ScienceDirect
Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Development of a Health-State Classification System for the Pediatric Quality-of-Life Inventory Version 4.0 Generic Core Scales for Preference-Based Valuation in Australia

Joseph Kwon, PhD, Rakhee Raghunandan, PhD, Son Hong Nghiem, PhD, Kirsten Howard, PhD, Emily Lancsar, PhD, Elisabeth Huynh, PhD, Martin Howell, PhD, Stavros Petrou, PhD, Sarah Smith, PhD

ABSTRACT

Objectives: Pediatric Quality-of-Life Inventory Version 4.0 Generic Core Scales (PedsQL GCS), comprising 23 items covering 4 subscales (physical, emotional, social, and school functioning), is a widely applied generic measure of childhood health-related quality of life but does not provide health utilities for cost-effectiveness-based decision making. This study aimed to develop a reduced item version of PedsQL GCS amenable to health utility derivation in Australia.

Methods: Data sources were 2 cohorts of the Longitudinal Study of Australian Children, including proxy responses for all PedsQL GCS versions (Toddlers, Young Children, Children, and Teens), and the CheckPoint sample containing child self-report to the Children version. Three analytic samples were CheckPoint sample (n = 1874); Mallinson sample containing 1 measurement per child from one of the Young Children, Children, or Teens versions (n = 7855); and Toddlers sample (n = 7401). Exploratory and confirmatory factor analyses assessed dimensionality. Psychometric analyses used Rasch and classical criteria on 3 randomly selected subsamples (n = 500) per sample. Item selection prioritized psychometric performance in the CheckPoint sample, also considering performance in other samples and conceptual content.

Results: Dimensionality assessments did not generate an alternative empirical structure for the measure, and psychometric analyses were conducted on the original 4 subscales. The selected items were: "Get aches and pains" for physical functioning; "Feel sad/blue" for emotional functioning; "Other kids not friends" for social functioning; and "Keeping up with school work" for school functioning.

Conclusions: The final 4-item set, pending further psychometric validation and valuation, can generate health utilities from the widely used PedsQL GCS to inform cost-effectiveness-based decision making.

Keywords: health-state classification system, health utility, item set selection, Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales, psychometric evaluation.

VALUE HEALTH. 2025; 28(1):88–98

Highlights

- The Pediatric Quality-of-Life Inventory 4.0 Generic Core Scales (PedsQL GCS) is one of the most widely used measures of childhood health-related quality of life. However, to date, it has not been possible to use it to estimate health utilities that can support health economic evaluations using quality-adjusted life year as outcome. To address this gap, a health-state classification system is developed from PedsQL GCS with which health utilities can be derived.
- Through robust dimensionality assessments and psychometric analyses, this study generated a reduced health-state classification system from the PedsQL GCS, which contains 4 items, 1 from each of the original subscales of the PedsQL GCS. Future research should create a value set to derive health utilities from this system.
- Once a value set is created, the reduced health-state classification system will greatly increase the availability of childhood health utility outcomes through the application of PedsQL GCS. This will, in turn, aid the conduct of cost-utility analyses that inform healthcare decision making in terms of the cost-effectiveness of alternative pediatric clinical and public health interventions.

Introduction

Scarce healthcare resources motivate health economic evaluations that compare interventions in terms of costs and consequences.[1] Many healthcare decision makers recommend cost-utility analysis as the form of health economic evaluation[2-5] because it uses a common outcome metric quality-adjusted life years, combining preference-based health utilities with the length of time in health states.[1] Generic multiattribute utility instruments (MAUIs), such as EQ-5D,[6,7] have been developed to derive health utilities. Typically, they consist of a small number of items that each represent a different dimension of the construct being measured. Together, items form a health-state classification system, which is valued by a relevant population using techniques such as time tradeoff.[8-10] Utilities are derived after weighting items by the value set. A small number of items in the system enables efficient valuations.[11] Recently, several MAUIs with small numbers of items have been derived from existing longer nonpreference-based instruments, although these have mostly been disease specific and none that target children.[12-19]

