ substantially, and performance with respect to reliability and validity issues is comparatively well.
- *Economy/efficiency:* As compared to the DCGM-37, a total number of 25 items were deleted in the short-form version, i.e. more than two-thirds (67.6%). The estimated patient response burden thus is lower. Item scaling and wording as well as response choice format are in line with the original scale. Answereing the DCGM-12 approximately requires only a half of the time usually needed to fill in the DCGM-37.

6.1.2 "Internal related" performance

Internal related performance of the short-form was demonstrated with respect to criteria of classical psychometric test theory, such as reliability and validity.

- With respect to the *internal structure* of the item pool, the 12-item version was not found to be unidimensional. Results from exploratory factor analysis, confirmatory factor analysis, as well as Rasch analysis all indicated 2 items derived from the medication facet to be measuring an aspect other than HRQOL. As the administration mode of the chronic-generic measure treats this factor as an "applicable" one (and not as a regular part), even in the original version this facet differs from regular facets due to its restricted general applicability. In addition to empirical reasons it can therefor be regarded as an additional tool.
- Focusing on *reliability*, the item pool was demonstrated to be sufficiently homogenous and furthermore the composite scale scores display stability. The extracted item pool was not strictly unidimensional, because items included were conceptually different and therefore impaired psychometric performance of the measure.
- Composite score display high *discriminatory power*, indicating a sufficient utilization of scale range (observed raw scores) concerning individual levels of HRQOL.
- Discriminant *validity* in terms of differentiating between known groups was demonstrated, as indicated by differences in magnitude of effect sizes as well as their relative power, as compared to other measures such as the CHQ-KINDL index. This result is of special relevance regarding the usefulness of this newly developed measure as compared to already existing measures. Magnitude of correlation coefficients between the short-form scores, and facet and sum scores of various generic measures assessing HRQOL suggests that the overall score covers the whole range of the underlying multifaceted construct. It thus satisfyingly balances the weighting of the different content areas included (physical, mental, and social issues on domain level).

6.2 Strengths of the measure

Beyond the proof of psychometric performance, a new measure complies with economy and efficiency, usability and usefulness, simplicity and practicability, etc. In addition, special advantages of the current, newly developed measure should be highlighted. These advantages are related to both, its general advantage in terms of being a “short-form” measure of an already existing assessment as well as to its special qualities themselves.

- *Economy/efficiency:* As compared to the DCGM-37 a total number of 25 items were deleted in the short-form version. Thus, the measure was reduced far more than two-thirds (67.6%). This resulted in a significant reduction of time required for filling in the DCGM-12 compared to the DCGM-37. Less than a half of time is needed for answering the DCGM-12. The short version nonetheless displays comparatively good psychometric performance, as indicated by its high degree of precision. Thus, this version is highly economic with respect to the ratio of time saved versus precision lost.
- *Simplicity/practicability:* The estimated patient response burden should be low. Item scaling and wording as well as response choice format are comparable with items of the original assessment. Scoring procedures are transferred from the “long-form” to the short-form measure, making scoring easy in terms of usability.
- *Representation of main content:* Although the final 12 item measure displays sufficient homogeneity each facet of the original measure is included with representing items (so-called “marker” items), referring to coefficients for item-total correlation at the facet level of the original “long-form” version.
- *Symmetric representation:* As item-facet assignment of the original “long-form” measure DCGM-37 is nearly balanced (with 5 facets comprised of 6 items and just one facet comprised of 7 items). The short-form measure represents a nearly complete symmetric representation of the original measurement approach with respect main content areas.
- *Splitting the test:* Moreover, including 2 items from every facet allows for computing test-half’s of the short-form, with the same amount of items, which are also equally balanced with one item derived from each facet with respect to the original measure. Preliminary results regarding internal consistencies of test-halfs (SR: .72/.71; PR: .79/.77), correlation of test-half’s (SR: .83; PR: .87), and split-half reliability (SR: .91; PR: .93) empirically support the usefulness of this approach. The performance of the 10 item index is comparable to the 37 item index, with even more homogeneos test-halfs (SR: .75/.73; PR: .82/.79). As a limitation, the paired items derived from each facet and domain does not measure exactly the same characteristic; thus, test-halfs are only comparable with respect to the general assessment of HRQOL.

6.3 Weaknesses of the measure

As a promising measure, it has some shortcomings, unsolved problems and remaining deficiencies regarding its development process, validation procedure, and psychometric performance. The *limitations* occurred are outlined below.

- *Problems with confirmatory fit:* Empirical data fitted the model to a sufficient degree only when excluding both "applicable" items derived from the "treatment/medication" facet. In addition, when specifying correlated residual variances of "paired" items of a common factor, the model fit improved significantly. Although this procedure does not yield a parsimonious model, this is a common practice within QOL research. For example, this procedure has already been applied to fitting the SF-12 structure model (e.g. Wilson et al., 2002). However, applying this procedure on fitting the model can at least be regarded as "controversial" (Büh-ner, 2006: p. 263).

- *Impaired Rasch scalability:* Governed by a statistical item extraction, the measure did not prove to be "unidimensional" in a strictly psychometric sense (Hattie, 1985). Although the measure displays homogeneity in terms of internal consistency (Green et al., 1977), limitations concerning the assumption of unidimensionality are indicated by its impaired Rasch scalability. Violation of this assumption primarily stem from including items derived from the "treatment/ medication" factor. Although Rasch homogeneity is usually seen as a prerequisite for calculating sum scores, some authors regard this assumption as too restrictive and recommend use of composite scores for sufficiently homogenous facets.

- *Inclusion of a second dimension:* With respect to health-related quality of life, medication does not only result in impairments caused by and related to condition or disease. Moreover, it is plausible that item wordings and items comprised to the "treatment/medication" factor share the main subject "medication", as opposed to other items of the measure which usually refer to "condition". This issue of increased covariance due to simple linguistic pairing (instead of true semantical pairing) has also been reported within the WHOQOL project (see Power, 2003). Moreover, it is of special relevance in the assessment of children, as within development, children utilize similarity of form rather than semantic linking for transfer and generalization in syntactic development (Ninio, 2006). Thus, developing semantic discri-mination of language is strongly associated with structral similarities of language.

- *Non-ordered response choice frequency patterns (medication items)*: Most medication items display impaired distribution characteristics in terms of non-ordered response choice frequency patterns. This is true for nearly all items of both versions of the original measure (self-and proxy-report). This problematic pattern is addressed in further refinement of the original measure, especially with regard to rethinking response choice options.

- *Strongly skewed items included*: A particular problem arises from the inclusion of item “Do you feel lonely because of your condition”. This item displayed high skewness (> 2). As skewness had not been explicitly defined as exclusion criterion for item selection, this item was not ommited. Reanalysing the extraction sub-sample confirmed the strongly skewed distribution of the item. It is more strongly skewed as compared to the a-priori validation sub-sample. This also holds true for the item characteristics of the original measure, based on the complete DISABKIDS field study sample (n = 1.153; see Simeoni et al., 2007), which included the combined sub-samples for extraction and a-priori validation. Item skewness causes problems in psychometric performance, especially regarding Rasch scalability of the measure, as outlined in detail above within the Rasch analysis results.
- *Item-total correlation for one item below critical value:* With respect to a critical value of about .40, one item (item 10) did not display sufficient item-total correlation. This item was problematic prior to inclusion, as it represents the marker item of the “inclusion” facet despite its low ranking in terms of its correlation with the DCGM-37 total score (low criterion validity). Despite of this discrepancy, further analyses showed that the item correlates sufficiently well for the proxy-report version and in various sub-samples of the self-report as well. Moreover, the critical values of (corrected) item-total correlations are defined from a pragmatic point of view. In the end including or excluding the items should not just depend on performance indicators. It has been suggested that expert consensus should be the final criterion concerning usefulness of items, even if psychometric perfomance indicates excluding the items due to psychometric concerns (Bühner, 2006; Yousfi, 2007).
- *No balance of protrait and contrait items:* Although the DCGM-12 included reversely scored items, “protrait” items (“positive” items, keyed in the same direction as the respective construct of interest, namely QOL) and “contrait” items (“negative” items, keyed in the opposite direction) are not balanced. This is also the case in the original version (DCGM-37). For the DCGM-37, most of the factors were defined by items exclusively representing just protrait or contrait items. Thus, true score varriance and method factor variance are potentially confounded.

6.4 Limitations of the study

In addition to weaknesses of the preliminary short-form measure in particular, some limitations concerning data material, study design, and applied analytic methodology are highlighted in the following section.

- *No independent validation sample:* A-priori validation was based on data from the long-form measure, in which the final item pool of the DCGM-12 is embedded. Thus, the short-form measure has so far not been tested using an independent sample, although this is a standard procedure for reporting measurements performance according to relevant literature on test economy, short-form development, and shortening of measures (Coste et al., 1997;

Hornke, 2006; Smith et al., 2000). For instance, in the current case position effects of the items could pose severe problems that in turn do not allow for relating the short-form to its original measure.

- *DCGM-37 already validated on same sample:* Moreover, the original "long-form" measure of the DISABKIDS chronic generic module (DCGM-37) had not been validated on an independent sample, either. Thus, neither the long-form nor the short-form measure has so far been cross-validated. This is a severe issue, as both measures have been preliminary validated based on the same data sample.
- *Comparability not adjusted for common error:* Comparability between short-form score performance and total score performance of the original measure was not adjusted for common error, as it was investigated using a set of 12 common items with data from the same sample. Although this practice has been crititized (Coste et al., 1997), there is no satisfactory alternative at hand (Smith & McCarthy, 2000).
- *Investigating criterion validity with a non-valid criterion*: A-posteriori analysis of the DISABKIDS chronic generic measurement approach indicated the "medication" facet as being different from the other facets representing HRQOL. Thus, inclusion of items assigned to the medication facet within the global score could be critizised. This raises the question of how meaningful a total score of the measure in general can be assumed to be.
- *Heterogeneity of sample:* The DISABKIDS field study sample comprises a variety of samples from seven countries with six languages, including children and adolescents with seven varying chronic conditions and different clinical severities of their respective conditions. Furthermore, the sample is neither balanced regarding equal group distributions nor regarding the seven disease groups (i.e. they do not apply equally to all seven countries). In addition to the heterogeneity of the sample, this unbalanced design poses problems in detecting subgroup variations. Confunding effects are not meaningfully addressed in a reliable and sufficient statistical manner.
- *Analysis of sample on aggregated level:* As a result of the heterogeneity of the sample, analysis was primarily done on aggregated level, combining sub-samples of different countries and disease groups. Although inclusion of these different groups was undertaken to account for cross-cultural development and a condition generic approach, it could cause severe problems, especially as the country by condition matrix is unbalanced.
- *Inconsistencies across different languages/cultures:* When examining psychometric performance at the level of language- and disease specific sub-samples, inconsistencies of psychometric performance were observable. As the number of respondents in sub-samples of the DISABKIDS field study sample rapidly decrease when using two-way tables with countries (language) and disease groups, it was not possible to clearly separate effects solely

resulting from small sample sizes (due to less power or larger variance), or from “true scores” as source of variance.

- *Missing investigation of latent classes:* No analyses of latent classes were conducted due to methodologogical problems such as the heterogeneity of the sample, the incomplete language x condition matrix etc. Therefore, screening for different types of underlying answering patterns (e.g. by applying latent class analyses or mixed Rasch models) remains a task for future investigations.

6.5 Recommendations for further improvement

Results of the performance of the measure so far provide important *recommendations for further improvement* of the short-form measure. Measurement development should always be regarded as work in progress in a dialectic process of assessment improvement and construct understanding (Smith & McCarthy, 1995). Issues of measurement refinement are suggested, divided into recommendations for (I) *basic improvement* and (II) *progressive improvement* of the (measurement approach) of the DISABKIDS chronic generic module in general as well as of its short-form measure in particular. Whereas “basic” improvement includes “small” refinement procedures that could easily be applied without changing the whole structure or measurement model of the DISABKIDS chronic generic module, progressive improvement results in rather invasive modifications, which could also affect the measurement approach of the original “long-form” measure (DCGM-37).

6.5.1 Basic improvement

Preliminary results indicate methodological limitations and shortcomings of the DCGM-12, which also provide important *recommendations for further improvement* of the short measure.

According to results from confirmatory factor anaylsis (CFA) and Rasch analysis, excluding items 11 and 12 from the score composition would increase the fit of the general and incremential fit indices (CFA) and of the Q-index statistics (Rasch model) of the remaining item pool for the self-report as well as for the proxy-report version. As both items were derived from the same facet of the DCGM-37 (“treatment/medication”), one can assume another factor comprising both items, providing additional variance from another source (Rost et al., 1999). Further analyses strongly support this suggestion. Thus, impairment of HRQOL by treatment in particular is not completely attributable to condition or disease related impairments, indicating special impairment of HRQOL resulting from treatment/medication, although treatment is strongly related to condition. Thus, the measure should be used as a 10 item index assessing HRQOL in particular. Consequently, the short-form could be renamed into DCGM-10+2. This subject of treatment-related QOL is also of relevance in more general terms, affecting conceptual,

technical, as well as empirical issues of QOL assessment and the DISABKIDS measurement approach to HRQOL in particular and will be outlined in detail below.

