



Comparing Generic Paediatric Health-Related Quality-of-Life Instruments: A Dimensionality Assessment Using Factor Analysis

Mina Bahrapour¹ · Renee Jones² · Kim Dalziel² · Nancy Devlin² · Brendan Mulhern¹ on behalf of the QUOKKA (Quality of Life in Kids: Key Evidence for Decision Makers in Australia) Team

Accepted: 11 April 2024 / Published online: 31 May 2024
© The Author(s) 2024

Abstract

Background Widely used generic instruments to measure paediatric health-related quality of life (HRQoL) include the EQ-5D-Y-5L, Child Health Utility 9 Dimension (CHU-9D), Paediatric Quality of Life Inventory (PedsQL) and Health Utilities Index (HUI). There are similarities and differences in the content of these instruments, but there is little empirical evidence on how the items they contain relate to each other, and to an overarching model of HRQoL derived from their content.

Objective This study aimed to explore the dimensionality of the instruments using exploratory factor analysis (EFA).

Methods Data from the Australian Paediatric Multi-Instrument Comparison (P-MIC) Study were used. EQ-5D-Y-5L, CHU-9D, PedsQL and HUI data were collected via proxy or child self-report data. EFA was used to investigate the underlying domain structure and measurement relationship. Items from the four instruments were pooled and domain models were identified for self- and proxy-reported data. The number of factors was determined based on eigenvalues greater than 1. A correlation cut-off of 0.32 was used to determine item loading on a given factor, with cross-loading also considered. Oblique rotation was used.

Results Results suggest a six-factor structure for the proxy-reported data, including emotional functioning, pain, daily activities, physical functioning, school functioning, and senses, while the self-report data revealed a similar seven-factor structure, with social functioning emerging as an additional factor.

Conclusion We provide evidence of differences and similarities between paediatric HRQoL instruments and the aspects of health being measured by these instruments. The results identified slight differences between self- and proxy-reported data in the relationships among items within the resulting domains.

Key Points for Decision Makers

This study identifies the domains that emerge from factor analysis of the items from the EQ-5D-Y-5L, Child Health Utility 9 Dimension (CHU-9D), Paediatric Quality of Life Inventory (PedsQL) and Health Utilities Index (HUI).

The study shows how items from different instruments relate to each other and to an overarching model of HRQoL derived by combined instrument content.

The study demonstrates that when determining HRQoL domains for children, proxy-reported data yield a slightly different model of HRQoL in comparison with self-reported data.

✉ Mina Bahrapour
mina.bahrapour@uts.edu.au

Renee Jones
reneej1@student.unimelb.edu.au

Kim Dalziel
kim.dalziel@unimelb.edu.au

Nancy Devlin
nancy.devlin@unimelb.edu.au

Brendan Mulhern
brendan.mulhern@uts.edu.au

¹ Centre for Health Economics Research and Evaluation (CHERE), University of Technology Sydney, Sydney, NSW, Australia

² Health Economics Unit, School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia

1 Introduction

Health-related quality of life (HRQoL) is an important outcome in the evaluation of healthcare interventions and treatments, and in allocating health resources [1]. Generic instruments measuring HRQoL are commonly used in economic evaluation to inform the estimation of quality-adjusted life-years (QALYs) [2]. QALYs measure health outcomes by combining quality of life with length of life. Preference-weighted HRQoL instruments have two components, a descriptive system and preference weights. The descriptive system usually comprises a number of domains, such as physical functioning, emotional well-being, etc., and the response levels associated with these domains. Health states, described by HRQoL instruments, may vary based on the domains and the way levels within each domain are defined.

A range of instruments with accompanying preference weights have been developed for use in the adult population, including the EQ-5D-3L [3], EQ-5D-5L [4], SF-6D v1 [5, 6] and SF-6D v2 [7]. Studies have found that these instruments might not be suitable for child and adolescent populations [8, 9]. Children have a different experience of health than adults. Evidence suggests that for children, the concepts of well-being and psychosocial health may be more important, whereas for adults, HRQoL tends to focus on the absence of illness or disability [10].

Due to differences in child development, experience and perspective, a range of paediatric-specific instruments have been developed for use in measuring child HRQoL [11]. Widely used and validated generic instruments of child HRQoL include the EQ-5D-Y-3L and EQ-5D-Y-5L [12], Child Health Utility 9 Dimension (CHU-9D) [13], Paediatric Quality of Life Inventory (PedsQL) [14] and the Health Utilities Index (HUI) [15]. Even though these instruments are all developed to measure generic HRQoL, they differ in important ways. There are differences in how the instruments were developed. For instance, the CHU-9D was developed de novo [16] but the EQ-5D-Y was adapted from the adult version of the instrument, the EQ-5D [17]. Consequently, instruments might differ in terms of the HRQoL domains measured. There is little empirical evidence on the dimensionality (the domains of HRQoL measured by a set of items from different instruments) between these instruments. Where they do measure the same construct, they use different item wording or different response modes (e.g., statements vs. Likert scale items). Moreover, they use different response scales (e.g., frequency vs. severity and different numbers of response levels) and different recall periods (e.g., 1 month vs. today vs. usual).

On the assumption that the generic instruments aim to measure the same construct of HRQoL, it is possible to

investigate what domains are being measured by each and how the instruments converge and diverge in the overall assessment of HRQoL. If HRQoL is conceptualized as the overall construct that is being measured, then the domain structure can be used to explore and define the dimensionality between instruments within the overall construct. Comprehending this allows for the further understanding of instrument content and construct validity, and identifying the relationship across different instruments.

Research assessing dimensionality across instruments has been conducted for generic adult HRQoL measures [18] but has not been conducted for measures of paediatric HRQoL. Therefore, this study aimed to explore the dimensionality between four paediatric HRQoL instruments—the EQ-5D-Y-5L, CHU-9D, PedsQL and HUI, in proxy- and self-reported data, using factor analysis.

2 Methods

2.1 Data Source

Data from the Paediatric Multi-Instrument Comparison (P-MIC) study (data cut 2, dated 10 August 2022) were used [19]. Data cut 2 includes approximately 94% of the total planned P-MIC participants. This study is part of the wider Quality of Life in Kids: Key Evidence for Decision Makers in Australia (QUOKKA) research programme in Australia. A detailed summary of the P-MIC data collection is available from Jones et al. [20, 21]. This study focused on 5- to 18-year-olds in the sample, as the instruments for use in children aged 2–4 years of age are experimental. Children and adolescents aged 7–18 years were invited to self-report their own HRQoL. Proxies were used if the child was not able to report their health themselves, either due to younger age (< 7 years) or health problems. In the current paper, ‘child’ refers to both children and adolescents.

2.2 Instruments Included

The P-MIC study included a number of different generic and condition-specific instruments. The instruments that have been used in this study are the EQ-5D-Y-5L, PedsQL, CHU-9D and HUI3 (hereafter HUI). The EQ-5D-Y-5L, PedsQL and CHU-9D were administered to the whole sample as these instruments were part of the P-MIC core instrument set [20]. HUI, PROMIS-25 and AQL6D each were administered to a subset (approximately one-third) of the online panel to minimize respondent burden [20]. As this paper aimed to explore relationships among commonly used instruments, and given that HUI is the most frequently used

instrument among the three subset instruments, it was also selected for inclusion in this study.