In children, there are particular challenges to measuring health utilities. For example, it is unclear at what age self-report by children is feasible[20] and how proxy and self-reports are related.[20-23] Bio-psychosocial development also means that

relevant health dimensions may differ between stages of childhood,[24] and the content of classification systems need to be different across age groups.[25] A recent systematic review identified 11 MAUIs specifically designed for use with children and 3 further MAUIs, whose target age ranges cover both children and adults.[26] These instruments (16D,[27] 17D,[28] AHUM,[29] AQoL-6D,[30] CH-6D,[31] CHSCS-PS,[32] CHU9D,[33,34] EQ-5D-Y-3L,[35] EQ-5D-Y-5L,[36] HUI2,[37] HUI3,[38] IQI,[39] QWB,[40,41] and TANDI[25,42]) either had inadequate content validity (eg, HUI2/3 and EQ-5D-Y), targeted narrow childhood age ranges (eg, age 7-11 years for CHU9D), or had both limitations.[43] Many also had gaps in their psychometric evidence, and there was no clear leading instrument in terms of psychometric robustness.[43] There is hence a need for an instrument that has been developed from a strong conceptual and psychometric base and measures health utilities in children across a wide range of age groups.

Pediatric Quality-of-Life Inventory 4.0 Generic Core Scales (PedsQL GCS) is a generic nonpreference-based measure of childhood health-related quality of life (HRQoL) developed from robust psychometric methods.[44,45] It includes 4 subscales: physical, emotional, social, and school functioning. Separate versions exist for Toddlers (age 2-4 years), Young Children (5-7 years), Children (8-12 years), Teens (13-18 years), and Young Adults (18-25 years). All age groups have both self- and proxy-reported versions and have the same 23 items, except for the Toddlers version, which is only proxy reported and has 3 rather than 5 items within the school subscale. Each item is reported on a 5-point Likert-type scale (except for the self-reported Young Children version which has a simplified 3-point scale).[26,44]

There are strong rationales for deriving a health-state classification system from PedsQL GCS that is amenable to valuation and health utility derivation. First, PedsQL GCS covers a wide range of childhood age groups using the same core constructs with only minor variations in item wording and response scales. Second, it has demonstrated acceptable performance on key psychometric criteria, including content validity in multiple settings.[46-53] Thus, any reduced classification system created from PedsQL GCS would have the strength of having selected the best performing items from a set that already has robust measurement properties. Third, PedsQL GCS is one of the most widely used measures of childhood HRQoL, with over 90 translations.[26,54,55] Health utilities could be retrospectively derived from previous applications if item-level responses are available. The continuing widespread use of PedsQL GCS likewise broadens the scope for prospective health utility derivation in childhood health contexts in which MAUI applications are not the norm.[56] This study aims to develop a health-state classification system based on PedsQL GCS for Australian children that is amenable to valuation and health utility derivation.

Methods

Data

Data sets

PedsQL GCS was developed as a generic instrument for use with healthy, acute, and chronically ill childhood populations.[55] This study analyses the Longitudinal Study of Australian Children (LSAC), which includes PedsQL GCS data across those populations. LSAC recruited a nationally representative cohort of 5000 children aged 0 to 1 years in 2003 and 2004 (cohort B) and a second cohort aged 4 to 5 years in 2003 and 2004 (cohort K). Both were followed longitudinally and age-appropriate PedsQL GCS versions applied. Around 20% of the cohorts have 1 or more common chronic health condition (asthma, anxiety/depression, ADHD, autism/Asperger's, epilepsy, and type 1 diabetes).[57] We obtained the latest version of LSAC from the Australian Data Archive.[58] We also use PedsQL GCS data from the LSAC Child Health CheckPoint study (CheckPoint), which conducted physical health and biomarker collection from cohort B children aged 11 to 12 years. Details on LSAC[59-61] and CheckPoint[62,63] are reported elsewhere.

Pediatric Quality-of-Life Inventory 4.0 Generic Core Scales

The CheckPoint data set contained self-reported data on Children version of PedsQL GCS, whereas LSAC cohorts contained proxy-reported data on all age versions of PedsQL GCS except Young Adults version. Hence, data on the following versions of PedsQL GCS were available for analysis: Toddlers (proxy reported), Young Children (proxy reported), Children (self- and proxy reported) and Teens (proxy reported). All versions used 1-month recall for the same 5-point response scale (0 = Never, 1 = Almost never, 2 = Sometimes, 3 = Often, and 4 = Almost always). Except for Toddlers version, all versions consist of 23 items representing 4 subscales of functioning: physical (8 items), emotional,[5] social,[5] and school.[5] There are minor differences in wording between versions, but the conceptual meaning of items is consistent. The Toddlers version consists of 21 items, with 2 fewer school items. Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.005 provides the full item wording. The analyses here used reversed ordinal responses rather than values on the transformed numerical 0 to 100 scales[44] because this is required for Rasch analysis.