Another potential improvement results from the indication of *non-ordered thresholds*. This is the case for item 8, 11, and 12 of the proxy-report version and for all items of the child-report version. Removing item 11 and 12 for reasons of item misfit and disordered thresholds would result in a 10 item pool with ordered thresholds for the proxy-report version. For the remaining 10 items of the self-report version, first thresholds are consistently disordered. A recommendation would be to merge two relevant categories into just one response choice option, resulting in 4 answering categories replacing the original response choice format of 5. However, compromises then have to be made with respect to the comparability of assessment and with respect to the proxy-version and the original measure (DCGM-37).

6.5.2 Progressive improvement

To ensure the further use of data from the original DCGM measures, all refinements and improvements of these measures should include the same items. Thus, no new items, no other items, and no modified items should be included. In consequence, all versions of DCGM measures should represent the same main content areas.

Progressive improvement

- *Response format*: As already noted above, another recommendation would be to merge first both categories of the answering scale into just one response choice option, resulting in 4 answering categories replacing the original response choice format of 5. Re-analyses of the DCGM-37 should examine the necessity/usability of applying this modification also to the original version to refine this measure and to ensure comparability of the long and the short version at the level of response choice options and scaling procedures.
- *Conceptual model:* As a fundamental agreement of the DISABKIDS measurement approach, HRQOL assessment following this approach refers to the biopsychosocial model of health as defined by the WHO. Thus, in operational terms and based on a first level of differentiation of the conceptual model, the DISABKIDS approach includes “physical”, ”mental”, and “social” dimensions. “Functioning” or “functionality” (functional limitation or impairment as seen from the perspective of deficiencies) is treated within the DISABKIDS approach as a dimension of HRQOL that is relate to all other dimensions of QOL. Therefore, “functionality” refers to ressources and deficiencies of all facets of HRQOL, with a special emphasis on condition related consequences for daily life and not just its impact on subjective health status or physical symptoms in particular. This issue becomes evident looking at linkages between DISABKIDS main content and respective ICF categories included (see Fava et al., 2009).

- *Functionality of "medication" ("treatment") factor*: For several reasons (see above for details) pertaining to the measurement approach, the "medication" factor should be treated as a second dimension of HRQOL. This dimension could be denoted to as "medication-related" or in more general terms "treatment-related" quality of life" (MRQOL/TRQOL), referring to burdens and the impact of treatment or medication on respondents' subjective quality of life in particular. Such a conceptual modification can be seen as the most profound change regarding the original conception of the DISABKIDS core module. The author *strongly recommended* treating this factor as an additional *module* instead of as an integral aspect. In fact, this facet is not regarded a regular facet for reasons of its limited general applicability. Therefore, even despite empirical reasons, from a simple administrative perspective it is already almost considered an additional tool. There are examples in the literature that could serve as a template, e.g. the additional "disease specific" module of the KINDL measure (Ravens-Sieberer & Bullinger, 1998), which is not included in total score computations. As analyses of the overall psychometric performance of the short-form measure regarding its internal structure (consistency, factor structure/dimensionality, Rasch homogeneity) as well as further anaylses of the "long" version indicate, the "medication" facet clearly represents an aspect different from mere HRQOL. Thus, it should be treated as such, also with respect to its consequences.
- The *"independence" facet* of the DISABKIDS condition generic module represents a general facet of HRQOL. The main content of this facet deals with resources regarding feelings (mental domain), contact (social domain), and behaviour (physical domain). Impact of condition on various areas of one's (quality of) life is represented by "exclusion" (social domain), "limitation" (physical domain), and "impact" (mental domain). Thus, impact is not limited to the mental domain, although it refers to psychological processes such as feeling, thinking, evaluating. Rathermore, it concerns all domains (*see item content*). "Independence" is very similar to "functionality" in terms of the ICF.
- *Substitution of "medication" facet*: Substituting the "medication" facet within the physical domain of the DISABKIDS chronic generic conceptual model could be done by including further physical items. Lacking "positive" physical items could easily be replaced by "independence" items strongly related to aspects of physical limitation (e.g. "Are you able to do everything you want to do even though you have your condition?").
- *No global score* (or at least excluding the "medication" factor): As our results provide little empirical support for merging all items into one global score, at least items assigned to the "medication" facet should be excluded from such a global composite score.
- To *balance the measure conceptually*: In addition to the symmetric structure of the measure with respect to two factors for each domain, those two facets assigned to one common domain should be balanced in addition. Consequently, one facet should focus on positive

aspects of the respective dimension ("resources", "gains", or "potentials"), the other facet should focus on negative aspects of the respective domain ("deficiencies", "losses", or "burdens"). This distiction between positive and negative aspects of a dimension is not thought to be the same as protrait and contrait items. Both types of items should be used for measuring positive as wells negative issues of all respective QOL areas, as this is a prerequisite for distinguishing between variance resulting from quality of life (true score variance) and variance comprised of methodological aspects in particular.

- *Rearranging factorial structure by balancing "negative" vs. "positive" facets*: It is recommended to add one positive facet to the "physical" domain for reasons of balancing the measure with respect to its number of "protrait" and "contrait" facets as well as two paired facets assigned to each dimension. As a result, each domain would be comprised of two facets, one representing "resources" the other one focusing on "impairments".
- *Proposal for reduced index ("ultra-short version")*: With respect to a 6 item solution including 1 item per facet and 2 items for each domain, one item should be negatively formulated (focusing burden and impact of disease) and the other one should be a positively formulated (focusing resources). First results based on iterative, selective, and thus non-representative selection resulted in an item pool with alpha coefficent of .82, which is furthermore Rasch homogenous, and does not display collapsing thresholds. As already analysed and described within this thesis, a weakness of such an very short index is its lack of sufficient association with the long-form measure in terms of correlation between scores or an unsatisfactory prediction of the total score by the item pool (meaningful critical value: $r^2 \geq .90$).
- *Refinements on item-level with respect to cross-cultural differences*: Analyses of the measurement's psychometric performance across groups at the cross-cultural level display varying item characteristics between groups. Inital analysis focussed on (corrected) item-total correlation at the item level as well as on indicators of reliability (internal consistency, split-half reliability) and discriminatory power (generalized delta coefficient) at the aggregated level. Despite small sample sizes, on a very general level of observation, first results indicate deviant item performance with respect to condition espescially for the sub-sample suffering from diabetes.´

6.6 Concluding remarks

In summary, despite some methodological limitations, the DCGM-12 represents a methodologically sound short measure to assess HRQOL in children and adolescents with different chronic health conditions, both as self- and proxy-report version. Results suggest that the DCGM-12 is a satisfactory instrument in terms of reliability and validity, although the 10-item

solution is more appropriate and therefore recommended for suggested purposes, this score could easily been computed in addition, even if all items were included. Consequently, the newly developed short-form alternative can be named “DCGM-10+2”.

The short-form of the DCGM-37 consists of 12 Likert-type scaled items, each selected from the original long form measure DCGM-37 without any changes regarding wording or scaling. Items were originally assigned to six facets and further conceptually linked to three domains (“mental”, “social”, and “physical”) respectively, to assess different areas of HRQOL of importance in the subjective experience of children and adolescents with chronic conditions (aged 8 to 16 years). Items should be combined to produce a total score. As compared to the DCGM-37, the DCGM-12 score conceptually displays a good structural representation, statistically high association with the total score, and comparability regarding psychometric properties. Therefore, the DCGM-12 can serve as an economic and sound alternative to the original longer form for use in clinical surveys. Nevertheless, it is not recommendable for individual assessment, as the short-form provides an overall impression but not an differentiated profile of the personal HRQOL status. The DCGM-12 should be used whenever a short measure or an index is needed for economical reasons in survey studies on children and adolescents with chronic health conditions or on the general child population with a filter question identifying a chronic condition. As the short-form is an integral part of the DISABKIDS modular system, the DCGM-12 is likely to be of potential use in studies where different modules (condition generic measure, condition specific measures) need to be combined in order to meet such different needs as efficiency, feasibility, precision, or comparability.

Nevertheless, future studies should provide more data for supporting the reported estimations concerning psychometric properties of the DCGM-12, based solely on applications of the DCGM-12 in independent samples. In addition, implementation studies should provide comprehensive reference data for specific groups, especially for different language versions for reasons of cross-cultural comparability as well as for various chronic health condition diseases for reasons of meaningful clinical comparisons.

Finally, studies should also focus and report on the 10-item solution in particular, as this “version” (I) fits all respondents, (II) displays undamped psychometric performance (as compared to the 12-item solution especially with respect to dimensionality), (III) and is expected to be more resistant against potential refinements of the DISABKIDS chronic generic measurement approach in general.

7 (I) Summary (English)

Objective: This thesis describes the development and a-priori validation of a short-form (index) of a quality of life instrument for children and adolescents with chronic conditions, aged 8 to 16 years (DCGM-37). Its aim was to reproduce the main content areas of the DCGM-37 as closely as possible, while simultaneously accounting for a shorter test length and merged facets.

Material and methods: Two data sets were available, both provided by the European DISABKIDS project (pilot study sample, field study sample). After initial preparatory analyses, pilot study data were not used for further analysis, as item characteristics differed compared to the “original” items with respect to administration mode and answering distributions. Final data analysis was performed based on two sub-samples of the split DISABKIDS field study sample (n = 1.153) to independendly compute (I) item extraction and (II) a-priori validation. Item selection combined both rationale and statistical procedures and were as follows: (I) to represent the conceptual model, every domain and facet of the original measure was represented by the same number of items within the short-form (index); (II) multivariate methods of item selection (conditioned total permutation for maximizing item pool homogeneity and iterative application of regression models for calculating predictions of the original measure total score) were used for extracting a final item pool for the short-form measure. A-priori-estimates of psychometric performance at the single item and composite scale level were calculated based on classic and modern test theory.

Results and conclusion: Results indicated that the DCGM-12 is a reliable and valid instrument. Twelve items in the pool conceptually reflected the main domains of the DCGM-37 and reliably predicted its total score. The DCGM-12 can serve as an economic alternative to the DCGM-37 for use in clinical surveys. However, limitations regarding unidimensionality (referring to impaired Rasch scalability) need to be taken into considerations. Recommendations for further improvement and validation of the short-form performance are outlined and alternative models are proposed.

7 (II) Zusammenfassung (deutsch)

Fragestellung: Die vorliegende Arbeit dokumentiert die Konstruktion und a-priori psychometrische Prüfung einer Kurzform des DCGM-37, ein Instrument zur Erfassung der gesundheitsbezogenen Lebensqualität von Kindern und Jugendlichen mit chronischen Erkrankungen im Alter von 8 bis16 Jahren. Ziel der Kurzformentwicklung war es, alle relevanten Bereiche des Originalfragebogens in einem Itempool so gut wie unter den Bedingungen verkürzter Testlänge und aggregierter Facetten zu reproduzieren.

Daten und Methoden: Grundlage für die Analyse bildeten zwei Datensätze der europäischen DISABKIDS-Projektgruppe (Pilotstudie, Feldstudie). Der Datensatz der Pilotstudie erwies sich allerdings nach initialen Anaylsen aufgrund administrativ und empirisch deutlich abweichender Itemcharakteristiken als wenig brauchbar für das Anliegen. Daher wurde der Datensatz der DISABKIDS Feldstudie (n = 1.152) gesplittet, um die getrennte Berechnung von Itemselektion und a-priori Validierung zu ermöglichen. Zunächst wurden komparative Analysen zur Nützlichkeit unterschiedlicher Itemselektionsstrategien durchgeführt. Die finale Itemselektion kombinierte rationale (expertenbasierte) und psychometrische Ansätze: Um die Repräsentation des konzeptuellen Modells des Originals zu gewährleisten, wurde jede Domäne und jeder Faktor der Originalversion mit der gleichen Anzahl an Items in der Kurzform (Index) repräsentiert und die Selektion der Items an rational definierte Bedingungen gebunden. Multivariate Methoden der Itemselektion wurden neben univariaten Kriterien eingesetzt, um einen adequaten Itempool für die Kurzform zu extrahieren. Auf Grundlage dieser Ergebnisse wurde expertenbasiert die finale Selektion der Items realisiert. A-priori-Schätzungen der internen Struktur (Dimensionalität) des Itempools und der psychometrischen Güte der potenziellen Kurzform auf Einzelitemebene und auf Scoreebene wurden auf der Grundlage klassischer und moderner Testtheorie berechnet.