EQ-5D-Y-5L: The EQ-5D-Y instruments, both the 3L and 5L, have shown validity and reliability in measuring HRQoL in children and adolescents [22]. Despite the P-MIC dataset including the EQ-5D-Y-5L and EQ-5D-Y-3L, this analysis opted to use only one of the EQ-5D-Y instruments, as both versions share the same dimensions. The decision to include EQ-5D-Y-5L over EQ-5D-Y-3L was supported by evidence indicating its psychometric advantages [23]. With the EQ-5D-Y-5L being the newer instrument, there is a need for evidence about how it relates to other generic instruments. The EQ-5D-Y-5L measures five dimensions of health using single items: ‘able to walk around’, ‘looking after myself’, ‘doing usual activities’, ‘having pain or discomfort’, and ‘feeling worried, sad, or unhappy’. Each dimension assesses severity across five severity response levels ranging from no problems to unable to/extreme problems [4]. Proxy- and self-report versions of the EQ-5D-Y-5L have been developed and were used according to instrument age recommendations for self-report above age 7 years.

CHU-9D: The CHU-9D is designed for measuring HRQoL [24] in children aged 7–11 years. It was developed with and for children as a de novo measure. The CHU-9D has nine dimensions and each dimension has five severity response categories (from ‘no’ (don’t feel) to ‘very’ in five items, and ‘no problem’ to ‘can’t do’ in four items). The dimensions are ‘worry’, ‘sadness’, ‘pain’, ‘tiredness’, ‘annoyed’, ‘school’, ‘sleep’, ‘daily routine’ and ‘joining in with activities’ [13, 25].

PedsQL: The PedsQL Generic Core 4.0 has been widely used to assess HRQoL in children and adolescents [26]. It includes versions for different age groups: 5–7, 8–12, and 13–18 years. Proxy versions are available for children aged 2–4, 5–7, 8–12, and 13–18. The instrument was designed to measure core health dimensions and is based on guidelines from the World Health Organization (WHO) [26]. The PedsQL is composed of 23 items that measure four broad domains, defined as ‘physical’, ‘emotional’, ‘social’ and ‘school’ functioning. Each item has five frequency levels (from never to almost always) [26]. PedsQL was included in the study even though it was not designed to be preference weighted, because it is the most comprehensive instrument in terms of scope of items, validation across all child ages, and potentially the most used generic paediatric QoL instrument [27].

HUI: The HUI Mark 2 and 3 (HUI2/3) can be used to assess a child’s HRQoL. The HUI2 was developed to address the global morbidity burden of childhood cancer, while the HUI3 was developed to resolve certain issues related to the definitions of HUI2, ensuring applicability in both clinical and general population studies, and was therefore selected for inclusion in this study. The HUI3 [15] has eight domains

measured across 15 items, each with five or six response levels. The domains are ‘vision’ (2 items), ‘hearing’ (2 items), ‘speech’ (2 items), ‘ambulation’ (2 items), ‘dexterity’ (1 item), ‘emotion’ (2 items), ‘cognition’ (2 items) and ‘pain’ (2 items) [28, 29]. HUI measures typically are designed for people aged 5 years and older.

As the HUI was administered to one-third of the online panel, fewer data are available for exploring dimensionality between the HUI and the other instruments.

2.3 Data Analysis

2.3.1 Convergence Assessment Using Correlations

Correlation among variables was used in factor analysis to model latent factors. To provide a basis for the dimensionality assessment, the convergence and divergence between the items and dimensions were assessed using Spearman’s correlations (that assumes a non-normal distribution). We prespecified that correlation scores < 0.3 were considered weak, scores between 0.3 and < 0.5 were considered moderate, and scores of ≥ 0.5 indicated a strong correlation [30]. The correlation was assessed to see if the instruments were measuring the same construct and for the presence of outliers.

2.4 Factor Analysis

Choice of Factor Analysis Approach Common methods to detect the dimensionality in variables include principal components analysis (PCA) and exploratory factor analysis (EFA). The main aim of each method is slightly different; PCA is a technique for reducing the dimensionality of the data, whereas EFA is a method for identifying and measuring latent variables or factors, which cannot be measured directly. As we aimed to investigate the dimensionality of the item pool, without imposing a pre-existing model framework, EFA was used. To assess the overall dimensionality, the items from the four instruments were pooled for the self- and proxy-reported data separately. The Stata default EFA method, principal factor, was used for the estimation method. As children were aged 5–18 years in the proxy-completed data and 7–18 years in the self-reported data, the EFA was also applied for proxy-reported data for children aged 7–18 years to see if the results differed.

Data Check for Suitability Kaiser–Meyer–Olkin (KMO) and Bartlett’s test of sphericity were used to assess the suitability of the data for EFA. KMO examines the strength of the partial correlation between variables, and ranges between 0 and 1; a value close to 1 suggests that the sum of the partial correlations is relatively small in comparison with the sum of the correlations. This indicates that the correlations tend to be concentrated and cluster among a few variables,

which is advantageous for factor analysis. A rule of thumb for interpreting the KMO is that values between 0.8 and 1 indicate the sample is adequate to run a factor analysis. The significance level for Bartlett's test should be below 0.05. A p -value < 0.05 on Bartlett's test indicates that individual variables are sufficiently correlated for a factor analysis to be accomplished.

Choosing Factors and Items When using EFA, a range of indicators could inform the identification of the most appropriate domain structure. The structure can be decided based on the number of factors included in the model, items representing each factor, and the correlation between items.

The number of factors in EFA can be decided using the number of eigenvalues. The eigenvalue shows the variance explained by factors. The rule of thumb for choosing factors is based on eigenvalues > 1 , but factor structures can be forced to extract a certain number of factors. This choice can be supported by scree plots (that plot the eigenvalues) and parallel analysis (determines the number of factors based on eigenvalues). To check a range of possible factor structures, 6, 7, 8, 9, 10, and 11 factor models were identified and assessed for interpretation. Hereafter, factors resulting from EFA will be referred to as domains.

Each factor consists of items; to choose the items representing each factor, loadings were used. Loadings are the correlation between item and factor and uniqueness is the variance that is unique to that item in the model, represented as $(1 - \text{loading}^2)$. Items were kept when loadings were more than 0.32 [31, 32] for each factor. Factor loadings < 0.32 usually indicate a poor correlation between the items and the factor [33]. If there was cross-loading, i.e., if an item had a loading > 0.32 on two factors, we chose the factor with the higher loading. Due to the assumption of correlation between factors, the oblique method of rotation (Promax) was used. Factor correlation was confirmed using a correlation matrix. All analyses were performed using Stata software version 17.0 (StataCorp LLC., College Station, TX, USA) [34].

3 Results

3.1 Sample

Of the 6787 participants available in the P-MIC data cut, 5949 were children aged 5–18 years, of whom a total of 1728 fully completed all four instruments of interest (EQ-5D-Y-5L, CHU-9D, PedsQL, and HUI) and their responses were used in this study. From those who fully completed all four instruments, 604 responses were completed by proxies and 1124 were self-reported. Comrey and Lee [35] suggested a sample size rating scale for a factor analysis sample size of 100 as poor, 200 as fair, 300 as good, 500 as very

Table 1 Demographic summary of respondents completing all four instruments

	Self-complete($n = 1124$)	Proxy complete ($n = 604$)
	n (frequency)	n (frequency)
Age (years)		
5–6	–	294(48.68%)
7–10	467 (41.55%)	114 (18.87%)
11–15	433 (38.52%)	130 (21.52%)
16–18	224 (19.93%)	66 (10.93%)
Gender of the child		
Male	587 (52.22%)	320 (52.98%)
Female	520 (46.26%)	277 (45.44%)
Other	17 (1.52%)	7 (1.16%)

good, and 1000 or more as excellent. The proxy- and self-reported data show a very good and excellent sample size for EFA, respectively. A demographic summary of respondents is provided in Table 1.