Analytic samples

Figure 1 shows the LSAC and CheckPoint data sources composing the 3 analytic samples used.

The CheckPoint sample comprised self-reported data on Children version from CheckPoint. For Young Children, Children, and Teens proxy-reported versions, a combined analytic sample was created using the Mallinson method for Rasch analysis of repeated measures.[64] Specifically, from the 3 measurement points for a child in the given cohort (children without three measurements were excluded), 1 was selected randomly such that in the constructed cross-sectional data set, each child appears once, and measurement points from the 3 proxy-reported versions are evenly represented. This was applied separately for cohorts B and K and the data sets combined to form the Mallinson sample. Children aged 4 years in the cohort B wave 3 and cohort K wave 1 samples formed the Toddlers sample.

Selection of Items for Health-State Classification System

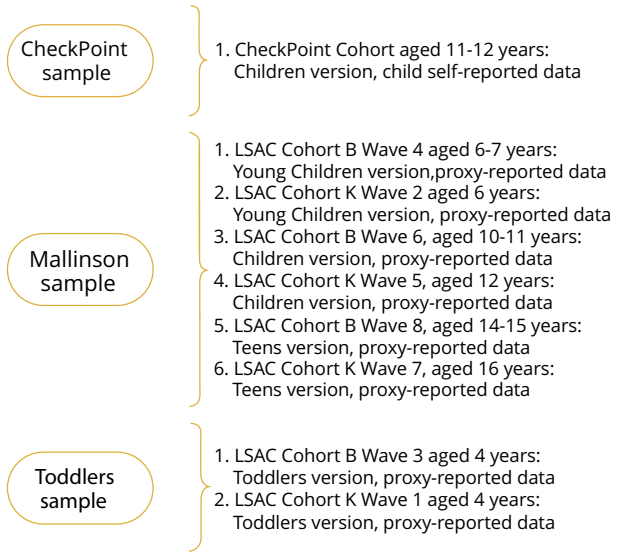
The aim was to identify 1 item per subscale/dimension that could be used in a health-state classification system for PedsQL GCS across all ages. The analysis comprised 3 stages (25-27)[1]: empirical determination of dimensionality,[2] psychometric analyses to understand item performance, and[3] selection of the best-performing item per subscale. RUMM2030[65] was used for Rasch analyses and STATA 17[66] for other analyses.

Stage 1: Determining dimensionality

Stage 1 aimed to understand the extent to which the 4 original subscales hypothesized for the PedsQL GCS[45] are empirically supported by LSAC data or whether there was an alternative structure. We used exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA).

We used principal axis factoring and varimax rotation for EFA, then the eigenvalue >1 rule and examination of scree plot to

Figure 1. Composition of main analytic samples derived from the LSAC and CheckPoint cohorts and waves.



LSAC indicates Longitudinal Study of Australian Children.

guide the extraction of factors. Subsequently, we specified 4 factors because 4 subscales were originally hypothesized by the developers of PedsQL GCS. Previous work has identified a possible 5 factor structure in which the fifth factor consisted of the 2 school absence items.[67] Because 2 items constitute only a very weak factor, our final EFA removed these 2 items and assessed whether a 4-dimensional structure could be identified. Items were considered to load on a factor if the loading was >0.4 and to cross-load if they loaded on multiple factors with the difference in loading coefficients <0.2.

CFA assessed how well the observed covariance matrix fitted the expected matrix for the following a priori structures[14]:

1. Original 4-factor structure of PedsQL GCS.
2. Five-dimensional structure identified by Varni and colleagues[67]: same as original structure but the last 2 school functioning items forming a separate fifth subscale. This was not applied to the Toddlers version because its school subscale contains 3 items.
3. Empirical structure obtained from EFA without constraint on the number of factors.
4. Empirical 4-factor structure from EFA.
5. Empirical structure from EFA after removing the last 2 school functioning items.