Ergebnisse und Schlussfolgerungen: Die Ergebnisse verweisen auf ein weitgehend reliables und valides Instrument. Die 12 Items des DCGM-12 repräsentieren alle Facetten des Originals und klären (unkorrigiert) etwa 90% der Varianz des Gesamtscores der Originalversion auf. Relativiert an Testlänge/Informationsverlust sind die psychometrischen Eigenschaften der Kurzform im Vergleich als ausgezeichnet und auf absolutem Niveau als gut bis sehr gut zu bezeichnen. Die Kurzform kann somit als eine ökonomische Alternative zum Einsatz des DCGM-37 in Surveys mit entsprechenden klinischen Populationen empfohlen werden (Kinder und Jugendliche im Alter von 8 bis 16 Jahren mit einer chronischen Erkrankung). Einschränkungen betreffen die interne Struktur des Itempools: Der Index erlaubt entgegen der ursprünglichen Konzeption keine direkte Profildarstellung mit expliziter Referenz auf die Domain-Ebene des DISABKIDS-Messansatzes. Zudem weist die als eindimensional konzipierte Kurzform auch Probleme hinsichtlich der Unidimensionalitätsannahme auf, wie dies durch faktorielle Analysen und die Verletzungen der Rasch-Modell-Annahmen indiziert wird. Aus

diesen Gründen wird die Nutzung der 10-Item Version empfohlen, die nicht diese Defizite aufweist, zumal diese Version für alle potenziellen Nutzer anwendbar erscheint, während die 12-Item-Lösung für ca. 15% der anvisierten Zielgruppe nicht komplett anwendbar ist. Aus diesen und weiteren Analysen ergeben sich Empfehlungen für die Weiterentwicklung des Fragebogens, die ausführlich dargestellt werden.

8 References

1. **Aaronson, N.K., Acquadro, C., Alonso, J., Apolone, G. et al.** (1992). International quality of life assessment (IQOLA) project. *Quality of Life Research, 1*, 349-351.

2. **Aaronson, N.K. ,Cull, A., Stein, K. & Sprangers, M**. (1994). The EORTC modular approach to quality of life assessment in oncology*. International Journal of Mental Health, 2*, 75-96.

3. **Baars, R.M., Atherton, C.I., Koopman, H.M., Bullinger, M., Power, M. & the European DISABKIDS Group** (2005). The European DISABKIDS project: development of seven condition-specific modules to measure health related quality of life in children and adolescents. *Health and Quality of Life Outcomes, 3*, 70.

4. **Barofsky, I., Erickson, P. & Eberhardt, M.** (2004). Comparison of a single global item and an index of a multi-item health status measure among persons with and without diabetes in the U.S. *Quality of Life Research, 13*, 1671-1681.

5. **Bortz, J.** (2004). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Berlin, Heidelberg, New York: Springer.

6. **Bowling, A.** (1997). *Measuring Health. A review of quality of life measurement scales* (2nd ed.). Buckingham: Open University Press.

7. **Bowling, A.** (2001). *Measuring Disease. A review of disease-specific quality of life measurement scales* (2nd ed.). Buckingham: Open University Press.

8. **Boyce, W., Torsheim, T., Currie, C. & Zambon, A.** (2006). The Family Affluence Scale as a measure of national wealth: validation of an adolescent self-report measure. *Social Indicators Research, 78*, 473-487.

9. **Brazier, J., Usherwood, T., Harper, R. & Thomas, K.** (1998). Deriving a preference-based single index from the UK SF-36 Health Survey*. Journal of Clinical Epidemiology, 51,* 1115-1128.

10. **Brooks, R., Rabin, R. & Charro, F. de** (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Amsterdam: Kluwer.

11. **Bühner, M.** (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktl. u. erw. Aufl.). München: Pearson.

12. **Bullinger, M.** (1991). Quality of life – definition, conceptualization and implications: a methodologist's view. *Theoretical Surgery, 6,* 143-146.

13. **Bullinger, M.** (1993). Indices versus profiles – advantages and disadvantages. In: S.R. Walker & R.M. Rosser (Eds.), *Quality of life assessment. Key issues in the 1990s* (pp. 209-220). Dordrecht, Boston, London: Kluwer.

14. **Bullinger, M.** (2003). International comparability of health interview surveys: An overview of methods and approaches. In A. Nosikov & C. Gudex (Eds.), *EUROHIS: developing common instruments for health surveys* (*Vol. 57*; pp. 1-11). Amsterdam: IOS Press.

15. **Bullinger, M., Anderson, R., Cella, D. & Aaronson, N.** (1995). Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. In S.A. Shumaker & R.A. Berzon (Eds.), *The international assessment of health-related quality of life* (pp. 83-91). New York: Oxford Rapid Communications.

16. **Bullinger, M., von Mackensen, S., Fischer, K., Khair, K., Petersen, C., Ravens-Sieberer, U., Rocino, A., Sagnier, P., Tusell, J. M., van den Berg, M. & Vicariot, M.** (2002). Pilot testing of the Haemo-QOL quality of life questionnaire for haemophiliac children in six European countries. *Haemophilia 8 (Supplement II)*: II-47-II-54.

17. **Bullinger, M., Power, M.J., Aaronson, N.K., Cella, D.F. & Anderson, R.T.** (1996). Creating and evaluating cross-cultural instruments. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed.). Hagerstown. MD: Lippincott-Raven.

18. **Bullinger. M., Petersen, C., Schmidt, S. & the European DISABKIDS Group** (2002-a). European Paediatric Health-Related Quality of Life Assessment. *Quality of Life Newsletter, 29,* 4-5.

19. **Bullinger, M. & Ravens-Sieberer, U.** (1995). Health-related quality of life assessment in children: a review of the literature. *European Review of Applied Psychology, 45,* 245-254.

20. **Bullinger, M. & Schmidt, S.** (2007). Cross-cultural quality of life assessment approaches and experiences from the health care field. In I. Gough & J.A. McGregor (Ed.), *Wellbeing in developing countries - from theory to research* (pp. 219-242). Cambridge: Cambridge University Press.

21. **Bullinger, M., Schmidt, S., Petersen, C. & the European DISABKIDS Group** (2002-b). Assessing quality of life of children with chronic health conditions and disabilities – a European approach. *International Journal of Rehabilitation Research, 25,* 197-206.

22. **Chan, K.S., Mangione-Smith, R., Burwinkle, T.M., Rosen, M. & Varni, J.W.** (2005). The PedsQL^{TM}: reliability and validity of the short-form generic core scales and asthma module. *Medical Care, 43,* 256-265.

23. **Chaplin, J., Koopman, H.M., Schmidt, S. & the DISABKIDS Group** (2008). DISABKIDS smiley questionnaire: the TAKE 6 assisted health-related quality of life measure for 4 to 7-year-olds. *Clinical Psychology & Psychotherapy, 15*, 173-180.

24. **Cohen, J.** (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

25. **Coste, J., Guillemin, F., Pouchot, J. & Fermanian, J.** (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50,* 247-252.

26. **Cronbach, L.J.** (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

27. **Cunny, K.A. & Perri, M.** (1991). Single-item vs. multi-item measures of quality of life. *Psychological Report, 69*, 127-130.

28. **Currie, C.E., Elton, R.A., Todd, J. & Platt, S.** (1997). Indicators of socio-economic status for adolescents: the WHO health behaviour in school-aged children survey. *Health and Education Research, 12*, 385-397.

29. **Derogatis, L.R.** (1992). *SCL-90-R. Administration, scoring, and procedures manual-II for the R(evised) version and other instruments of the Psychopathology Rating Scale Series*. Townson: Clinical Psychometric Research Inc.

30. **Derogatis, L.R**. (1993). *Brief Symptom Inventory (BSI). Administration, scoring, and procedures manual* (3d ed.). Minneapolis: National Computer Services.

31. **The DISABKIDS Group Europe** (2006). *The DISABKIDS Questionnaires – Quality of life questionnaires for children with chronic conditions.* Handbook. Lengerich: Pabst.

32. **Ellert, U., Thomas, C., Ravens-Sieberer, U., Kosinski, M., Bjorner & J., Ware J.** (2001). Using item response theory to improve QOL measures for children – first results from the child dynamic health assessment project. *Quality of Life Research, 10,* 203.

33. **Farquhar, M.** (1995). Elderly people definitions of quality of life. *Social Science and Medicine, 41*, 1439-1446.

34. **Fava, L., Muehlan, H.M. & Bullinger, M.** (2009). Linking the DISABKIDS modules for health- related quality of life assessment with the international classification of functioning, disability and health. *Disability & Rehabilitation*, *20*, 1-12

35. **Ferguson, G.A**. (1949). On the theory of test discrimination. *Psychometrika, 14*, 61-68.

36. **Franke, G.H.** (1995). *SCL-90. Symptom Checkliste von Derogatis – Deutsche Version*. Manual. Göttingen: Beltz.

37. **Franke, G.H.** (2000). *BSI. Brief Symptom Inventory – Deutsche Version*. Manual. Göttingen: Beltz.

38. **Franke, G.H.** (2002). *SCL-90. Symptom Checkliste von Derogatis – Deutsche Version*. Manual (2. Aufl.). Göttingen: Beltz.

39. **Goodman, R.** (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38,* 581-586.

40. **Green, S.B., Lissitz, R.W. & Mulaik, S.A.** (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-838.

41. **Hankins, M.** (2007). Questionnaire discrimination: (re)-introducing coefficient *δ*. *BMC Medical Research Methodology, 7*, 19.

42. **Hardt, J., Egle, U.T., Kappis, B., Hessel, A. & Brähler, E.** (2004). Die Symptom-Checkliste SCL-27. Ergebnisse einer deutschen Repräsentativbefragung. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 54*, 214 - 223.

43. **Harfst, T., Koch, U., Aschoff, K. von, Nutzinger, D. O., Rüddel, J. & Schulz, H.** (2002). Entwicklung und Validierung einer Kurzform der SCL-90-R. *DRV-Schriften, 27*, 71-73.

44. **Harkness, J.** (Ed.) (1998). Cross-cultural survey equivalence. *ZUMA-Nachrichten Spezial 3*. Mannheim: ZUMA.

45. **Hattie, J.** (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

46. **Haywood, K.L., Garratt, A.M. & Fitzpatrick, R.** (2005). Quality of life in older people: A structured review of generic self-assessed health instruments. *Quality of Life Research, 14,* 1651-1668.

47. **Hornke, L.F. (2006).** Testökonomie (Test Economy). In M. Eid & F. Petermann (Hrsg.), *Handbuch der psychologischen Diagnostik* (S. 434-446). Göttingen: Hogrefe.

48. **Hurny, Chr., Bernhard, J., Coates, A., Peterson, H.F., Castiglione-Gertsch, M., Gelber, R.D., Rudenstam, C.-M., Collins, J., Lindtner, J., Goldhirsch, A. & Senn, H.-J.** (1996). Responsiveness of a single-item indicator versus a multi-item scale: assessment of emotional well-being in an international adjuvant breast cancer trial. *Medical Care, 34*, 234-248.

49. **The KIDSCREEN Group Europe** (2007). *The KIDSCREEN questionnaires. Quality of life questionnaires for children and adolescents*. Lengerich: Pabst.

50. **Klaghofer, R. & Brähler, E.** (2001). Konstruktion und teststatistische Prüfung einer Kurzform der SCL-90-R. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie, 7*, 115-124.

51. **Koopman, H.M., Theunissen, N.C., Vogels, A.G., Kamphuis, R.P. & Verris, G.H. (1998).** The DUC-25: a short-form questionnaire for measuring health related quality of life of children with a chronic illness [Abstract]. *Quality of Life Research, 7*, 619.

52. **Landgraf J.M., Maunsell E., Nixon-Speechley K., Bulllinger M., Campbell S., Abetz L. & Ware J.E.** (1998). Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. *Quality of Life Research, 7,* 433-445.

53. **Landgraf, J.M., Ravens-Sieberer, U. & Bullinger, M.** (1997). Quality of life research in children: methods and instruments. *Dialogues in Pediatric Urology, 20,* 5-7.

54. **Lipsey, M.W. & Wilson, D.B.** (2001). *Practical Meta-Analysis*. Sage.

55. **Mackensen, S.v., Bullinger, M. & The HAEMO-QOL Group (**2004). Development and testing of an instrument to assess the Quality of Life of Children with Haemophilia in Europe (Haemo-QOL). *Haemophilia, 12*, 384-392.

56. **Muehlan, H., Schmidt, S., Simeoni, M.-C., Petersen, C., Bullinger, M. & The European DISABKIDS Group** (in prep.). The DISABKIDS chronic generic short form measure (DCGM-12) – A tool for assessing quality of life in children and adolescents with different chronic conditions in cross-cultural context.

57. **Ninio, A.** (2006). *Language and the learning curve. A new theory of syntactic development*. Oxford: Oxford University Press.

58. **Nosikov, A. & Gudex, Cl.** (2003). *EUROHIS. Developing common intruments for health surveys*. Amsterdam: IOS Press.

59. **Petersen, C., Schmidt, S., Bullinger, M.** & The European DISABKIDS Group (2004) Brief report: development and pilot testing of a coping questionnaire for children and adolescents with chronic health conditions. *Journal of Pediatric Psychology, 29*, 635-640.

60. **Petersen, C., Schmidt, S., Bullinger, M. & The European DISABKIDS Group** (2006) Coping with a chronic pediatric health condition & health-related quality of life. *European Psychologist, 11*, 50-56.

61. **Petersen, C., Schmidt, S., Power, M. & The European DISABKIDS Group** (2005). Developing and pilot testing a chronic generic quality of life questionnaire for children with chronic conditions*. Quality of Life Research, 73*, 1065-1075.