Appendix Table 2 lists the brief item descriptions.

3.2 Convergence Assessment

Figures 1 and 2 display the results of the correlation between the pooled items for self- and proxy-reported data. The scale at the bottom of each figure presents the correlation, which ranges from 0 to 1. Rectangles separated by the red lines indicate a correlation between items from different instruments. Self- and proxy-reported results indicated a moderate correlation between most of the items and highlight items related to senses as outliers. In terms of both self- and proxy-reported data, the pain items from EQ-5D-Y-5L and CHU-9D exhibited the strongest correlations, with coefficients of 0.71 and 0.77, respectively. Detailed correlation tables can be found in Online Resource Table 1–6.

3.2.1 Self-Reported Data

Items regarding senses (which are all from HUI) were correlated with each other and had a very weak correlation with other items. Sleeping items from PedsQL and CHU-9D showed a strong correlation of 0.67, while items related to pain had a high correlation. Emotional functioning items from different instruments showed a strong correlation of < 0.6 , especially the items 'sad' and 'worried' from CHU-9D and 'worried, sad and unhappy' from EQ-5D-Y-5L. 'Taking a bath' from PedsQL had a strong correlation with 'looking after self' from EQ-5D-Y-5L, which also asks about washing and dressing (see Fig. 1).

Some items correlated moderately with more than one item; for instance, 'hurting' from PedsQL had a high

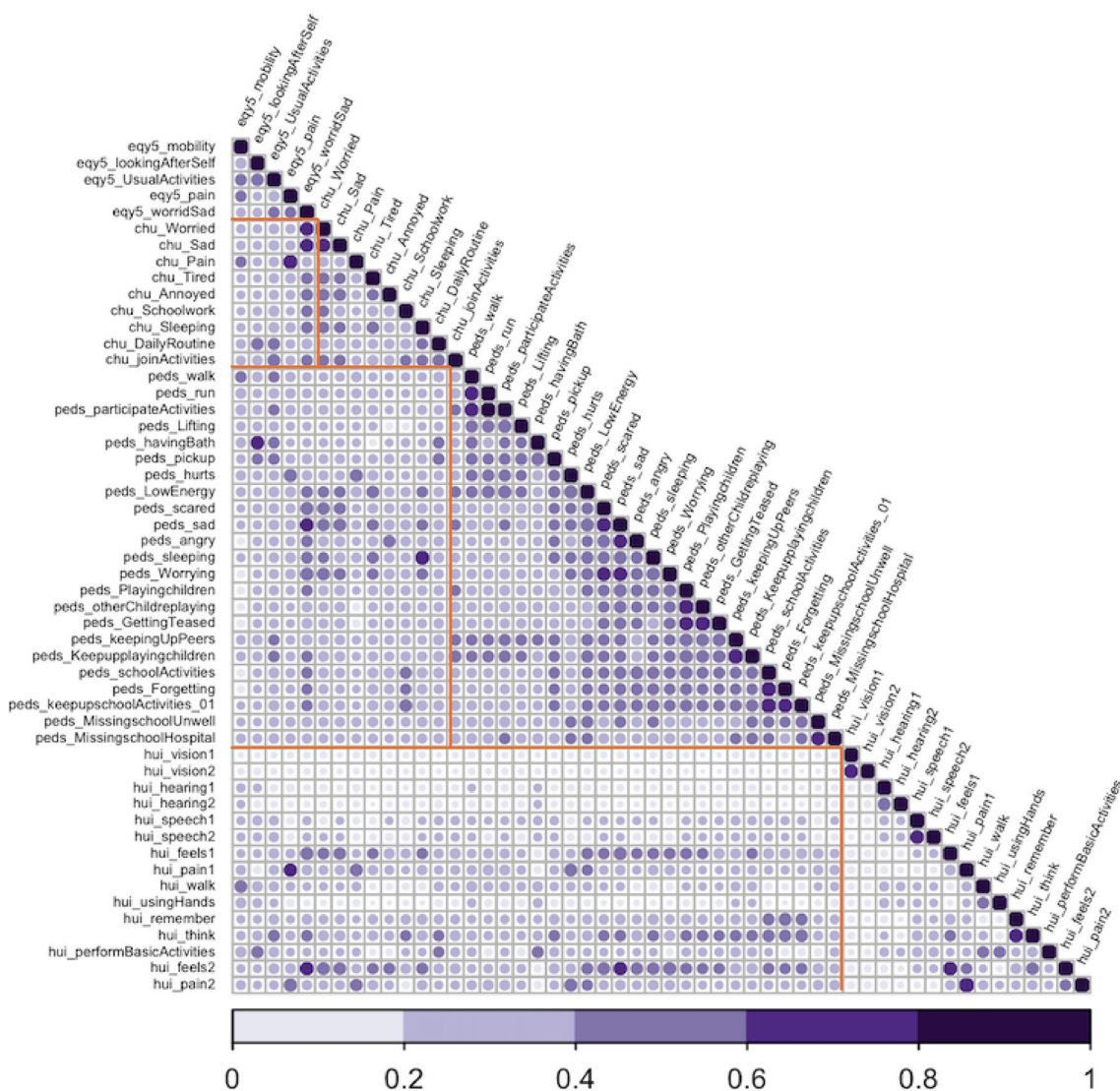


Fig. 1 Correlation matrix results from spearman correlation self-reports (children 7-18 years, $n = 1124$)

correlation with some of the physical function items from PedsQL.

3.2.2 Proxy-Reported Data

The correlation results from proxy-reported data shared similarities with the self-reported results. Items related to pain showed strong correlations. ‘Join activities’ from the CHU-9D and ‘usual activities’ from the EQ-5D-Y-5L also exhibited a strong correlation (0.61). The items related to senses and vision showed a weak correlation, however correlation was stronger compared with the self-reported data. Items related to cognition had a strong correlation, for instance the PedsQL item ‘forgetting’ and the HUI items ‘remember’ and ‘think’ (see Fig. 2).

3.3 Dimensionality Assessment

The data were tested for suitability prior to running the EFA. The KMO was 0.937 and 0.951 for proxy- and self-reported data, respectively, and both groups had a significant Bartlett’s test (p -value < 0.001). Thus, the data were suitable for EFA.

3.3.1 Pooled Item Model—Self-Reported Data

The EFA resulted in seven eigenvalues > 1 (17.62, 3.03, 2.30, 1.76, 1.29, 1.16, 1.01), indicating seven factors to extract according to the Kaiser criterion. Parallel analysis also suggested seven factors. The variance explained by the seven factors was 91.09%. Thus, seven factors were chosen for the self-reported pooled data. Figure 3 presents