When assessing empirical fits, the root mean square error of approximation and the standardized root mean square residual of less than 0.05 indicated good fit, whereas values between 0.05 and 0.08 indicated reasonable fit. The comparative fit index and the Tucker and Lewis index values above 0.95 indicated good fit, whereas values between 0.90 and 0.95 indicated reasonable fit.[14]

The final decision on the dimensional structure for psychometric analysis was taken via group deliberation involving all coauthors who together have expertise in health economics, pediatrics, and psychometrics. This considered the EFA and CFA results plus face validity and conceptual coverage of the proposed structure. Literature on the initial development of PedsQL GCS[45] was also considered to assess the content validity of the original structure.

Stage 2: Psychometric analysis

Item analyses were undertaken to understand the performance of each item in relation to the subscale to which it was supposed to belong. We followed the approach in previous work[11,14] and used both Rasch and classical psychometric methods. Rasch analysis compares the observed questionnaire responses with what would be expected if the desirable characteristics of an interval scale were met.[68,69] Respondents are located on a latent scale representing their underlying levels on a dimension. The probability of a respondent endorsing a particular item response is assumed to be a logistic function of the relative distance between their position on the unidimensional latent scale and the position of the item response.[70]

As Rasch models assume unidimensionality, a separate model was evaluated for each of the subscales identified from dimensionality assessment. Unidimensionality was tested by applying a principal component analysis on the residuals of the Rasch model, and then conducting a series of t tests on the difference in the latent scale location of each person as measured by 2 subsets of items that have, respectively, positive and negative correlations between items and the first residual factor.[71] If statistically significant differences are found for more than 5% of the sample (ie, 95% confidence interval of the estimated proportion should contain at most 5%), the assumption of unidimensionality is no longer reasonable and the Rasch analysis less appropriate.

The overall fit to the Rasch model was examined by considering item and person fit residuals (criterion: mean residual approximately 0 with a standard deviation of approximately 1), item-trait interaction (nonsignificant χ^2 test, $P > .01$), and person separation index ≥0.7.[11] The following criteria were used to evaluate how well each item met the Rasch expectations:

(a) Threshold ordering: the item response levels should sequentially show the highest probability of endorsement as one progresses along the unidimensional latent scale. This generates an ordered set of response thresholds for item. The category probability curve was plotted for each item to identify disordered thresholds. Where disordered, adjacent levels were collapsed to improve the overall model fit. Items requiring adjustment were retained in the Rasch models but not considered for selection.
(b) Differential item functioning (DIF): DIF occurs when different groups within the sample show different scores for the same amount of the construct being measured. We evaluated DIF by age (CheckPoint and Mallinson samples) and sex (all samples). Uniform DIF refers to the between-group difference being constant (main effect); nonuniform DIF refers to the difference varying (interaction). Items with uniform DIF were adjusted and retained but not considered for selection.
(c) Item-level goodness of fit: poorly fitting items, with fit residuals outside the ± 2.5 range and statistically significant χ^2 test result, were retained but not considered for selection. Bonferroni correction for the number of items in the subscale was applied on the base-case significance threshold of P value < .01.

Item threshold range and mean location were also considered, but no explicit criteria were set. Several classical psychometric criteria[11,14,72,73] that address more descriptive elements of item performance were assessed: the rate of missing data (criterion: rate ≤ 5%[72]), internal consistency (corrected item-total correlations >0.2 and <0.8[73]); and floor/ceiling effects (rate <80%[72]).

For robustness, the Rasch analyses were repeated for 3 subsamples of size n = 500 randomly selected from each analytic sample using the random sampling function in RUMM2030. The size of n = 500 has been recommended for Rasch analysis to avoid type I error.[14,74,75] Classical psychometric analyses were conducted on the full analytic samples.

Stage 3: Item selection

To select the best item for each subscale, we eliminated items that failed the Rasch criteria for threshold ordering, DIF, or item-level goodness of fit in 2 out of 3 subsamples within each analytic sample. The classical psychometric criteria were then considered among the remaining items. The results from CheckPoint sample with self-reported data were prioritized because this was considered to best represent children's own views. The CheckPoint sample results were subsequently compared with those from the other samples to assess whether the psychometric performance of items was consistent across age groups and respondent types. Finally, the conceptual meaning of the items was considered, to ensure that the selected item adequately represents the subscale content.

Results

Sample Characteristics

Table 1 shows the characteristics of the 3 analytic samples.

Dimensionality Results

Exploratory factor analysis

Table 2 shows the EFA results for the Children self-report data from the CheckPoint sample with no factor number constraint. This produced 5 factors with eigenvalue >1 and the scree plot supported 5 factors. Broadly, the first 4 factors were similar to the original structure, although physical item 7 "Get aches and pains" loaded on factor 3 alongside emotional items. The fifth factor consisted of 2 school absence items.