62. **Pollak, E., Muehlan, H., Mackensen, S. von, Bullinger, M. & the Haemo-QOL Group** (2006). The Haemo-QOL Index. Developing a short measure for health-related quality of life assessment in children and adolescents with haemophilia. *Haemophilia, 12*, 384-392.

63. **Power, M.J.** (2003). Development of a common instrument for quality of life. In: A. Nosikov & C. Gudex (Eds.), *EUROHIS: developing common instruments for health surveys (Vol. 57*). Amsterdam: IOS Press.

64. **Power, M. J., Bullinger, M., Harper, A. & The WHOQOL Group** (1998). The World Health Organization WHOQOL-100: tests of the universality of quality of life in fifteen different cultural groups world-wide. *Health Psychology, 18,* 495-505.

65. **Power, M., Quinn, K., Schmidt, S. & The WHOQOL-OLD Group** (2005). Development of the WHOQOL-OLD module. *Quality of Life Research, 14*, 2197-2214.

66. **Ravens-Sieberer, U., Auquier, P., Erhart, M., Gosch, A., Rajmil, L., Bruil, J., Power, M., Duer, W., Cloetta, B., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J.** (2007). 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, 16*, 1347-1356.

67. **Ravens-Sieberer, U. & Bullinger, M.** (1997). Current approaches to measure the health-related quality of life in children. *Quality of Life Newsletter, 18,* 12-13.

68. **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.

69. **Ravens-Sieberer, U., Gosch, A., Abel, T., Auquier, P., Bellach, B.M., Bruil, J., Dür, W., Power, M., Rajmil, L. & The European KIDSCREEN Group** (2001). Quality of life in children and adolescents: a European public health perspective. *Sozial- und Präventivmedizin, 46,* 294-302.

70. **Robitail, St., Ravens-Sieberer, U., Simeoni, M.-C., Rajmil, L., Bruil, J., Power, M., Duer, W., Cloetta, B., Czemy, L., Mazur, J. Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J. & Auquier, P.** (2007). Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Quality of Life Research, 16*, 1335-1345.

71. **Rockwood, T.H.** (2007). Impact of fetal incontinence on quality of life. In: C. Ratto & G.B. Doglietto (Eds.), Fecal *incontinence, Diagnosis and treatment* (pp. 73-78). Milan, Berlin, Heidelberg, New York: Springer.

72. **Rose, M.** (2003). *Quality of life in chronic health conditions*. Habilitation thesis. Medical Department (Charité), Humboldt-University of Berlin [in German].

73. **Rost, J.** (2004). *Textbook test theory – test construction* (2nd., rev. and ext. ed.). Bern: Huber [in German].

74. **Rost, J., Carstensen, C. & von Davier, M.** (1999). Sind die Big-Five Rasch-skalierbar? *Diagnostica, 45*, 119-127.

75. **Sapin, C., Antoniotti, S., Simeoni, M.-C., Clément, A., El Khammar, M. & Auquier, P.** (2005-a). Shortening the VSP-A: Preliminary development of the VSP-A12, a 12 item short-form. *Quality of Life Research, 13*, 235-241.

76. **Sapin, C., Simeoni, M.-C., El Khammar, M., Antoniotti, S. & Auquier, P.** (2005-b). Reliability and validity of the VSP-A, a health-related quality of life instrument for ill and healthy adolescents. *Journal of Adolescent Health, 36,* 327-336.

77. **Schmidt, S. & Bullinger, M.** (2003). Current issues in cross-cultural quality of life instrument development. *Archives of Physical Medicine and Rehabilitation, 84 (Supplement II),* S29-S34.

78. **Schmidt, S. & Bullinger, M.** (2007). Cross-cultural quality of life assessment approaches and experiences from the health care field. In: Gough, I. & McGregor, A. (Eds.), *Wellbeing in Developing Countries* (pp. 219-241). Cambridge University Press.

79. **Schmidt, S., Debensason, D., Muehlan, H., Petersen, C., Power, M., Simeoni, M.-C., Bullinger, M. & the European DISABKIDS Group** (2006-a). The DISABKIDS generic quality of life instrument showed cross-cultural validity. *Journal of Clinical Epidemiology, 59,* 587-598.

80. **Schmidt, S., Mühlan, H, & Power, M.** (2006-b). The EUROHIS-QOL 8 item index. Psychometric results of a cross-cultural study. *European Journal of Public Health, 16*, 240-248.

81. **Schmidt, S., Thyen., U., Chaplin, J., Mueller-Godeffroy, E. & the European DISABKIDS Group (2007).** Cross-cultural testing of a child health care instrument on satisfaction, utilization, and needs. *Ambulatory Paediatrics. 7,* 374-382.

82. **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*, 393-400.

83. **Simeoni, M.-C., Schmidt, S., Muehlan, H., Debensason, D., Bullinger, M. & the European DISABKIDS Group** (2007). Field testing of a European quality for life instrument for children and adolescents with chronic conditions: The 37 DISABKIDS chronic generic module. *Quality of Life Research*, *16,* 881-893.

84. **Skevington S.M.** (2004). Advancing cross-cultural research on quality of life: Observations drawn from the WHOQOL development. *Quality of Life Research*, *11*, 135-144.

85. **Skevington S.M., Sartorius, N., Amir, M. & The WHOQOL Group** (2004). Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL instruments. *Social Psychiatry and Psychiatric Epidemiology, 39*, 1-8.

86. **Sloan, J.A., Aaronson, N.K., Capelleri, J.C., Fairclough, D.L. & Varricchio, C.** (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinical Proceedings, 77*, 479-487.

87. **Slocum-Gori, S., Zumbo, B.D., Michalos, A.C. & Diener, E**. (2009). A note on the Diemnsionaity of quality of life scales: an illustration with the satisfaction with life scale (SWLS). *Social Indicators Research, 92*, 489-496.

88. **Smith, G.T. & McCarthy, D.M.** (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7,* 300-308.

89. **Smith, G.T., McCarthy, D.M. & Anderson, K.G.** (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102-111.

90. **Stein, R.E.K. & Jessop, D.J** (1990). Functional Status II(R): A measure of child health status. *Medical Care, 28,* 1041-1055.

91. **Stewart, A.L., Sherbourne, C.D. & Brod, M.** (1996). Measuring health-related quality of life in older and demented populations. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed.; pp. 819-830). Hagerstown. MD: Lippincott-Raven.

92. **Steinbüchel, von N., Lischetzke, T., Gurny, M., Winkler, I. & The WHOQOL-OLD Group** (2005). Erfassung gesundheitsbezogener Lebensqualität älterer Menschen mit dem WHOQOL-BREF-Fragebogen. *Zeitschrift für Medizinische Psychologie, 14*, 13 - 23.

93. **Steyer, R. & Eid, R.** (1993). *Messen und Testen*. Berlin, Heidelberg, New York: Springer.

94. **Varni, J.W., Birwinkle, T.M., Seid, M. & Skarr, D.** (2003). The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambulatory Pediatrics, 3,* 329-341.

95. **Von Davier, M.** (1996). *WINMIRA V1.74 Users Manual. A program system for analyses with the Rasch Model, with Latent Class Analysis and with Mixed Rasch Model*. Kiel: Institute for Science Education.

96. **Ware, J.E.** (1993). *The SF-36 Health Survey – manual and interpretation guide*. Boston: Nimrod.

97. **Ware, J.E., Kosinski, M. & Keller, S.D.** (1996). A 12 item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care, 34,* 220-233.

98. **Ware, J.E., Kosinski. M., Dewey, J.E. & Gandek, B.** (1999). *How to score and interpret single-item health status measures: an manual for users of the SF-8 Health Survey* (Manual).

99. **Wilson, D.**B. (2001). *Effect size determination program*. Maryland: University of Maryland, College Park.

100. **Wilson, D., Tucker, G. & Chittleborough, C.** (2002). Rethinking and rescoring the SF-12. *Sozial- und Präventivmedizin, 47*, 172-177.

101. **The WHOQOL Group** (1998). Development of the World Health Organization WHOQOL-BREF Quality of Life assessment. *Psychological Medicine, 28*, 551-558.

102. **Youngblut, J.M. & Casper, G.R.** (1993). Single-item indicators in nursing research. *Research in Nursing and Health, 16*, 459-465.

103. **Zumbo, B.** (1999). *A handbook of the theory and methods of differential item functioning (DIF). Logistic regression modelling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa, ON: Directorate of Human Ressources Research and Evaluation, Department of National Defense.

9 List of tables

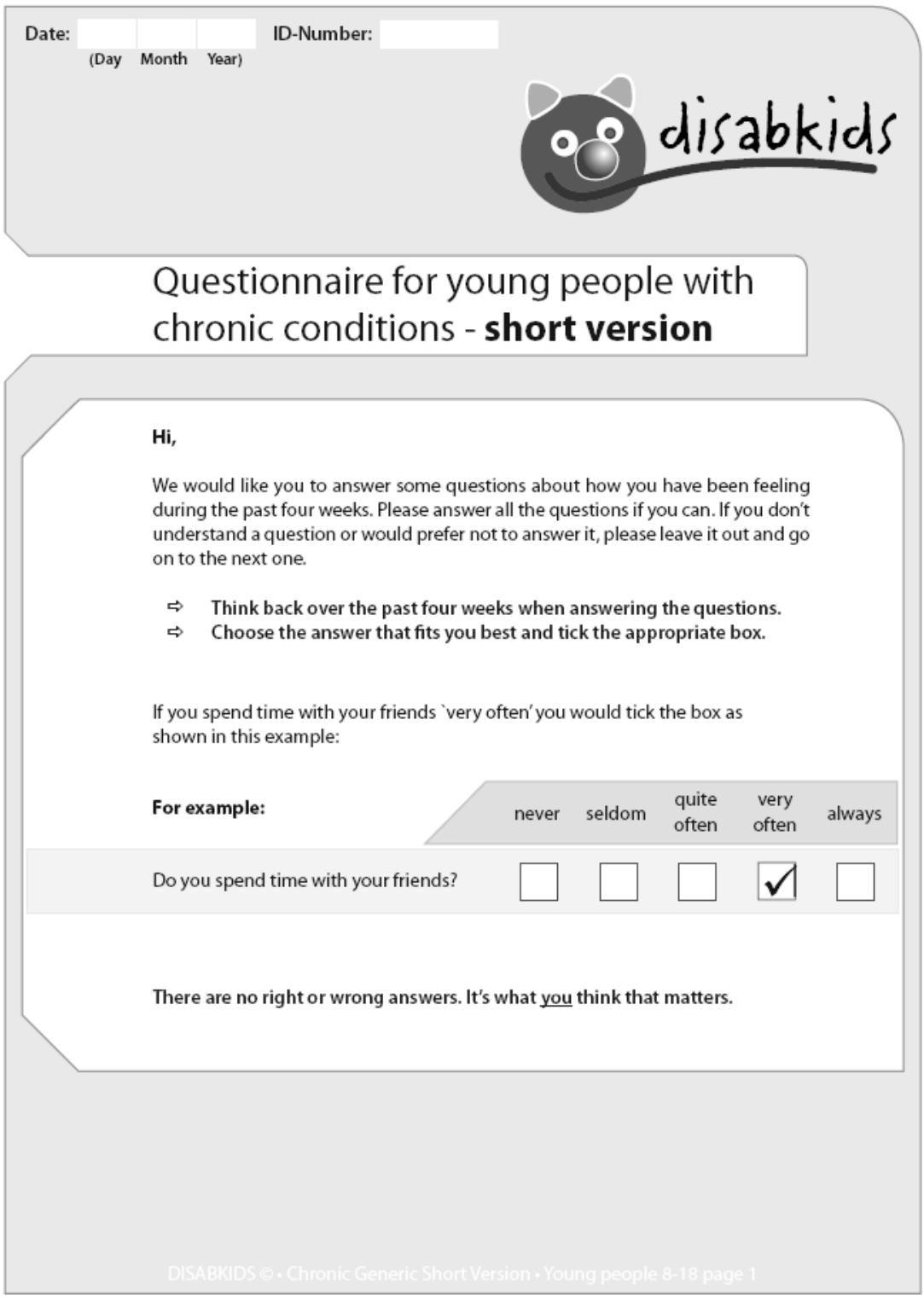
10 List of figures

Note on technical environment: Written in Word Office on Microsoft XP Home Edition 2001; fond style serial “Arial”, with varying fond sizes, standard text fond size 10 pts. Operating system installed and running on Fujitsu Siemens Computers, AMILO Series, Model M 1420 and on IBM ThinkPad R51.

Appendix

A.1 DISABKIDS condition generic short-form measure(s)

A.1a The DISABKIDS condition generic short-form measure – self-report version



Date: (Day Month Year) ID-Number:

disabkids

Questionnaire for young people with chronic conditions - **short version**

Hi,

We would like you to answer some questions about how you have been feeling during the past four weeks. Please answer all the questions if you can. If you don't understand a question or would prefer not to answer it, please leave it out and go on to the next one.