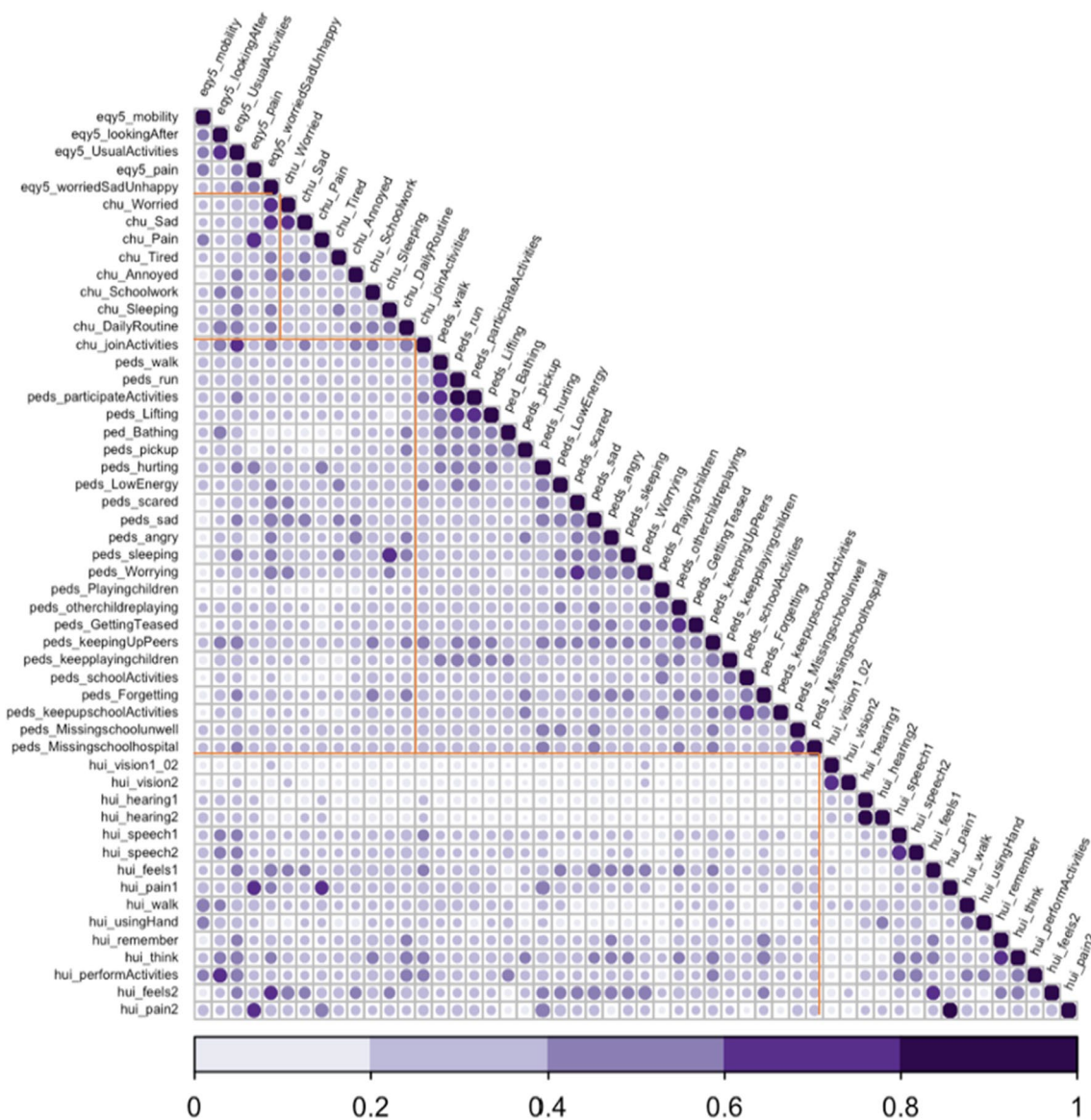


Fig. 2 Correlation matrix results from spearman correlation proxy-reports (children 5-18 years, $n = 604$)

the domain structure identified, including the factor loadings for each item. The domains were defined as follows.

1. *Emotional functioning*, which included one EQ-5D-Y-5L item, i.e. ‘worried, sad or unhappy’, alongside four PedsQL items, six CHU-9D items and two HUI items. The factor loadings ranged from 0.32 to 0.90. The items with the highest loadings were ‘sad’ and ‘worried’ from the CHU-9D, whereas the item with the lowest loadings was ‘angry’ from the PedsQL.
2. *Daily activities* included three items from the EQ-5D-Y-5L, five HUI items, and one item each from PedsQL and

CHU-9D. The item with the highest loading was EQ-5D-Y-5L ‘looking after self’, followed by the HUI item focused on ‘performing basic activities’. The usual activity items from the other instruments were also included, alongside broader activity-related constructs such as dexterity and communication. The loadings ranged from 0.34 to 0.81.

3. *Cognition/school functioning* did not include any of the EQ-5D-Y-5L items but included the ‘schoolwork’ item from CHU-9D, five items from PedsQL, and two cognition-related items from HUI. Factor loadings ranged

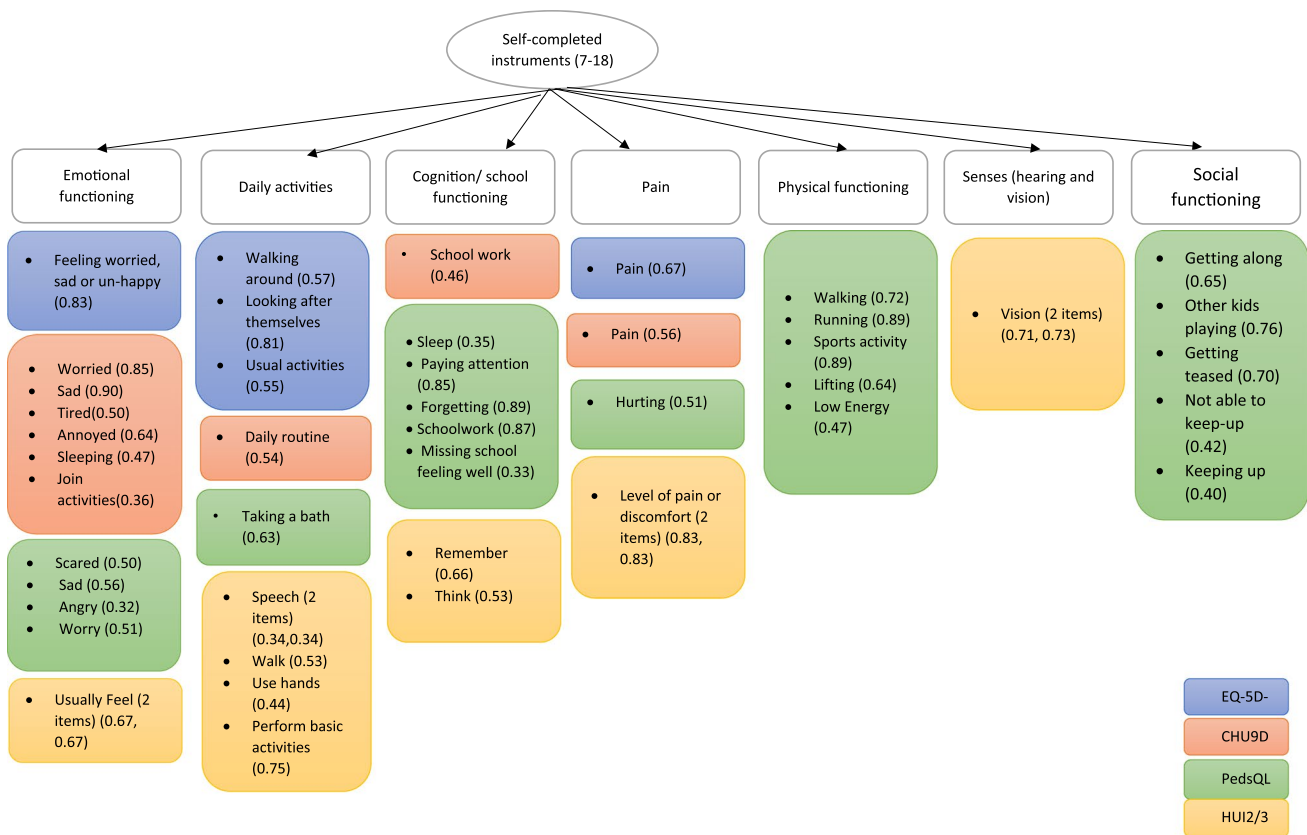


Fig. 3 Conception model of exploratory factor analysis results for items pooled from all instruments and item loadings -using self-reports

from 0.33 to 0.89; the lowest loading was related to ‘missing school’ from the PedsQL, and the highest was related to ‘forgetting’, also from the PedsQL.