Appendix Tables 2 and 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.005, respectively, show the EFA results after imposing a 4-factor constraint and after removing the 2 school absence items. In the former, the original structure of PedsQL GCS was reproduced, except for physical item 7, which did not load on any factor. In the latter, the physical items were split across 2 factors to create 5 factors. The EFA results for the proxy-reported data in the Mallinson and Toddlers samples are not displayed and are available from the corresponding author. In summary, although the original structure of PedsQL GCS was not completely replicated, this is largely due to the 2 school absence items loading on a fifth factor. Moreover, the 4-factor constraint replicated the original structure.

Confirmatory factor analysis

Table 3 summarizes the CFA results. For each sample, the best-performing structure for each fit statistic is indicated by †. For the CheckPoint sample, which we prioritized, the original structure

Table 1. Characteristics of analytic samples.

Characteristic	CheckPoint sample (N = 1874)	Mallinson sample (N = 7855)	Toddlers sample (N = 7401)
Mean age (SD) [Range]	11.5 (0.5) [11-12]	9.6 (3.6) [6-16]	4.0 (0.0) [4]
Female N (%)	919 (49.0)	3809 (48.5)	3603 (48.7)
SEIFA disadvantage quintile N (%)			
Most disadvantaged	156 (8.3)	804 (21.7)*	1,572 (21.2)
2nd most	277 (14.8)	748 (20.2)	1,480 (20.0)
Middle	351 (18.7)	786 (21.2)	1760 (23.8)
2nd least	439 (23.4)	683 (18.4)	1126 (15.2)
Least disadvantaged	651 (34.7)	685 (18.5)	1463 (19.8)
Remoteness area N (%)			
Major cities	1318 (70.3)	5093 (64.8)	4855 (65.6)
Inner regional	379 (20.2)	1710 (21.8)	1515 (20.5)
Outer regional	162 (8.6)	903 (11.5)	874 (11.8)
Remote	9 (0.5)	110 (1.4)	118 (1.6)
Very remote	6 (0.3)	28 (0.4)	32 (0.4)
Missing	0 (0.0)	11 (0.1)	7 (0.1)
Mean BMI (SD)	19.2 (3.4)	18.7 (4.0)	16.3 (1.7)
Asthma N (%)	252 (13.5)	2248 (28.6)	1519 (20.5)
Diabetes N (%)	7 (0.4)	16 (0.2)	Data not available
Hearing loss N (%)	22 (1.2)	44 (1.0)†	Data not available
Vision loss N (%)	12 (0.6)	168 (3.4)‡	Data not available
Chronic pain N (%)	39 (2.1)	150 (3.0)§	Data not available

ABS indicates Australian Bureau of Statistics; BMI, body mass index; LSAC, Longitudinal Study of Australian Children; SD, standard deviation; SEIFA, socio-economic indexes for areas.
*SEIFA disadvantage data were available only for cohort B wave 4 and cohort K wave 2 children (N = 3706).
†Children reporting moderate/severe hearing problem; these data were available only for cohort B waves 4 and 6 and cohort K wave 5 (N = 4327).
‡Children reporting moderate/severe vision problem; these data were available only for cohort B waves 4 and 6 and cohort K waves 5 and 7 (N = 4973).
§Children reporting moderate/severe pain in the abdomen or other parts of the body; these data were available only for cohort B waves 4 and 6 and cohort K waves 5 and 7 (N = 4978).

Table 2. Rotated factor loadings and unique variances without constraint on factor number for PedsQL GCS Children self-report version, CheckPoint sample.