- ⇨ **Think back over the past four weeks when answering the questions.**
- ⇨ **Choose the answer that fits you best and tick the appropriate box.**

If you spend time with your friends `very often' you would tick the box as shown in this example:

For example:	never	seldom	quite often	very often	always
Do you spend time with your friends?	☐	☐	☐	☑	☐

There are no right or wrong answers. It's what you think that matters.

DISABKIDS © • Chronic Generic Short Version • Young people 8-18 page 1

(The DISABKIDS condition generic short-form measure – self-report version / continued)

Some questions about you

A. Are you male or female? ☐ female ☐ male

B. How old are you? ☐ years

C. Which condition do you have?

☐ asthma ☐ arthritis ☐ dermatitis

☐ cerebral palsy ☐ diabetes ☐ cystic fibrosis

☐ epilepsy ☐ other → Which? ☐

In some questions we use the word „condition". When you see condition, please think about what you filled in above.

About your life

Think about the past four weeks!

		never	seldom	quite often	very often	always
1.	Do you feel like everyone else even though you have your condition?	☐	☐	☐	☐	☐
2.	Are you free to lead the life you want even though you have your condition?	☐	☐	☐	☐	☐
3.	Is your life ruled by your condition?	☐	☐	☐	☐	☐
4.	Does your condition bother you when you play or do other things?	☐	☐	☐	☐	☐
5.	Are you unhappy because of your condition?	☐	☐	☐	☐	☐
6.	Does your condition get you down?	☐	☐	☐	☐	☐
7.	Do you feel lonely because of your condition?	☐	☐	☐	☐	☐

(The DISABKIDS condition generic short-form measure – self-report version / continued)

About your life

Think about the past four weeks!

	never	seldom	quite often	very often	always
8. Do you feel different from other children/adolescents?	☐	☐	☐	☐	☐
9. Do you think that you can do most things as well as other children/adolescents?	☐	☐	☐	☐	☐
10. Do your friends enjoy being with you?	☐	☐	☐	☐	☐

About your medical treatment

Do you take any medicine for your condition? (by medicine we mean tablets, cream, spray, insulin or any other medicine) yes ☐ no ☐

Think about the past four weeks!

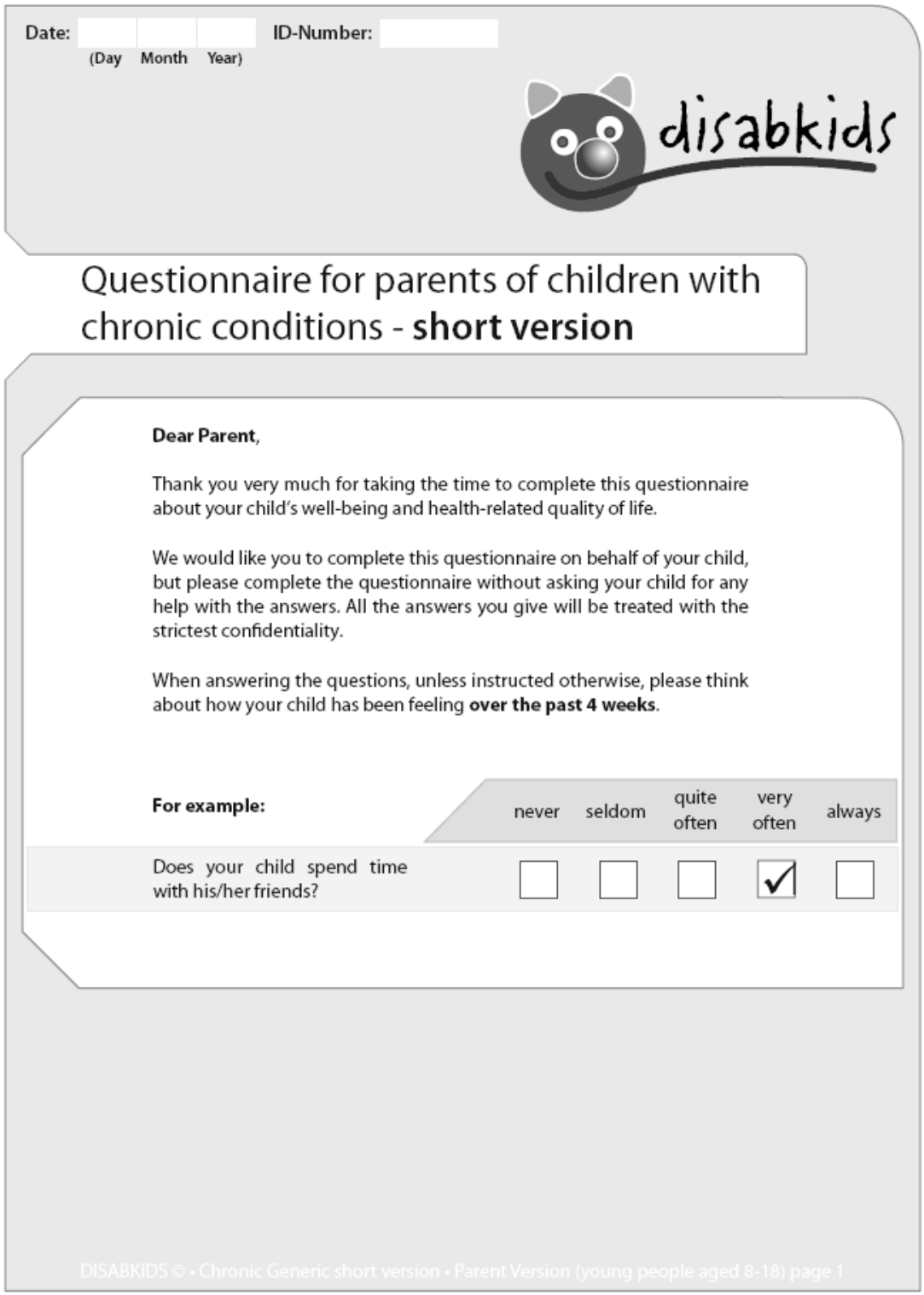
If yes, please fill in the following questions, if no, please skip this section.

	never	seldom	quite often	very often	always
11. Does taking medication bother you?	☐	☐	☐	☐	☐
12. Do you hate taking your medicine?	☐	☐	☐	☐	☐

Thank you for your assistance!

DISABKIDS © - Chronic Generic Short Version - Young people 8-18 page 3

A.1b The DISABKIDS condition generic short-form measure – proxy-report version



Date: (Day Month Year) ID-Number:

disabkids

Questionnaire for parents of children with chronic conditions - **short version**

Dear Parent,

Thank you very much for taking the time to complete this questionnaire about your child's well-being and health-related quality of life.

We would like you to complete this questionnaire on behalf of your child, but please complete the questionnaire without asking your child for any help with the answers. All the answers you give will be treated with the strictest confidentiality.

When answering the questions, unless instructed otherwise, please think about how your child has been feeling **over the past 4 weeks**.

For example:	never	seldom	quite often	very often	always
Does your child spend time with his/her friends?	☐	☐	☐	☑	☐

DISABKIDS © • Chronic Generic short version • Parent Version (young people aged 8-18) page 1

(The DISABKIDS condition generic short-form measure – proxy-report version / continued)

Some questions about your child

A. Is your child male or female? ☐ female ☐ male

B. How old is your child? ☐ years

C. What is your child's condition?

☐ asthma ☐ arthritis ☐ dermatitis

☐ cerebral palsy ☐ diabetes ☐ cystic fibrosis

☐ epilepsy ☐ other → Which? ☐

D. Who is filling in the questionnaire?

☐ mother ☐ father ☐ stepmother/father's partner

☐ stepfather/ mother's partner ☐ others → Who? ☐

In some questions we use the word „condition". When you see „condition", please think about what you filled in above.

About your child's life

Think about the last four weeks!

		never	seldom	quite often	very often	always
1.	Does your child feel like everyone else even though they have their condition?	☐	☐	☐	☐	☐
2.	Does your child feel free to lead the life they want even though they have their condition?	☐	☐	☐	☐	☐
3.	Does your child feel that their life is ruled by their condition?	☐	☐	☐	☐	☐
4.	Does your child's condition bother them when they play or do other activities?	☐	☐	☐	☐	☐

(The DISABKIDS condition generic short-form measure – proxy-report version / continued)

About your child's life

		Think about the past four weeks!				
		never	seldom	quite often	very often	always
5.	Does your child feel unhappy because of his/her condition?	☐	☐	☐	☐	☐
6.	Does your child's condition get him/her down?	☐	☐	☐	☐	☐
7.	Does your child feel lonely because of his/her condition?	☐	☐	☐	☐	☐
8.	Does your child feel different from other children/adolescents?	☐	☐	☐	☐	☐
9.	Does your child think that he/she can do most things as well as other children?	☐	☐	☐	☐	☐
10.	Does your child feel that their friends enjoy being with them?	☐	☐	☐	☐	☐

About your child's medical treatment

Does your child take any medicine for their condition? (by medicine we mean tablets, cream, spray, insulin or any other medicine) yes ☐ no ☐

If yes, please fill in the following questions, if no, please skip this section.		Think about the past four weeks!				
		never	seldom	quite often	very often	always
11.	Does taking medication bother your child?	☐	☐	☐	☐	☐
12.	Does your child hate taking his/her medicine?	☐	☐	☐	☐	☐

Thank you for your assistance!

A.2a: DISABKIDS chronic generic short-form score if "medication/treatment" is applicable (self-report) – DCGM-12-S-SF (DISABKIDS field study sample, N = 1.153)

DCGM-12-S SCORE SELF-REPORT		Total sample 8-12 (n = 461)		Total sample 13-16 (n =375)		Total sample Overall (n = 836)		Females 8-12 (n = 206)		Females 13-16 (n = 194)		Females Overall (n = 400)		Males 8-12 (n = 255)		Males 13-16 (n = 181)		Males Overall (n = 436)	
RS	**TRS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**
<=25	**27**	1	19	1	21	1	20	2	21	1	21	1	22	1	17	-	-	1	17
26	**29**	1	19	1	22	1	21	2	21	3	22	2	23	1	17	-	-	1	17
27	**31**	1	22	2	23	1	23	2	23	3	24	2	24	1	17	1	21	1	21
28	**33**	1	22	2	25	2	24	2	23	3	26	3	25	1	17	1	21	1	21
29	**35**	2	24	2	26	2	25	2	23	3	29	3	25	1	23	1	23	1	23
30	**38**	3	26	2	27	3	26	4	27	4	30	4	28	2	24	1	23	1	24
31	**40**	3	27	4	28	3	28	4	27	5	31	5	29	2	25	3	26	2	26
32	**42**	3	28	5	30	4	29	4	27	7	32	5	30	2	27	3	26	3	27
33	**44**	4	29	6	31	5	30	5	31	7	34	6	32	3	28	4	29	3	28
34	**46**	5	31	7	32	6	31	7	32	8	34	7	33	3	28	6	30	4	29
35	**48**	6	32	7	33	7	33	8	33	9	35	9	34	4	31	6	30	5	31
36	**50**	7	33	9	35	8	34	9	34	11	36	10	35	6	32	7	32	6	32
37	**52**	9	34	11	36	10	35	10	36	13	37	11	37	8	33	9	34	8	33
38	**54**	11	36	13	37	12	36	14	37	15	39	14	38	9	34	11	35	10	35
39	**56**	12	37	16	38	14	38	15	38	18	40	16	39	11	36	13	36	12	36
40	**58**	15	38	17	40	16	39	17	39	20	41	18	40	13	37	15	38	14	37
41	**60**	17	39	21	41	19	40	19	41	23	42	21	41	15	38	18	39	16	38
42	**63**	20	41	23	42	21	41	22	42	26	44	24	43	18	39	19	40	19	40
43	**65**	22	42	24	43	23	43	24	43	27	45	25	44	21	41	21	41	21	41
44	**67**	26	43	28	45	27	44	29	44	30	46	30	45	24	42	25	43	24	42
45	**69**	28	44	31	46	29	45	31	46	35	47	33	46	26	43	28	44	27	44
46	**71**	31	46	36	47	33	46	35	47	41	49	38	48	28	45	32	45	30	45
47	**73**	36	47	40	48	38	47	39	48	46	50	43	49	33	46	33	47	33	46
48	**75**	38	48	45	50	41	49	40	49	52	51	46	50	36	47	37	48	36	47
49	**77**	41	49	49	51	45	50	43	50	57	52	50	51	40	48	39	49	40	49
50	**79**	45	51	53	52	48	51	47	52	62	54	55	53	43	50	43	50	43	50
51	**81**	50	52	60	53	55	52	53	53	68	55	60	54	48	51	51	52	49	51
52	**83**	57	53	65	55	60	54	61	54	73	56	67	55	54	52	55	53	54	52
53	**85**	62	54	70	56	66	55	67	55	79	57	73	56	57	53	61	54	59	54
54	**88**	69	56	74	57	71	56	74	57	82	59	78	58	64	55	66	55	65	55
55	**90**	74	57	80	58	77	57	80	58	87	60	83	59	69	56	74	57	71	56
56	**92**	78	58	84	60	81	59	85	59	91	61	88	60	73	57	77	58	75	58
57	**94**	82	59	89	61	85	60	89	60	93	62	91	61	77	59	85	59	80	59
58	**96**	90	61	92	62	91	61	94	62	94	64	94	63	87	60	90	61	88	60
59	**98**	94	62	96	63	95	62	97	63	97	65	97	64	91	61	95	62	93	61
60	**100**	100	63	100	65	100	64	100	64	100	66	100	65	100	62	100	63	100	63