- Pain:** This domain included all the pain items from all the instruments, with one item each from EQ-5D-Y-5L, CHU-9D and PedsQL, and two HUI items. The loadings ranged from 0.51 to 0.83. The item with the lowest loading was the PedsQL ‘hurting’.
- Physical functioning:** All items loading on this domain were from the PedsQL. Item loadings ranged from 0.47 to 0.89. The item with the lowest loading was ‘low energy’.
- Senses** included two items related to ‘vision’ and had a loading > 0.7 . Another sense item was ‘hearing’ but as the loading was < 0.32 , it did not load on this factor. All items were from the HUI.
- Social functioning** had five items from PedsQL, with the loadings ranging from 0.40 to 0.76. The lowest loading was related to ‘keeping up’ and the highest was ‘other kids playing’.

Three domains (emotional functioning, daily activities, and pain) had items from all instruments. Items in the physical functioning domain were all from PedsQL,

demonstrating that the EQ-5D-Y-5L ‘walking around’ and HUI ‘walk’ items did not have a strong relationship with the PedsQL physical functioning items.

Three items had cross-loadings; PedsQL ‘hurting’ cross-loaded on the pain and physical functioning domains, ‘feeling afraid or scared’ and ‘feeling sad or blue’ had loadings on both the emotional functioning and social functioning domains, and the PedsQL ‘hurting’ item loaded on both pain and physical functioning domains. Two of the PedsQL items, i.e. ‘doing chores’ and ‘missing school due to doctor or hospital visit’, did not load on any of the factors at a level of 0.32 or above.

The results indicated that modelling more factors divides items related to the same domain, such as emotional functioning or senses, into multiple domains.

3.3.2 Pooled Item Model—Proxy-Reported Data

The EFA resulted in six eigenvalues > 1 (17.21, 3.41, 3.12, 2.07, 1.56, 1.09), indicating six factors to extract according to the Kaiser criterion. Parallel analysis also suggested six factors. The variance explained by the six factors was 86.30%. Therefore, the six-factor model was identified and used as the best fitted. All PedsQL items, except

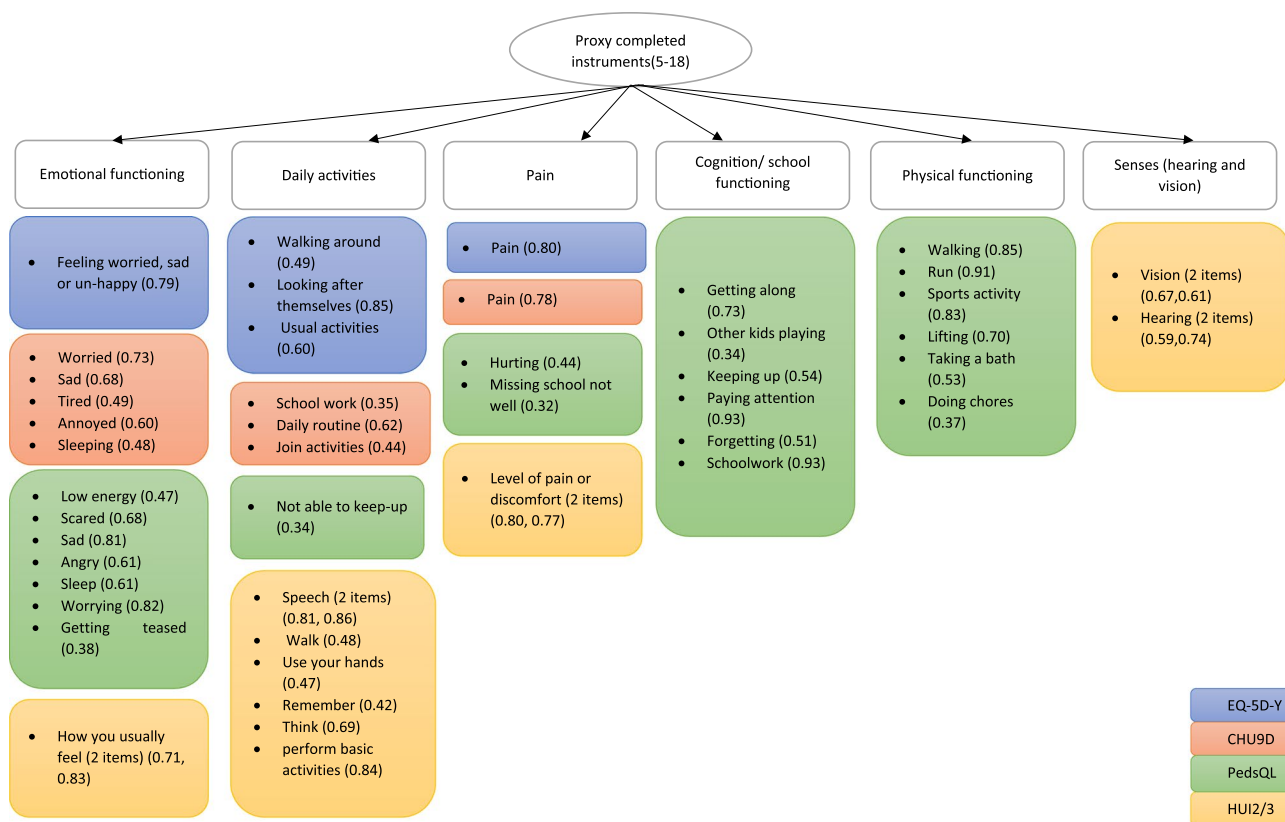


Fig. 4 Conception model of exploratory factor analysis results for items pooled from all instruments and item loadings- using proxy-reports

‘missing school due to doctor or hospital visit’, loaded for this model. Figure 4 presents the domain structure identified, with the factor loadings for each item included. The domains were defined as follows.

1. *Emotional functioning*, which included one EQ-5D-Y-5L item ‘worried, sad or unhappy’, alongside seven PedsQL, five CHU-9D, and two HUI items. The factor loadings ranged from 0.38 to 0.83. The item with the lowest loading was PedsQL ‘getting teased’.
2. *Daily activities* included three items from the EQ-5D-Y-5L, three items from CHU-9D, one item from PedsQL, and seven items from HUI. The loadings ranged from 0.34, which was the PedsQL ‘not able to keep up’ item, to 0.86, which was the PedsQL item regarding speech.
3. *Cognition/school functioning* included seven items from PedsQL; the highest loading was 0.93 for two items, i.e. ‘schoolwork’ and ‘paying attention’, and the lowest loading was 0.34 for ‘other kids playing’.
4. *Pain*: This domain included all items related to pain across all instruments, with one item from EQ-5D-Y-5L, one item from CHU-9D, two items from PedsQL,

and two HUI items. The loadings ranged from 0.32 to 0.80. The item with the lowest loading was the PedsQL item ‘missing school feeling unwell’. HUI ‘pain’ items and EQ-5D-Y-5L ‘pain’ had the highest loadings.

5. *Physical functioning* items were all from PedsQL. The loadings ranged from 0.37 to 0.91; the item with the lowest loading was ‘doing chores’ and the item with the highest loading was ‘run’.
6. *Senses* included HUI items related to hearing and vision, and the loadings ranged from 0.59 to 0.74.