Original subscale	Item*	Factor 1†	Factor 2	Factor 3	Factor 4	Factor 5
Physical	1. Walk a few houses	**0.646**‡	0.098	0.069	0.177	0.063
	2. Run	**0.754**‡	0.226	0.137	0.030	0.179
	3. Play sport/do exercise	**0.763**‡	0.247	0.129	0.049	0.144
	4. Lifting something heavy	**0.517**‡	0.130	0.220	0.282	0.011
	5. Have a bath or shower	**0.488**‡	-0.084	0.051	0.292	-0.200
	6. Help around the house	**0.438**§,‡	0.090	0.097	**0.496**§	-0.001
	7. Get aches and pains	0.333	0.139	**0.414**‡	0.038	0.327
	8. Have low energy	**0.499**‡	0.180	0.389	0.069	0.216
Emotional	1. Feel afraid/scared	0.194	0.124	**0.716**‡	0.162	0.078
	2. Feel sad/blue	0.117	0.298	**0.692**‡	0.168	0.070
	3. Feel angry	0.068	0.247	**0.643**‡	0.261	0.113
	4. Trouble sleeping	0.151	0.146	**0.585**‡	0.206	0.083
	5. Worry about what will happen	0.117	0.274	**0.665**‡	0.160	0.103
Social	1. Trouble getting along with other kids	0.164	**0.731**‡	0.209	0.219	0.092
	2. Other kids not friends	0.133	**0.821**‡	0.155	0.145	0.091
	3. Other kids tease	0.095	**0.743**‡	0.229	0.144	0.089
	4. Cannot do things other kids can do	**0.407**§	**0.521**‡,§	0.144	0.223	0.112
	5. Keep up when playing	**0.443**§	**0.584**‡,§	0.188	0.137	0.134
School	1. Pay attention	0.072	0.194	0.198	**0.767**‡	0.156
	2. Forget things	0.094	0.223	0.291	**0.598**‡	0.249
	3. Keeping up with school work	0.111	0.256	0.178	**0.748**‡	0.208
	4. Away from school, feel sick	0.129	0.102	0.117	0.227	**0.770**‡
	5. Away from school, to doctor	0.118	0.130	0.087	0.146	**0.800**‡

PedsQL GCS indicates Pediatric Quality of Life Inventory 4.0 Generic Core Scales.
*See Appendix Table 1 in Supplemental Materials found at https://dx.doi.org/10.1016/j.jval.2024.08.005 for the full item wording.
†Factor loadings greater than 0.4 are in bold text.
‡Items in each factor.
§Cross-loading items.

performed best according to root mean square error of approximation and standardized root mean square residual but not according to comparative fit index and Tucker and Lewis index. For all versions and structures, no statistic approached level indicating good fit.

Because the results of the EFA and CFA did not support a robust alternative, the original subscales (physical, emotional, social, and school functioning) were taken forward to psychometric analyses.

Psychometric Results

Overall model psychometric performance

Thirty-six Rasch models were fitted (3 analytic samples, each with 3 subsamples and 4 subscales). Of these, 33 (91.7%) were unidimensional, and 17 (47.2%) showed good overall fit to the Rasch model.

Item-level psychometric performance

Table 4 shows the results of psychometric assessments (Rasch and classical) for the subscales in CheckPoint. For each subsample, items that failed the criterion for disordered threshold, DIF, or poor item-level fit are indicated by ||| and marked by × in the final column. The corresponding results for the Mallinson and Toddlers samples are shown in Appendix Tables 4 and 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.005, respectively.

In the physical subscale for CheckPoint, item 7 was the best-performing item (rejected on Rasch criteria in only 1 subsample). It also had the lowest ceiling effect (20.9%). In the Mallinson sample, all items failed at least 1 Rasch criterion in multiple subsamples. In the Toddlers sample, items 3, 7, and 8 each failed the Rasch criteria in only 1 subsample. Items 3, 7, and 8 were therefore retained for consideration.

In the emotional subscale for CheckPoint, items 2, 4, and 5 failed none of the Rasch criteria in any subsample. Item 2 performed best in terms of the classical criteria and had the widest threshold range. For Mallinson, item 5 failed none of the Rasch criteria in any subsample, whereas items 2 to 4 failed the criteria in one subsample each. For Toddlers, items 1, 3, and 5 failed the criteria in 1 subsample each. All items were retained for consideration, with a particular focus on items 2, 4, and 5, which performed well for CheckPoint.

In the social subscale for CheckPoint, only item 4 failed the Rasch criteria in multiple subsamples. Items 1 and 2 had relatively low ceiling effects and high corrected item-total correlations. For Mallinson, all items except item 3 failed the Rasch criteria in multiple subsamples. For Toddlers, items 2 and 3 failed in none of the subsamples, whereas the other 3 items failed in all subsamples. Items 1, 2, 3, and 5 were retained for consideration.

In the school subscale for CheckPoint, items 4 and 5 were rejected for failing the Rasch criteria in multiple subsamples, whereas items 1, 2, and 3 failed in none. Item 2 had a relatively

Table 3. Results of confirmatory factor analysis by PedsQL GCS version.