Notes: *Calculation of scores was restricted to complete cases. RS = raw score; TRS = transformed raw score (range 0-100); PR = percentile: TS = T-score (M = 50, SD = 10).*

A.2b: DISABKIDS chronic generic short-form score if "medication/treatment" is not applicable (self-report) – DCGM-12-S-SF (V-10) (DISABKIDS field study sample, N = 1.153)

DCGM-12-S (V-10) SCORE (V-10) SELF-REPORT		Total sample 8-12 (n = 514)		Total sample 13-16 (n =425)		Total sample Overall (n = 939)		Females 8-12 (n = 231)		Females 13-16 (n = 213)		Females Overall (n = 444)		Males 8-12 (n = 283)		Males 13-16 (n = 212)		Males Overall (n = 495)	
RS	**TRS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**
<= 20	**<= 25**	1	14	1	15	1	14	0	14	1	17	1	16	1	13	1	13	1	13
21	**27.5**	1	14	1	15	1	14	0	14	1	17	1	16	1	13	1	13	1	13
22	**30**	1	20	1	21	1	20	0	14	1	23	1	22	1	19	1	19	1	19
23	**32.5**	1	21	1	23	1	22	1	22	2	24	2	23	1	19	1	19	1	19
24	**35**	2	23	2	24	2	23	3	23	3	26	3	25	1	22	1	19	1	22
25	**37.5**	3	24	3	26	3	25	3	25	3	26	3	26	2	23	2	24	2	24
26	**40**	3	26	4	27	3	26	4	26	5	29	4	28	2	23	2	24	2	24
27	**42.5**	4	27	4	29	4	28	4	28	6	30	5	29	3	27	3	27	3	27
28	**45**	4	29	5	30	5	29	5	29	7	32	6	30	4	28	3	28	3	28
29	**47.5**	6	30	7	32	6	31	7	31	9	33	8	32	4	30	5	30	5	30
30	**50**	6	32	9	33	7	32	8	32	10	35	9	33	4	30	8	31	6	31
31	**52.5**	8	33	9	35	9	34	9	34	10	36	10	35	7	33	9	33	8	33
32	**55**	10	35	11	36	10	35	10	35	11	37	11	36	9	34	10	34	10	34
33	**57.5**	13	36	13	37	13	37	13	37	15	39	14	38	12	36	12	36	12	36
34	**60**	14	38	17	39	15	38	16	38	18	40	17	39	13	37	15	37	14	37
35	**62.5**	16	39	20	40	18	40	17	40	23	42	19	41	16	39	17	39	16	39
36	**65**	19	41	22	42	20	41	19	41	25	43	22	42	19	40	19	40	19	40
37	**67.5**	22	42	26	43	24	43	22	43	31	45	26	44	23	42	21	42	22	42
38	**70**	26	44	29	45	28	44	26	44	34	46	30	45	27	43	24	43	26	43
39	**72.5**	29	45	33	46	31	46	29	46	39	48	34	47	28	45	26	45	28	45
40	**75**	33	47	37	48	35	47	34	47	43	49	39	48	31	46	31	46	31	46
41	**77.5**	36	48	42	49	39	49	37	49	48	51	42	50	35	48	36	48	35	48
42	**80**	41	50	49	51	44	50	43	50	54	52	48	51	39	49	43	49	41	49
43	**82.5**	45	51	52	52	48	52	47	52	57	54	52	53	44	51	47	51	45	51
44	**85**	54	53	58	54	56	53	55	53	65	55	60	54	53	52	51	52	52	52
45	**87.5**	60	54	65	55	62	55	61	55	74	57	67	56	59	54	57	54	58	54
46	**90**	68	56	73	57	70	56	71	56	82	58	77	57	65	55	64	55	64	55
47	**92.5**	74	57	80	58	77	58	77	58	88	60	82	58	71	57	73	57	72	57
48	**95**	83	59	85	60	84	59	86	59	92	61	89	60	80	58	79	58	79	58
49	**97.5**	91	60	92	61	92	61	95	60	94	63	95	61	89	60	90	60	89	60
50	**100**	100	62	100	63	100	62	100	62	100	64	100	63	100	61	100	61	100	61

Notes: *Calculation of scores was restricted to complete cases. RS = raw score; TRS = transformed raw score (range 0-100); PR = percentile: TS = T-score (M = 50, SD = 10).*

A.2c: DISABKIDS chronic generic short-form score if "medication/treatment" is appicable (proxy-report) – DCGM-12-P-SF (DISABKIDS field study sample, N = 1.153)

DCGM-12-P SCORE PROXY-REPORT		Total sample 8-12 (n = 446)		Total sample 13-16 (n = 326)		Total sample Overall (n = 772)		Females 8-12 (n = 202)		Females 13-16 (n =160)		Females Overall (n = 362)		Males 8-12 (n = 242)		Males 13-16 (n = 164)		Males Overall (n = 406)	
RS	**TRS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**
<= 25	**27**	0	21	2	24	1	22	-	-	2	23	1	22	1	21	1	10	1	17
26	**29**	0	21	2	25	1	23	-	-	3	25	1	23	1	21	1	10	1	17
27	**31**	1	23	3	26	2	25	1	23	4	26	2	25	1	23	1	10	1	21
28	**33**	1	25	3	27	2	26	2	25	4	26	3	26	1	23	1	27	1	21
29	**35**	3	26	3	29	3	27	2	26	4	29	3	27	3	26	2	28	3	23
30	**38**	3	27	4	30	4	28	3	27	4	29	4	28	4	27	2	29	3	24
31	**40**	4	28	5	31	4	30	3	27	4	29	4	28	5	28	4	30	4	26
32	**42**	5	30	6	32	5	31	3	27	5	32	4	31	6	30	6	32	6	27
33	**44**	5	31	7	33	6	32	4	31	6	34	5	32	6	31	7	33	7	28
34	**46**	7	32	8	35	7	33	6	32	7	35	7	33	7	32	8	34	7	29
35	**48**	8	33	10	36	8	35	8	34	9	36	8	35	7	33	10	35	8	31
36	**50**	9	35	11	37	10	36	9	35	10	37	10	36	8	35	12	36	10	32
37	**52**	10	36	13	38	11	37	11	36	13	39	12	37	10	36	13	37	11	33
38	**54**	12	37	15	39	13	38	13	37	15	40	14	39	11	37	15	39	13	35
39	**56**	15	39	19	41	17	39	17	39	19	41	18	40	13	38	18	40	15	36
40	**58**	18	40	21	42	19	41	20	40	22	42	21	41	15	40	20	41	17	37
41	**60**	20	41	23	43	21	42	22	41	25	44	24	42	18	41	21	42	20	38
42	**63**	24	42	26	44	25	43	26	43	30	45	28	44	23	42	23	43	23	40
43	**65**	26	44	32	45	28	44	27	44	33	46	30	45	25	43	31	45	27	41
44	**67**	30	45	33	47	31	46	31	45	35	47	33	46	29	45	31	46	30	42
45	**69**	34	46	40	48	37	47	37	46	44	49	40	47	32	46	35	47	34	44
46	**71**	39	47	45	49	41	48	41	48	46	50	43	49	37	47	43	48	39	45
47	**73**	43	49	49	50	46	49	43	49	51	51	46	50	43	48	47	49	45	46
48	**75**	48	50	55	51	51	51	50	50	59	52	54	51	46	50	51	51	48	47
49	**77**	53	51	59	53	55	52	55	52	63	54	59	53	51	51	55	52	53	49
50	**79**	58	52	64	54	61	53	58	53	68	55	63	54	57	52	60	53	58	50
51	**81**	62	54	71	55	66	54	65	54	76	56	70	55	60	53	65	54	62	51
52	**83**	69	55	74	56	71	56	70	56	79	57	74	56	68	55	68	55	68	52
53	**85**	73	56	78	57	75	57	74	57	84	59	78	58	72	56	72	57	72	54
54	**88**	77	58	82	59	79	58	77	58	86	60	81	59	78	57	77	58	78	55
55	**90**	79	59	86	60	82	59	80	59	89	61	84	60	79	58	82	59	80	56
56	**92**	78	58	84	60	81	59	85	59	91	61	88	60	73	57	77	58	75	58
57	**94**	82	59	89	61	85	60	89	60	93	62	91	61	77	59	85	59	80	59
58	**96**	90	61	92	62	91	61	94	62	94	64	94	63	87	60	90	61	88	60
59	**98**	94	62	96	63	95	62	97	63	97	65	97	64	91	61	95	62	93	61
60	**100**	100	63	100	65	100	64	100	64	100	66	100	65	100	62	100	63	100	63

Notes: *Calculation of scores was restricted to complete cases. RS = raw score; TRS = transformed raw score (range 0-100); PR = percentile: TS = T-score (M = 50, SD = 10).*

A.2d: DISABKIDS chronic generic short-form score if "medication/treatment" is not appicable (proxy-report) – DCGM-12-P-SF (V-10) (DISABKIDS field study sample, N = 1.153)

DCGM-12-P (V-10) SCORE (V-10) PROXY-REPORT		Total sample 8-12 (n = 494)		Total sample 13-16 (n = 366)		Total sample Overall (n = 860)		Females 8-12 (n = 228)		Females 13-16 (n =173)		Females Overall (n = 401)		Males 8-12 (n = 264)		Males 13-16 (n = 191)		Males Overall (n = 455)	
RS	**TRS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**	**PR**	**TS**
<= 20	**<= 25**	1	20	1	20	1	21	1	20	1	20	1	20	0	19	1	10	0	19
21	**27.5**	1	21	1	20	1	22	1	21	1	20	1	22	1	21	1	10	1	22
22	**30**	1	21	1	20	1	22	1	21	1	20	1	22	1	21	1	10	1	22
23	**32.5**	2	24	1	26	2	25	1	21	2	26	2	25	3	24	1	10	2	25
24	**35**	2	26	3	28	3	26	1	21	3	27	2	26	3	26	3	27	3	26
25	**37.5**	3	27	4	29	4	28	2	27	5	29	3	28	4	27	3	28	4	28
26	**40**	4	28	5	30	4	29	3	29	5	30	4	29	5	29	4	30	4	29
27	**42.5**	5	30	6	32	6	31	4	30	6	32	5	31	6	30	6	31	6	31
28	**45**	6	31	8	33	7	32	5	31	8	33	6	32	7	31	8	33	7	32
29	**47.5**	7	33	9	35	8	34	7	33	9	35	8	34	7	33	9	34	8	33
30	**50**	8	34	12	36	10	35	8	34	12	36	10	35	9	34	12	35	10	35
31	**52.5**	10	36	13	37	11	37	11	36	12	38	12	37	10	36	13	37	11	36
32	**55**	12	37	15	39	13	38	12	37	13	39	13	38	13	37	16	38	14	38
33	**57.5**	14	39	18	40	16	39	15	39	17	41	16	40	14	39	18	40	15	39
34	**60**	19	40	20	42	19	41	21	40	22	42	21	41	17	40	18	41	18	40
35	**62.5**	21	41	24	43	22	42	23	42	26	44	24	43	20	41	21	42	20	42
36	**65**	25	43	28	45	26	44	26	43	30	45	28	44	24	43	25	44	24	43
37	**67.5**	28	44	32	46	30	45	29	45	33	47	31	46	27	44	30	45	28	45
38	**70**	32	46	39	47	35	47	35	46	42	48	38	47	30	46	36	47	33	46
39	**72.5**	38	47	43	49	40	48	39	48	46	50	42	48	37	47	41	48	39	47
40	**75**	42	49	51	50	46	49	44	49	54	51	49	50	40	49	47	49	43	49
41	**77.5**	47	50	56	52	51	51	50	50	62	53	55	51	45	50	51	51	47	50
42	**80**	53	52	61	53	57	52	56	52	66	54	60	53	52	51	57	52	54	52
43	**82.5**	60	53	67	55	63	54	61	53	73	56	66	54	58	53	61	54	60	53
44	**85**	65	55	69	56	67	55	64	55	77	57	70	56	66	54	62	55	65	55
45	**87.5**	71	56	77	57	74	57	70	56	83	59	75	57	73	56	71	56	72	56
46	**90**	75	57	81	59	78	58	75	58	86	60	80	59	75	57	75	58	75	57
47	**92.5**	80	59	86	60	83	59	83	59	92	62	87	60	78	59	81	59	79	59
48	**95**	85	60	91	62	87	61	87	61	95	63	91	62	83	60	86	61	84	60
49	**97.5**	92	62	96	63	94	62	94	62	98	65	96	63	91	61	95	62	92	62
50	**100**	100	63	100	65	100	64	100	64	100	66	100	65	100	63	100	63	100	63

Notes: *Calculation of scores was restricted to complete cases. RS = raw score; TRS = transformed raw score (range 0-100); PR = percentile: TS = T-score (M = 50, SD = 10).*