Similar to the self-reported data, the proxy-reported data also had three domains (emotional functioning, daily activities, and pain) consisting of items from all instruments.

The cognition/school functioning domain only had items from PedsQL, and the senses domain had items only from HUI. The observation that PedsQL items load on the same factor may reflect that the PedsQL has more items, but could also be because these items originated from the same instrument and share the same wording and framing.

The ‘walking around’ item from the EQ-5D-Y-5L had a cross-loading on the pain domain (loading 0.35), and ‘taking

a bath' from PedsQL had cross-loadings between the physical functioning and daily activities domains. 'Keeping up' cross-loaded between the school functioning and physical functioning domains. This could be attributed to the wording of the item, i.e. 'keeping up when playing with other children', which can be understood as either keeping up during play as a leisure activity or as keeping up while playing games or sports at school.

When the analysis was conducted for proxy-reported data for children aged 7–18 years, it resulted in the same six domain structures as the model based on proxy-reported data for children aged 5–18 years; however, five items were loaded on different domains. Items such as 'schoolwork' from CHU-9D and 'doing chores' from PedsQL loaded on cognition/school functioning, while 'keeping up' loaded on physical functioning and 'remember' loaded on emotional functioning. One of the hearing items did not load on any domain (Appendix Fig. 5, for children aged 7–18 years).

All EFA factor loadings for the proxy-reported data 6 factor and self-reported data 7 factor can be found in Online Resource Tables 7 and 8.

4 Discussion

This study reports the use of EFA to explore the dimensionality across four commonly used paediatric HRQoL instruments, using data from both proxy- and child self-reports. This study builds on earlier work assessing the dimensionality of item pools [18] and is unique in applying EFA to explore the domain structure of HRQoL instruments used for paediatric populations and comparing the results between self- and proxy-reported data. Results show a seven- and six-factor structure for self- and proxy-reports, respectively. A different number of domains resulting from the self- and proxy-reported data indicates varying models of HRQoL depending on whose perspective is being considered.

Due to the inclusion of diverse population groups in the study (e.g. 'proxy-reported' data, compared with 'self-reported' data), structural differences exist in the selection of items for health domains. This suggests a disparity in the experience of health between children and proxies, aligning with previous research [36, 37]. The findings from these studies indicate that children's perception of QoL diverges from that of their parents, particularly in domains focused on social and psychosocial aspects. Items related to socializing with peers from PedsQL, such as 'other kids not wanting to play with him or her' and 'getting teased by other children' loaded on social functioning for self-reported data,

whereas in the proxy-reported data, these items loaded on the cognition and school activities domain. This might be due to varied interpretation of the items; for instance, children are more likely to associate socializing with what others are thinking and feeling during an interaction [38]. Proxy reports are often found unreliable when it comes to domains that require interpretation, such as social functioning. Proxy views have been found to be more reliable for observable concepts [39, 40].

The results have implications for the assessment of HRQoL using these paediatric instruments. In both self- and proxy-reported data, items from all instruments load onto three domains: emotional functioning, pain, and daily activities. These three domains are included in almost all the generic HRQoL measures for children, as they are essential components of HRQoL assessment, each contributing to a comprehensive understanding of a child's health. Other domains resulting from EFA were physical functioning, cognitive/school functioning, and senses, with self-reported data having an extra domain defined as social functioning. The domains show how different instruments broaden the HRQoL concepts being measured.

The results indicate what the instruments measure when pooling items from different HRQoL instruments. For instance, in this study, the EQ-5D-Y-5L does not measure five distinct domains but combines the physical items together (e.g. mobility, usual activities, and looking after self) with separate coverage of mental health, and, in certain models, pain has also been identified elsewhere for the adult instrument [41]. Similarly, the CHU-9D, which was developed as a nine-domain preference-weighted measure, mainly loads onto emotional functioning. This may indicate that the CHU-9D is more appropriate for assessing the emotional functioning of HRQoL. PedsQL also differed from its intended conceptual framework by loading onto more than four domains. This could be attributed to factors being driven by the pool of items from different instruments.

When interpreting the results, it is important to acknowledge that these issues could be caused by differences in item wording or response levels across the instruments. For instance, items related to 'sleep' load on different domains. The item from PedsQL loads on cognition/school functioning, while the item from CHU-9D loads on emotional functioning, even though both pertain to the same aspect of sleep; or the physical functioning items from PedsQL load onto a different factor to the physical functioning items from other instruments, which mainly loaded on daily activities. The way in which items have been framed within the measures, commonly known as the framing effect, can wield a significant influence [42]. An additional consideration arises

regarding how items interrelate and affect different aspects of health, even though they load on the same domains [41, 43, 44]. Item characteristics play an important role in how an item loads on a factor and what it is measuring in the context of HRQoL. Further research may help to better understand the underlying causes of these variations.

Cognition/school functioning are known to impact the HRQoL in children and adolescents [45]. Many of the commonly used generic HRQoL instruments for children have a domain related to cognition or school activities. Cognitive/school functioning, which is mostly related to concentration and school functioning, was a domain resulting from the EFA in both self- and proxy-reported datasets. In the self-reported analyses, all instruments (except EQ-5D-Y-5L) included items loading on cognition/school functioning; however, for the proxy-reported data, all items included in cognition/school functioning were from PedsQL, while items from other instruments loaded on daily activities. Items related to cognition from HUI loaded on the daily activities domain in proxy-reported data, but loaded on the cognition/school functioning domain for the self-reported data. This suggests that proxies may perceive cognitive functioning as influencing a child's 'usual activities', while children themselves consider cognitive function as directly linked to school-related activities when responding to cognitive ability-related items [45]. This could be due to differences in how children and proxies interpret HRQoL domains.

A method that can help to explore dimensionality could be confirmatory factor analysis (CFA), especially when exploring the conceptual overlap between items. We did not use it here, as the aim of this study was to explore the dimensionality across instruments without imposing a pre-existing model framework.

A limitation of this study is that it does not include disease-specific instruments in the analysis, which might lead to the omission of some important domains related to specific

conditions. However, on the other hand, their inclusion may create a bias towards the HRQoL impacts of a particular condition and would therefore need repeating across condition groups. Another potential limitation of this study might be the absence of age-stratified analyses due to an insufficient number of participants within each age group. We also did not consider children below 5 years of age because measurement of HRQoL in this age group is challenging and many HRQoL instruments in this age group remain experimental. This could be the subject of further research.

Overall, this study underscores the importance of considering the HRQoL instrument's design and structure when interpreting factor loadings. The similarity of items within factors may not just reflect underlying constructs but may also be influenced by the similarities in item characteristics and conceptual framing within an instrument. Recognizing these provides valuable insights into the relationship between instruments and factor analysis outcomes.

5 Conclusion

This study provides new evidence regarding the dimensionality exploration of four commonly used paediatric generic instruments. This is important information for users of these instruments because by clarifying the interrelationships among these instruments, users can gain insights into their strengths, limitations, and potential interactions. Slight differences in the domains have resulted from self- and proxy-reported data, with self-report having an extra domain for social functioning. Items within each domain also differ, especially regarding daily activities and cognition/school functioning, which shows the different views of children and proxies.