Test statistics	Prespecified structures				
	Original 4 subscales	Varni 5 subscales	EFA no constraint	EFA 4-factor constraint	EFA no school absence items
CheckPoint sample: Children version, self-reported data					
RMSEA*	0.103†	0.103	0.103	0.107	0.106
SRMR*	0.250†	0.253	0.254	0.254	0.259
CFI‡	0.723	0.725	0.725†	0.701	0.693
TLI‡	0.695	0.697	0.698†	0.672	0.665
Mallinson sample: Young Children, Children, and Teens versions, proxy-reported data					
RMSEA	0.120	0.116	0.107†	0.113	0.107
SRMR	0.238	0.239	0.231	0.228†	0.231
CFI	0.696	0.720	0.759†	0.732	0.749
TLI	0.666	0.692	0.735†	0.706	0.726
Toddlers sample: Toddlers version, proxy-reported data					
RMSEA	0.118	N/A§	0.099	0.092†	0.097
SRMR	0.189	N/A	0.183	0.167†	0.184
CFI	0.608	N/A	0.722	0.764†	0.715
TLI	0.565	N/A	0.691	0.738†	0.686

CFI indicates comparative fit index; EFA, exploratory factor analysis; N/A, not applicable; PedsQL GCS, Pediatric Quality of Life Inventory 4.0 Generic Core Scales; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker and Lewis index.
*Estimate less than 0.05 indicates a good fit and between 0.05 and 0.08 a reasonable fit.
†The structure that performs best for each statistic.
‡Estimate greater than 0.95 indicates a good fit and between 0.90 and 0.95 a reasonable fit.
§Not applied because removing the 2 school absence items would leave a single item for the school subscale.

low ceiling effect. For Mallinson, all items except item 4 failed the Rasch criteria in multiple subsamples. For Toddlers, only 3 items were available (corresponding to items 3, 4, and 5 in the other age versions). Of these, only item 3 failed the Rasch criteria in multiple subsamples. All items were retained for consideration, with particular focus on items 1, 2, and 3, which performed well for CheckPoint.

Item Selection Results

Physical functioning

Item 7 (Get aches and pains) was the only retained item according to the Rasch criteria in the prioritized CheckPoint sample, had reasonable classical psychometric performance, and was also retained in the Toddlers sample. Item 7 was selected.

Emotional functioning

Items 2 (Feel sad/blue), 4 (Trouble sleeping), and 5 (Worrying about what will happen) were retained based on Rasch performance in CheckPoint and had comparable classical performance. Item 5 was the only item retained according to the Rasch criteria in all 3 analytic samples. However, it was considered to be conceptually different to other emotional functioning items in that it is future oriented and may be less easily understood by younger children. There was also a concern that item 4 on its own does not capture emotional functioning. Item 2 was selected.

Social functioning

Items 1 (Trouble getting along with other kids), 2 (Other kids not friends), 3 (Other kids tease), and 5 (Keeping up when playing) were retained according to the Rasch criteria in CheckPoint and had comparable classical performance. Item 3 was the only item retained according to the Rasch criteria in all 3 analytic samples, whereas item 2 was retained in the CheckPoint and Toddlers samples. Items 2 and 3 were conceptually similar in depicting negative peer relationships. However, it was noted that item 3 (Other kids tease) captures a more extreme end of peer relationships and hence may not represent the intended content of the social functioning subscale in isolation. Item 2 was selected.

School functioning

Items 1 (Pay attention), 2 (Forget things), and 3 (Keeping up with school work) were retained according to the Rasch criteria in CheckPoint and had comparable classical performance. Items 1 and 2 are not available in the Toddlers version and hence were ruled out. Item 3 was selected.

Discussion

This study identified 4 items that represent each of the 4 subscales of PedsQL GCS to form a reduced health-state classification system. This is amenable for valuation using techniques such as discrete choice experiments to derive health utilities and can be applied in future health economic evaluations. We recommend that PedsQL GCS is still administered in full and that the value set is applied to the relevant items afterward. Retrospective derivation of utilities would also be possible based on responses to the above 4 items in previous studies using PedsQL GCS. Given that PedsQL GCS is one of the most widely used

Table 4. Psychometric analysis results for the CheckPoint sample.