A.3a: A-priori estimations for selected descriptive item characteristics of the DISABKIDS chronic generic item pool (self-report; sub-sample II from the DISABKIDS field study sample, n = 578)

		n	**Missing**	**1**	**2**	**3**	**4**	**5**	**M**	**SD**	**Variance**	**Skewness**	**Curtosis**
				never	*seldom*	*quite often*	*very often*	*always*					
No.	*Facet*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*					
01.	**IND**	552	26 (4.50)	**2.42**	6.92	22.15	28.20	35.81	3.92	1.059	1.121	-.728	-.186
02.	**IND**	564	14 (2.42)	**1.56**	**3.29**	11.25	28.89	**52.60**	4.31	.916	.839	-1.414	1.778
03.	**IND**	561	17 (2.94)	**3.81**	9.86	20.76	31.66	30.97	3.78	1.117	1.248	-.685	-.308
04.	**IND**	563	15 (2.60)	6.06	6.75	15.40	25.61	43.60	3.96	1.201	1.444	-1.037	.135
05.	**IND**	563	15 (2.60)	**4.15**	6.57	16.44	26.82	43.43	4.01	1.127	1.270	-1.031	.265
06.	**IND**	564	14 (2.42)	**3.46**	**4.84**	13.32	32.18	43.77	4.11	1.046	1.093	-1.224	1.023
07.	**LIM**	565	13 (2.25)	5.36	5.02	11.42	24.39	**51.56**	4.14	1.151	1.325	-1.353	.976
08.	**LIM**	564	14 (2.42)	5.36	9.34	24.57	22.32	35.99	3.76	1.201	1.443	-.629	-.554
09.	**LIM**	562	16 (2.77)	5.54	9.17	18.69	22.15	41.70	3.88	1.224	1.498	-.824	-.369
10.	**LIM**	561	17 (2.94)	9.17	9.69	20.42	18.17	39.62	3.71	1.338	1.790	-.672	-.752
11.	**LIM**	565	13 (2.25)	**3.29**	**4.33**	14.19	19.38	**56.57**	4.24	1.070	1.146	-1.378	1.136
12.	**LIM**	563	15 (2.60)	**2.77**	6.57	24.57	23.18	40.31	3.94	1.091	1.191	-.725	-.299
13.	**EMO**	563	15 (2.60)	**1.73**	**3.11**	16.61	23.88	**52.08**	4.25	.966	.934	-1.200	.903
14.	**EMO**	564	14 (2.42)	**4.15**	6.92	17.47	23.88	45.16	4.01	1.145	1.311	-.997	.112
15.	**EMO**	560	18 (3.11)	**3.11**	6.92	22.66	29.24	34.95	3.89	1.077	1.159	-.742	-.123
16.	**EMO**	561	17 (2.94)	**2.94**	7.44	23.70	21.63	41.35	3.94	1.117	1.248	-.733	-.371
17.	**EMO**	558	20 (3.46)	**2.77**	**3.63**	15.22	20.42	**54.50**	4.25	1.036	1.072	-1.332	1.123
18.	**EMO**	560	18 (3.11)	**2.25**	**3.81**	14.71	26.12	50.00	4.22	.994	.989	-1.254	1.082
19.	**EMO**	557	21 (3.63)	7.96	12.11	20.07	17.82	38.41	3.69	1.327	1.761	-.597	-.867

(A. 3a continued)

		n	Missing	1	2	3	4	5	M	SD	Variance	Skewness	Curtosis
				never	*seldom*	*quite often*	*very often*	*always*					
No.	*Facet*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*					
20.	**EXCL**	559	19 (3.29)	**1.73**	**1.73**	8.48	12.11	**72.66**	**4.57**	.862	.743	**-2.244**	**4.835**
21.	**EXCL**	563	15 (2.60)	**2.25**	**4.15**	10.90	17.13	**62.98**	4.38	.997	.994	-1.640	1.992
22.	**EXCL**	562	16 (2.77)	**3.81**	**4.84**	15.92	19.90	**52.77**	4.16	1.111	1.234	-1.232	.686
23.	**EXCL**	563	15 (2.60)	**1.04**	**3.29**	7.09	19.90	**66.09**	**4.51**	.853	.727	-1.920	3.436
24.	**EXCL**	561	17 (2.94)	**1.90**	**2.94**	9.52	16.26	**66.44**	4.47	.929	.864	-1.873	3.056
25.	**EXCL**	562	16 (2.77)	**3.98**	**3.98**	16.96	22.32	50.00	4.14	1.097	1.204	-1.204	.740
26.	**INCL**	552	26 (4.50)	5.88	10.73	18.34	29.24	31.31	3.73	1.205	1.451	-.699	-.459
27.	**INCL**	559	19 (3.29)	10.90	7.61	19.03	20.76	38.41	3.70	1.356	1.839	-.734	-.662
28.	**INCL**	563	15 (2.60)	**2.25**	**3.81**	10.38	24.57	**56.40**	4.33	.973	.946	-1.551	1.983
29.	**INCL**	563	15 (2.60)	**3.11**	5.71	15.05	26.82	46.71	4.11	1.070	1.146	-1.141	.608
30.	**INCL**	560	18 (3.11)	**0.52**	**2.08**	7.09	24.39	**62.80**	**4.52**	.768	.590	-1.753	3.143
31.	**INCL**	563	15 (2.60)	12.28	13.84	23.01	21.11	27.16	**3.38**	1.356	1.837	-.343	-1.061
32.	**MED**	492	**86 (14.88)**	5.02	**3.81**	15.05	14.71	46.54	4.10	1.194	1.425	-1.196	-.458
33.	**MED**	480	**98 (16.96)**	**3.63**	**2.77**	10.73	13.84	**52.08**	4.30	1.094	1.196	-1.578	1.694
34.	**MED**	491	**87 (15.05)**	11.76	12.80	18.17	15.57	26.64	**3.38**	1.414	2.000	-.325	-1.195
35.	**MED**	494	**84 (14.53)**	**1.90**	**3.98**	11.76	16.78	**51.04**	4.30	1.017	1.034	-1.405	1.233
36.	**MED**	493	**85 (14.71)**	12.28	8.13	18.51	12.11	34.26	3.56	1.452	2.108	-.531	-1.075
37.	**MED**	492	**86 (14.88)**	15.57	5.19	15.92	15.40	33.04	3.53	1.500	2.250	-.584	-1.080

Notes: Facet abbreviations are as follows: IND = “Independence”; EMO = “Emotion”; EXCL = “Exclusion”; INCL = “Inclusion”; LIM = “Limitation”; MED = “Medication”. In **bold** print: missing data rate > 5%; response rates > 50% / < 5%. ‡ Reversed scored items (recoded before analysis). * Missing data rates are systematically overestimated, due to includion of non-applicable rates of respondents without any medical treatment.

A.3b: A-priori estimations for selected descriptive item characteristics of the DISABKIDS chronic generic item pool (proxy-report; sub-sample II from the DISABKIDS field study sample, n = 578)

		n	**Missing**	**1**	**2**	**3**	**4**	**5**	**M**	**SD**	**Variance**	**Skewness**	**Curtosis**
				never	*seldom*	*quite often*	*very often*	*always*					
No.	*Facet*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*					
01.	**IND**	498	**34 (6.39)**	**1.32**	**3.76**	17.29	36.65	34.59	4.06	.916	.839	-.880	.558
02.	**IND**	504	**28 (5.26)**	**0.56**	**1.50**	10.34	41.54	40.79	4.27	.764	.584	-1.041	1.483
03.	**IND**	500	**32 (6.02)**	**2.26**	7.71	13.16	41.54	29.32	3.94	.997	.994	-.956	.516
04.	**IND**	499	**33 (6.20)**	**2.07**	5.83	14.29	35.34	36.28	4.04	.993	.986	-1.015	.613
05.	**IND**	493	**39 (7.33)**	**2.44**	6.95	14.10	35.34	33.83	3.98	1.028	1.057	-.971	.420
06.	**IND**	504	**28 (5.26)**	**1.69**	**4.89**	12.41	36.84	38.91	4.12	.946	.895	-1.124	1.027
07.	**LIM**	504	**28 (5.26)**	**3.57**	5.45	9.96	31.95	43.80	4.13	1.059	1.122	-1.317	1.193
08.	**LIM**	501	**31 (5.83)**	**4.14**	11.28	34.21	25.19	19.36	**3.47**	1.080	1.166	-.246	-.510
09.	**LIM**	502	**30 (5.64)**	5.45	10.34	22.74	29.51	26.32	3.65	1.164	1.355	-.584	-.468
10.	**LIM**	501	**31 (5.83)**	**4.89**	8.83	20.68	27.82	31.95	3.78	1.164	1.354	-.712	-.330
11.	**LIM**	503	**29 (5.45)**	**2.44**	5.83	16.92	27.44	41.92	4.06	1.049	1.100	-.991	.307
12.	**LIM**	500	**32 (6.02)**	**2.26**	6.58	24.44	28.01	32.71	3.88	1.044	1.091	-.630	-.276
13.	**EMO**	497	**35 (6.58)**	**0.75**	**4.89**	22.18	28.95	36.65	4.03	.955	.913	-.637	-.372
14.	**EMO**	498	**34 (6.39)**	**3.20**	6.77	23.68	29.70	30.26	3.82	1.069	1.144	-.674	-.146
15.	**EMO**	497	**35 (6.58)**	**2.44**	6.39	27.63	34.96	21.99	3.72	.983	.966	-.525	-.028
16.	**EMO**	499	**33 (6.20)**	**2.82**	9.40	30.08	21.62	29.89	3.71	1.108	1.227	-.381	-.704
17.	**EMO**	495	**37 (6.95)**	**1.50**	**4.14**	17.67	29.32	40.41	4.11	.968	.938	-.941	.376
18.	**EMO**	497	**35 (6.58)**	**1.32**	7.52	23.31	27.82	33.46	3.91	1.024	1.050	-.578	-.489
19.	**EMO**	495	**37 (6.95)**	**3.57**	10.53	27.07	24.44	27.44	3.66	1.128	1.273	-.430	-.640

(A.3b continued)

		n	**Missing**	**1**	**2**	**3**	**4**	**5**	**M**	**SD**	**Variance**	**Skewness**	**Curtosis**
				never	*seldom*	*quite often*	*very often*	*always*					
No.	*Facet*	*n*	*n (%)*	*%*	*%*	*%*	*%*	*%*					
20.	**EXCL**	498	**34 (6.39)**	**0.94**	**4.89**	11.84	19.74	**56.20**	4.34	.955	.913	-1.363	1.043
21.	**EXCL**	499	**33 (6.20)**	**1.50**	**2.44**	13.72	20.11	**56.02**	4.35	.933	.871	-1.421	1.547
22.	**EXCL**	501	**31 (5.83)**	**3.57**	7.14	20.30	23.68	39.47	3.94	1.131	1.278	-.829	-.165
23.	**EXCL**	502	**30 (5.64)**	**1.13**	**3.38**	11.47	26.32	**52.07**	4.32	.907	.822	-1.360	1.474
24.	**EXCL**	500	**32 (6.02)**	**0.56**	**3.38**	9.40	22.56	**58.08**	4.43	.857	.734	-1.521	1.815
25.	**EXCL**	501	**31 (5.83)**	**1.69**	**4.32**	21.80	30.26	36.09	4.01	.981	.962	-.765	.074
26.	**INCL**	492	**40 (7.52)**	**3.38**	11.28	21.43	37.03	19.36	3.62	1.058	1.119	-.572	-.283
27.	**INCL**	498	**34 (6.39)**	5.26	8.46	16.17	27.44	36.28	3.87	1.188	1.412	-.872	-.159
28.	**INCL**	499	**33 (6.20)**	**1.32**	5.26	12.41	25.56	49.25	4.24	.976	.953	-1.221	.823
29.	**INCL**	501	**31 (5.83)**	**0.94**	5.83	11.09	33.65	42.67	4.18	.936	.877	-1.117	.750
30.	**INCL**	499	**33 (6.20)**	**0.19**	**1.69**	6.95	29.51	**55.45**	4.47	.729	.531	-1.413	1.948
31.	**INCL**	497	**35 (6.58)**	9.21	13.72	21.43	23.31	25.75	**3.46**	1.299	1.688	-.406	-.937
32.	**MED**	438	**94 (17.67)**	**3.01**	**4.32**	8.27	16.73	50.00	4.29	1.079	1.164	-1.558	1.612
33.	**MED**	444	**88 (16.54)**	8.83	13.35	25.38	17.48	18.42	**3.28**	1.265	1.601	-.195	-.929
34.	**MED**	446	**86 (16.17)**	**1.32**	**4.14**	13.72	25.00	39.66	4.16	.975	.951	-1.063	.560
35.	**MED**	448	**84 (15.79)**	8.83	8.46	24.62	19.17	23.12	**3.47**	1.277	1.632	-.433	-.771
36.	**MED**	448	**84 (15.79)**	8.65	6.77	20.30	21.99	26.50	3.60	1.284	1.649	-.632	-.595
37.	**MED**	447	**85 (15.98)**	**4.14**	**4.89**	15.41	25.38	34.21	3.96	1.126	1.267	-1.000	.310

Notes: Facet abbreviations are as follows: IND = “Independence”; EMO = “Emotion”; EXCL = “Exclusion”; INCL = “Inclusion”; LIM = “Limitation”; MED = “Medication”. In **bold** print: missing data rate > 5%; response rates > 50% / < 5%. ‡ Reversed scored items (recoded before analysis). * Missing data rates are systematically overestimated, due to inclusion of non-applicable rates of respondents without any me dical treatment.