6 Appendix

See Table 2 and Fig. 5

Table 2 Abbreviation table

Domain	Items full name	Abbreviation used in paper	
EQ-5D-Y-5L			
Mobility	Walking around	Walking around	
Self-care	Looking after themselves	Looking after themselves	
Usual activities	Doing usual activities	Usual activities	
Pain and discomfort	Pain	Pain	
Anxiety /depression	Feeling worried, sad or un-happy	Feeling worried, sad or unhappy	
CHU9D			
Worry	Worried	Worried	
Sadness	Sad	Sad	
Pain	Pain	Pain	
Tiredness	Tired	Tired	
Annoyance	Annoyed	Annoyed	
School	School work	School work	
Sleep	Sleeping	Sleeping	
Daily routine	Daily routine	Daily routine	
Activities	Join Activities	Join Activities	
PedsQL			
Physical	Walking more than one block	Walk	
	Running	Run	
	Participating in sports activity or exercise	Sport activities	
	Lifting something heavy	Lifting	
	Taking a bath or shower by him or herself	Taking bath	
	Doing chores around the house	Doing chores	
	Having hurts or aches	Hurting	
	Low energy level	Low energy	
	Emotional	Feeling afraid or scared	Scared
		Feeling sad or blue	Sad
Feeling angry		Angry	
Trouble sleeping		Sleep	
Social	Worrying about what will happen to him or her	Worry	
	Getting along with other children	Getting along	
	Other kids not wanting to be his or her friend	Other kids playing	
	Getting teased by other children	Getting teased	
	Not able to do things that other children his or her age can do	Not able to keep up	
School	Keeping up when playing with other children	Keeping up	
	Paying attention in class	Paying attention	
	Forgetting things	Forgetting	
	Keeping up with schoolwork	Schoolwork	
	Missing school because of not feeling well	Missing school unwell	
	Missing school to go to the doctor or hospital	Missing school hospital	
HUI			
Vision	Ability to see well enough to read ordinary newsprint	Vision 1	
	Ability to see well enough to recognize a friend on the other side of the street	Vision 2	
Hearing	Ability to hear what was said in a group conversation with at least three other people,	Hearing 1	
	Ability to hear what was said in a conversation with one other person in a quiet room	Hearing 2	
Speech	Ability to be understood when speaking your own language with people who do not know you	Speech 1	
	Ability to be understood when speaking with people who know you well	Speech 2	
Emotion	How you usually feel	Feel 1	
Pain	Level of pain or discomfort	Pain 1	

Table 2 (continued)

Domain	Items full name	Abbreviation used in paper
Ambulation	Ability to Walk	Walk
Dexterity	Ability to use your hands and fingers	Use hands
Cognition	Ability to remember things	Remember
Cognition	Ability to think and solve day to day problems	Think
Ambulation	Ability to perform basic activities	perform basic activities
Emotion	How you usually feel	Feel 2
Pain	Level of pain or discomfort	Pain 2

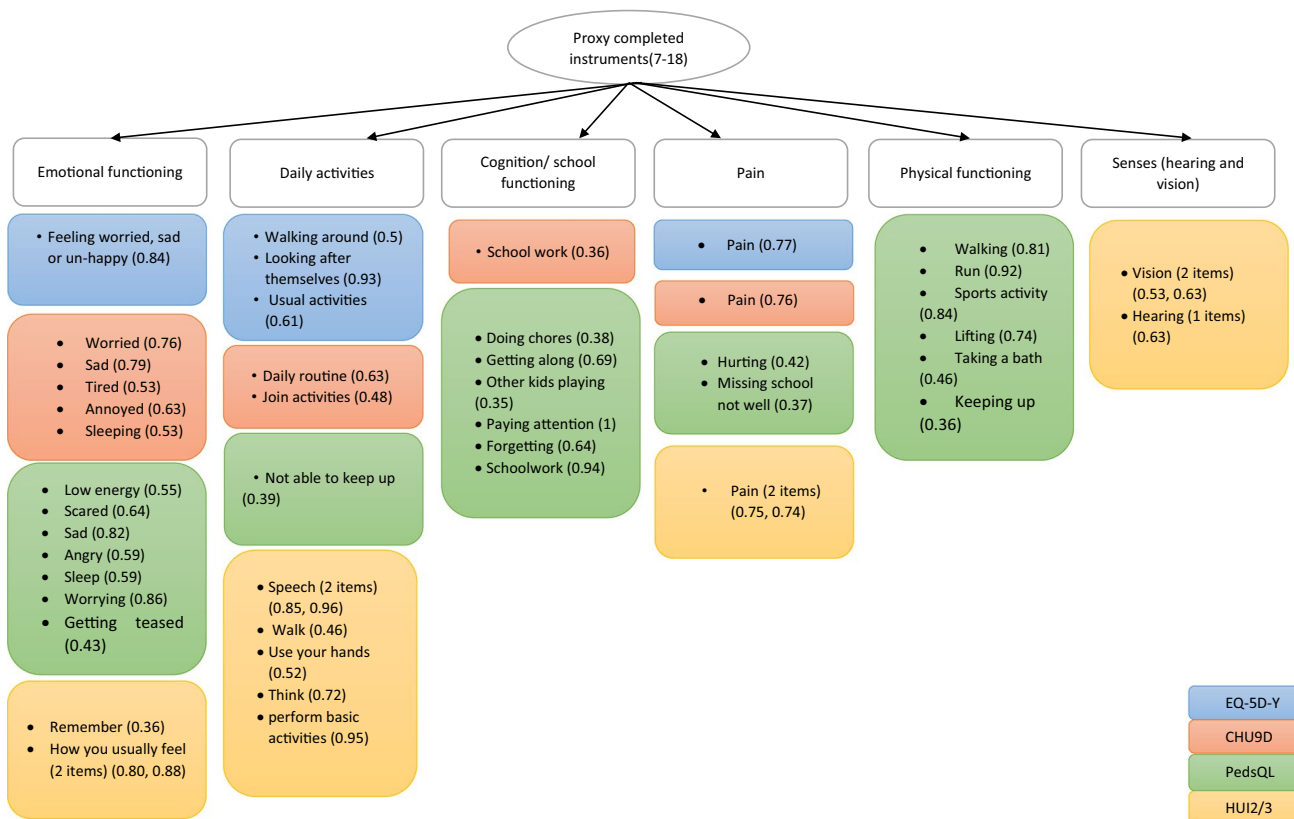


Fig. 5 Conception model of EFA results for items pooled from all instruments (proxy-complete) and item loadings (7–18 year old)

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-024-01382-y>.

Acknowledgements The authors would like to thank the QUOKKA study investigators, and research fellows and research assistants, especially the P-MIC team for designing and completing data collection. They would also like to thank the children and families for taking the time to complete the study, as well as members of QUOKKA’s consumer advisory group and decision-making panel for their input.

Author Contributions BM, MB, ND and KD conceptualized the study; RJ and MB worked on data validation; BM and MB worked on the methodology and analysis; and MB prepared the first draft and all authors reviewed and edited the draft. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This research was funded by an Australian Government Medical Research Futures Fund (MRFF) grant (1200816) and a EuroQol Research Foundation grant (361-RA). The views expressed in this paper are those of the authors and are not necessarily those of the EuroQol Research Foundation.

Disclosure This article is published in a journal supplement wholly funded by the Australian Government MRFF funded QUOKKA research grant 1200816, the University of Melbourne, the EuroQol Research Foundation, the University of Technology Sydney and Flinders University.

Declarations

Ethics approval Data collection for this study was approved by The Royal Children's Hospital (RCH) Human Research Ethics Committee (HREC/71872/RCHM2021) on 20 May 2021.

Registration This study was prospectively registered with the Australia New Zealand Clinical Trials Registry (ANZCTR) [ACTRN12621000657820].

Data Availability The datasets generated and/or analyzed during the current study are available from the P-MIC study group on reasonable request.