A.4a: **Country specific results for reliability analysis (internal consistency, split-half reliability, corrected item-total correlations) of the DISABKIDS chronic generic shortform (DCGM-12; self-report version; sub-sample II from the DISABKIDS field study sample, n = 578)**

Country (Language)	**Germany**		**Nether-lands**		**United Kingdom**		**France**		**Greece**		**Sweden**		**Austria**		(German language) *	
Version	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**
n	101	[130]	129	[134]	49	[54]	32	[34]	25	[34]	90	[96]	45	[52]	146	[182]
α	.832	[.855]	.815	[.832]	.862	[.891]	.802	[.805]	.841	[.860]	.912	[.906]	.857	[.820]	.843	.849
split-half	.903	[.881]	.883	[.842]	.921	[.937]	.914	[.894]	.838	[.843]	.953	[.938]	.882	[.820]	.936	.929
01.	.600	[.629]	.548	[.585]	.556	[.609]	.683	[.745]	.526	[.590]	.691	[.736]	.501	[.554]	.570	.609
02.	.482	[.555]	.559	[.645]	.812	[.858]	.671	[.746]	**.385**	[.491]	.751	[.768]	.780	[.785]	.561	.607
03.	.493	[.569]	**.390**	[.453]	.636	[.620]	.546	[.440]	.510	[.547]	.542	[.554]	.436	[.432]	.486	.540
04.	.596	[.637]	**.318**	**[.380]**	.586	[.575]	**.074**	**[.120]**	.496	[.588]	.675	[.647]	.404	[.478]	.508	.460
05.	.550	[.513]	.587	[.530]	.670	[.648]	.614	[.615]	.613	[.574]	.774	[.771]	**.380**	**[.281]**	.563	.538
06.	.547	[.574]	.619	[.629]	.707	[.661]	.492	[.486]	.450	**[.382]**	.742	[.740]	.727	[.654]	.597	.596
07.	.526	[.523]	.587	[.626]	.752	[.702]	**.252**	**[.271]**	.620	[.636]	.581	[.604]	.671	[.584]	.542	.600
08.	.532	[.601]	.565	[.561]	.780	[.785]	.584	[.549]	.714	[.753]	.771	[.781]	.740	[.656]	.592	.618
09.	.471	[.556]	.455	[.518]	.489	[.554]	.541	[.570]	.618	[.734]	.634	[.641]	**.238**	**[.268]**	.427	.499
10.	.440	[.467]	**.339**	**[.350]**	**.126**	**[.259]**	**.282**	**[.311]**	.466	[.460]	.440	[.484]	.507	[.546]	.469	.491
11.	**.387**	**-**	.440	**-**	**.194**	**-**	.459	**-**	.422	**-**	.658	**-**	.574	**-**	.445	**-**
12.	.407	**-**	**.392**	**-**	**.264**	**-**	.683	**-**	**.336**	**-**	.645	**-**	.635	**-**	.479	**-**

Notes: * Computations for “german language” based on merged austrian and german sub-samples. Reliability coefficients (internal consistency, split-half reliability) < .80 and coefficients for corrected item-total correlations < .40 are marked **in bold**.

A.4b: **Country specific results for reliability analysis (internal consistency, split-half reliability, corrected item-total correlations) of the DISABKIDS chronic generic shortform (DCGM-12; proxy-report version; sub-sample II from the DISABKIDS field study sample, n = 578)**

Country (Language)	**Germany**		**Nether-lands**		**United Kingdom**		**France**		**Greece**		**Sweden**		**Austria**		(German language) *	
Version	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**
n	76	[102]	128	[131]	43	[49]	29	31	21	24	88	91	42	49	118	[151]
α	.885	[.908]	.887	[.892]	.841	[.904]	.859	[.873]	.887	[.878]	.921	[.921]	.924	[.911]	.895	.907
split-half	.936	[.928]	.929	[.913]	.902	[.928]	.933	[.934]	.965	[.931]	.935	[.919]	.941	[.935]	.900	.870
01.	.670	[.783]	.634	[.706]	.683	[.678]	.726	[.700]	.599	[.596]	.757	[.788]	.759	[.707]	,570	,609
02.	.683	[.670]	.662	[.710]	.794	[.854]	.604	[.621]	.665	[.673]	.583	[.633]	.794	[.749]	,561	,607
03.	.477	[.589]	.681	[.664]	.515	[.613]	.493	[.531]	.468	[.488]	.696	[.677]	.682	[.628]	,486	,540
04.	.635	[.676]	.522	[.523]	.440	[.513]	**.365**	**[.393]**	.779	[.724]	.689	[.691]	.503	[.523]	,542	,600
05.	.703	[.687]	.724	[.714]	.717	[.675]	.601	[.664]	.746	[.667]	.810	[.799]	.731	[.629]	,508	,460
06.	.655	[.669]	.671	[.727]	.764	[.745]	.616	[.683]	.626	[.606]	.772	[.776]	.731	[.650]	,597	,596
07.	.629	[.702]	.639	[.710]	.785	[.823]	.745	[.695]	**.393**	[.503]	.730	[.737]	.637	[.716]	,563	,538
08.	.618	[.725]	.689	[.689]	.671	[.736]	.716	[.681]	.765	[.664]	.797	[.801]	.751	[.764]	,592	,618
09.	.564	[.670]	.538	[.550]	.501	[.564]	.552	[.640]	.776	[.758]	.592	[.646]	.666	[.739]	,427	,499
10.	.513	[.525]	**.345**	**[.325]**	.410	[.453]	.465	[.482]	.489	[.461]	.554	[.571]	.670	[.687]	,469	,491
11.	.502	**-**	.585	**-**	**-.020**	**-**	**.364**	**-**	.412	**-**	.542	**-**	.533	**-**	,469	**-**
12.	.525	**-**	.471	**-**	**-.014**	**-**	**.356**	**-**	.703	**-**	.681	**-**	.452	**-**	,479	**-**

Notes: * Computations for “german language” based on merged austrian and german sub-samples. Reliability coefficients (internal consistency, split-half reliability) < .80 and coefficients for corrected item-total correlations < .40 are marked **in bold**.

A.5a: **Condition specific results for reliability analysis (internal consistency, split-half reliability, corrected item-total correlations) of the DISABKIDS chronic generic shortform (DCGM-12; self-report version; sub-sample II from the DISABKIDS field study sample, n = 578)**

Chronic Condition	***Asthma***		***Arthritis***		***Diabetes***		***Skin Disease***		***Cerebral palsy***		***Cystic Fibrosis***		***Epilepsy***	
Version	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**
n	178	[193]	56	[72]	25	[26]	99	[100]	13	[39]	18	[19]	82	[85]
α	.805	[.830]	.836	[.886]	**.791**	[**.786**]	.850	[.826]	.900	[.864]	.871	[.851]	.888	[.895]
split-half	.853	[.842]	.903	[.911]	**.789**	[**.737**]	.915	[.859]	.917	[.872]	.906	[.836]	.956	[.948]
01.	.534	[.612]	.608	[.646]	.594	[.658]	.504	[.604]	.755	[.551]	.468	[.529]	.595	[.631]
02.	.584	[.687]	.449	[.562]	.464	[.532]	.643	[.658]	.732	[.726]	.533	[.506]	.711	[.751]
03.	.414	[.456]	**.343**	[.508]	.705	[.534]	**.375**	[**.375**]	.668	[.603]	.519	[.621]	.635	[.669]
04.	.462	[.516]	.516	[.583]	.481	[.575]	.586	[.569]	.443	[.506]	.729	[.742]	.646	[.622]
05.	.535	[.469]	.592	[.656]	.700	[.588]	.621	[.558]	.491	[.415]	.403	[**.340**]	.634	[.650]
06.	.482	[.458]	.624	[.720]	.459	[.602]	.629	[.593]	.872	[.649]	**.374**	[**.368**]	.746	[.762]
07.	.481	[.529]	.722	[.699]	**.095**	[**.125**]	.484	[.455]	.846	[.661]	.829	[.804]	.632	[.647]
08.	.604	[.623]	.621	[.684]	.463	[**.346**]	.561	[.540]	.594	[.650]	.790	[.757]	.723	[.732]
09.	.449	[.516]	.659	[.688]	**.235**	[**.338**]	.459	[.492]	.631	[.653]	.461	[.462]	.551	[.593]
10.	**.310**	[**.340**]	.414	[.515]	**.061**	[**.125**]	**.305**	[**.311**]	**.372**	[.403]	.543	[.536]	**.309**	[**.335**]
11.	.429	-	**.339**	-	.488	-	.647	-	.514	-	.545	-	.403	-
12.	**.292**	-	**.342**	-	**.333**	-	.529	-	.646	-	.669	-	.601	-

Notes: * Reliability coefficients (internal consistency, split-half reliability) < .80 and coefficients for corrected item-total correlations < .40 are marked **in bold**.

A.5b: **Condition specific results for reliability analysis (internal consistency, split-half reliability, corrected item-total correlations) of the DISABKIDS chronic generic shortform (DCGM-12; proxy-report version; sub-sample II from the DISABKIDS field study sample, n = 578)**

Chronic Condition	***Asthma***		***Arthritis***		***Diabetes***		***Skin Disease***		***Cerebral palsy***		***Cystic Fibrosis***		***Epilepsy***	
Version	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**	**V-12**	**[V-10]**
n	174	[183]	56	[70]	19	[20]	96	[97]	15	[37]	11	[11]	56	[59]
α	.874	[.886]	.883	[.894]	**.786**	[.859]	.882	[.868]	.901	[.902]	.820	[.866]	.937	[.948]
split-half	.930	[.922]	.918	[.907]	.941	[.937]	.892	[.864]	.958	[.928]	.893	[.899]	.963	[.958]
01.	.621	[.652]	.684	[.779]	**.346**	[.487]	.618	[.651]	.700	[.651]	.640	[.582]	.820	[.864]
02.	.612	[.661]	.725	[.714]	**.298**	[.476]	.571	[.572]	.606	[.583]	.449	[.494]	.790	[.828]
03.	.607	[.645]	.632	[.678]	**.261**	[.438]	.493	[.477]	.714	[.697]	.501	[.572]	.691	[.679]
04.	.696	[.676]	.493	[.526]	.787	[.828]	.532	[.544]	**.276**	[.467]	.585	[.606]	.762	[.776]
05.	.673	[.645]	.650	[.640]	.778	[.795]	.723	[.689]	.892	[.723]	.871	[.793]	.887	[.877]
06.	.671	[.719]	.583	[.631]	.721	[.728]	.673	[.645]	.701	[.750]	.520	[.530]	.811	[.817]
07.	.528	[.588]	.627	[.622]	.636	[.552]	.643	[.677]	.816	[.859]	**.398**	[.604]	.841	[.869]
08.	.636	[.607]	.694	[.762]	.690	[.694]	.670	[.685]	.647	[.751]	.748	[.823]	.804	[.798]
09.	.505	[.561]	.686	[.677]	**.397**	[.510]	.552	[.585]	.493	[.590]	.696	[.781]	.679	[.723]
10.	.434	[.420]	**.356**	[**.372**]	**.104**	[**.134**]	.403	[.406]	.690	[.508]	**-.142**	[**-.053**]	.548	[.585]
11.	.453	**-**	.493	**-**	**.123**	**-**	.554	**-**	.420	**-**	**.223**	**-**	.540	**-**
12.	**.394**	**-**	.516	**-**	**.157**	**-**	.629	**-**	.534	**-**	.436	**-**	.558	**-**

Notes: * Reliability coefficients (internal consistency, split-half reliability) < .80 and coefficients for corrected item-total correlations < .40 are marked **in bold.**

Verifications (Nachweise)

- *Erklärung an Eides statt I* (deutsch)
- *Erklärung an Eides statt II* (deutsch)

Erklärung an Eides statt (I)

Nach § 3 Abs. 2 Nr. 9 der Übergangsordnung für die Promotion zum Doktor der Philosophie der Universität Hamburg vom 17. September 1969:

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst sowie andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die wörtlich und inhaltlich übernommenen Stellen als solche kenntlich gemacht habe.

Hamburg, den 01.06.2010

Holger Muehlan.

Erklärung an Eides statt (II)

nach § 3 Abs. 2 Nr. 7 der Übergangsordnung für die Promotion zum Doktor der Philosophie der Universität Hamburg vom 17. September 1969:

Hiermit erkläre ich an Eides statt, dass ich mich nicht schon anderwärts der Doktorprüfung unterzogen oder um Zulassung zu ihr beworben habe.

Hamburg, den 01.06.2010.