Conflict of interest ND and BM are members of the EuroQol Group who is the developer and copyright holder of the instruments included in this study. Views expressed in this paper are those of the authors and are not necessarily those of the EuroQol Research Foundation.

Consent for Publication Not applicable.

Code Availability Not applicable.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Kromm SK, Bethell J, Kraglund F, Edwards SA, Laporte A, Coyte PC, Ungar WJ. Characteristics and quality of pediatric cost-utility analyses. *Qual Life Res.* 2012;21(8):1315–25.
- Richardson J, Schlender M. Health technology assessment (HTA) and economic evaluation: efficiency or fairness first. *J Mark Access Health Policy.* 2018;7(1):1557981–1557981.
- Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33(5):337–43.
- Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20(10):1727–36.
- Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002;21(2):271–92.
- Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care.* 2004;42(9):851–9.
- Brazier JE, Mulhern BJ, Bjorner JB, Gandek B, Rowen D, Alonso J, Vilagut GJE, Ware, and the SF-6Dv2 International Project Group. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care.* 2020;58(6):557–65.
- Matza LS, Swensen AR, Flood EM, Secnik K, Leidy NK. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value in Health.* 2004;7(1):79–92.
- Kreimeier S, Greiner W. EQ-5D-Y as a health-related quality of life instrument for children and adolescents: the Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. *Value in Health.* 2019;22(1):31–7.
- Ungar W. *Economic evaluation in child health.* Oxford: Oxford University Press; 2009.
- Kwon J, Freijser L, Huynh E, Howell M, Chen G, Khan K, Daher S, Roberts N, Harrison C, Smith S. Systematic review of conceptual, age, measurement and valuation considerations for generic multidimensional childhood patient-reported outcome measures. *Pharmacoeconomics.* 2022;40(4):379–431.
- Kwon J, Freijser L, Huynh E, Howell M, Chen G, Khan K, et al. Systematic review of conceptual, age, measurement and valuation considerations for generic multidimensional childhood patient-reported outcome measures. *Pharmacoeconomics.* 2022;40(4):379–431.
- Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qual Health Res.* 2010;20(3):340–51.
- Varni JW, Burwinkle TM, Seid M. The PedsQL™ as a pediatric patient-reported outcome: reliability and validity of the PedsQL™ Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res.* 2005;5(6):705–19.
- Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health Qual Life Outcomes.* 2003;1(1):54.
- Stevens KJ. *The development of a preference based paediatric health related quality of life measure for use in economic evaluation.* Sheffield: University of Sheffield; 2009.
- Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res.* 2010;19(6):875–86.
- Finch AP, Brazier JE, Mukuria C, Bjorner JB. An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: the EQ-5D Case Study. *Value in Health.* 2017;20(10):1362–75.
- Jones R, Mulhern B, Devlin N, Hiscock H, O'Loughlin R, McGregor K, Shilana Y, Dalziel K, on behalf of the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team. Australian Paediatric Multi-Instrument Comparison (P-MIC) Study. Technical Methods Paper [online]. Version 2. 2022.
- Jones R, B Mulhern, K McGregor, S Yip, R O'Loughlin, N Devlin, et al., on behalf of the Quality of Life in Kids Key Evidence to Strengthen Decisions in Australia (QUOKKA) project. Psychometric performance of HRQoL measures: an Australian Paediatric Multi-Instrument Comparison Study Protocol (P-MIC). *Children.* 2021;8(8):714. <https://doi.org/10.3390/children8080714>.
- Jones R, O'Loughlin R, Xiong X, Bahrapour M, Devlin N, Hiscock H, on behalf of the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia Project, et al. Comparative Psychometric Performance of Common Generic Paediatric Health-Related Quality of Life Instrument Descriptive Systems: Results from the Australian Paediatric Multi-Instrument Comparison Study. *Pharmacoeconomics.* 2023. <https://doi.org/10.1007/s40273-023-01330-2>.
- Golicki D, Młyńczak K. Measurement Properties of the EQ-5D-Y: A Systematic Review. *Value in Health.* 2022;25(11):1910–21.
- Pérez-Sousa MÁ, Olivares PR, Ramírez-Vélez R, Gusi N. Comparison of the Psychometric Properties of the EQ-5D-3L-Y and EQ-5D-5L-Y Instruments in Spanish Children and Adolescents. *Value in Health.* 2021;24(12):1799–806.

24. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual Life Res.* 2012;21(4):717–25.
25. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian Adolescent Population. *Value in Health.* 2012;15(8):1092–9.
26. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: reliability and validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in healthy and patient populations. *Med Care.* 2001;39(8):800–12. PMID: 11468499. <https://doi.org/10.1097/00005650-200108000-00006>.
27. Upton P, Eiser C, Cheung I, Hutchings HA, Jenney M, Maddocks A, Russell IT, Williams JG. Measurement properties of the UK-English version of the Pediatric Quality of Life Inventory™ Version 4.0 generic core scales. *Health Qual Life Outcomes.* 2005;3(1):1–7.
28. Gemke RJ, Bonsel GJ. Reliability and validity of a comprehensive health status measure in a heterogeneous population of children admitted to intensive care. *J Clin Epidemiol.* 1996;49(3):327–33.
29. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care.* 2002;40(2):113–28.
30. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–9.
31. Tabachnick BG, Fidell LS, Ullman JB. *Using multivariate statistics*, vol. 5. Boston: Pearson; 2007.
32. Hair JF, WC Black, BJ Babin, RE Anderson, RL Tatham. *Multivariate data analysis*. Vol. 6. 2006.
33. Comrey AL, HB Lee. *A first course in factor analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
34. StataCorp. *Stata statistical software: release 16*. College Station: StataCorp LLC; 2019.
35. Comrey AL, Lee HB. *A first course in factor analysis*. Berlin: Psychology press; 2013.
36. Jardine J, Glinianaia SV, McConachie H, Embleton ND, Rankin J. Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics.* 2014;134(4):e1129–48.
37. Pickard AS, Topfer L-A, Feeny DH. A structured review of studies on health-related quality of life and economic evaluation in pediatric acute lymphoblastic leukemia. *JNCI Monographs.* 2004;2004(33):102–25.
38. Klien G. Children and adults process social interactions differently: study reveals key differences in brain activation. *Neuroscience news*. Bangor University; 2023.
39. US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services, FDA Center for Biologics Evaluation and Research US Department of Health and Human Services FDA Center for Devices and Radiological Health. *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance*. *Health Qual Life Outcomes.* 2006;4(1):79.
40. Germain N, Aballéa S, Toumi M. Measuring the health-related quality of life in young children: how far have we come? *J Mark Access Health Policy.* 2019;7(1):1618661.
41. Gamst-Klaussen T, Gudex C, Olsen JA. Exploring the causal and effect nature of EQ-5D dimensions: an application of confirmatory tetrad analysis and confirmatory factor analysis. *Health Qual Life Outcomes.* 2018;16(1):153–153.
42. Peng J, Li H, Miao D, Feng X, Xiao W. Five different types of framing effects in medical situation: a preliminary exploration. *Iran Red Crescent Med J.* 2013;15(2):161–5.
43. Feng YS, Jiang R, Kohlmann T, Pickard AS. Exploring the internal structure of the EQ-5D using non-preference-based methods. *Value Health.* 2019;22(5):527–36.
44. Shah KK, Mulhern B, Longworth L, Janssen MF. Views of the UK general public on important aspects of health not captured by EQ-5D. *Patient.* 2017;10(6):701–9.
45. Ludwig K, Surmann B, Racker E, Greiner W. Developing and testing a cognitive bolt-on for the EQ-5D-Y (Youth). *Qual Life Re.* 2022;31(1):215–29.