Item*,†	Subsample	Disordered threshold?	DIF‡	Item fit residual§	Item χ^2 test significant‖?	Threshold range	Item mean location	Missing data (%)**	Floor effect (%)**,††	Ceiling effect (%)**,‡‡	Corrected item-total correlation**	Outcome (× = reject)§§
Physical functioning												
1. Walk a few houses	1	Yes‖		−0.408		−0.514 to 0.903	0.199	1.5	0.8	78.3	0.530	×‖
	2			−1.000		−1.278 to 0.459						
	3	Yes‖		−1.035		−0.679 to 0.558	−0.202					×‖
	Summary									Rejected in 2/3 subsamples		
2. Run	1			**−3.147**‖	Yes‖	−2.525 to 2.193	0.007	1.6	0.5	60.5	0.649	×‖
	2			**−3.947**‖	Yes‖	−1.285 to 1.739	0.040					×‖
	3			**−3.455**		−1.563 to 1.749	0.052					
	Summary									Rejected in 2/3 subsamples		
3. Play sport/do exercise	1			**−3.562**‖	Yes‖	−3.344 to 1.563	−0.553	1.7	0.4	66.8	0.670	×‖
	2			**−4.264**‖	Yes‖	−2.929 to 1.234	−0.496					×‖
	3	Yes		**−3.816**‖	Yes‖	−1.189 to 1.196	−0.320					×‖
	Summary									Rejected in 3/3 subsamples		
4. Lifting something heavy	1			0.442		−2.133 to 3.114	0.295	1.5	1.0	34.5	0.505	
	2	Yes‖		0.907		−1.391 to 2.583	0.490					×‖
	3	Yes‖		0.609		−0.916 to 2.810	0.605					×‖
	Summary									Rejected in 2/3 subsamples		
5. Have a bath or shower	1	Yes‖		−0.264		−7.340 to 5.033	−1.322	1.5	0.4	94.4	0.349	×‖
	2	Yes‖		−1.030		−5.796 to 6.928	−0.442					×‖
	3	Yes‖		−1.053		−1.622 to 0.946	−0.632					×‖
	Summary									Rejected in 3/3 subsamples		
6. Help around the house	1		Uniform (sex)‖	0.967		−1.305 to 1.468	0.098	1.5	0.6	57.6	0.471	×‖
	2			0.959		−1.119 to 1.729	0.198					
	3		Uniform (age)‖	0.979		−2.092 to 1.545	−0.211					×‖
	Summary									Rejected in 2/3 subsamples		
7. Get aches and pains	1			2.038	Yes	−0.539 to 3.524	1.517	1.6	1.5	20.9	0.422	
	2			1.628		−1.202 to 3.523	1.110					
	3			**2.645**‖	Yes‖	0.144 to 1.314	0.916					×‖
	Summary									Rejected in 1/3 subsamples		
8. Have low energy	1			−0.381		−0.515 to 3.072	0.835	1.5	0.7	41.5	0.538	
	2	Yes‖		−0.744		−1.145 to 2.496	0.296					×‖
	3		Uniform (sex)‖	−1.039		−1.039 to 2.356	0.146					×‖
	Summary									Rejected in 2/3 subsamples		
Emotional functioning												
1. Feel afraid/scared	1	Yes‖		−0.685		−2.084 to 2.608	−0.182	1.7	0.7	30.4	0.607	×‖
	2			−0.605		−4.293 to 3.170	−0.335					
	3	Yes‖		0.577		−1.768 to 2.289	−0.180					×‖
	Summary									Rejected in 2/3 subsamples		
2. Feel sad/blue	1			−0.790		−4.107 to 3.135	−0.462	1.9	0.5	19.3	0.639	
	2			−0.032		−2.700 to 4.106	0.433					
	3			−0.079		−2.727 to 3.294	−0.157					
	Summary									Rejected in 0/3 subsamples		

continued on next page

Table 4. Continued

Item*,‡	Subsample	Disordered threshold?	DIF‡	Item fit residual§	Item χ^2 test significant‖?	Threshold range	Item mean location	Missing data (%)**	Floor effect (%)**,††	Ceiling effect (%)**,‡‡	Corrected item-total correlation**	Outcome (× = reject)§§
3. Feel angry	1		Uniform (sex)¶¶	0.680		−2.711 to 3.403	0.192	2.0	0.5	18.2	0.593	×¶¶
	2		Uniform (age)¶¶	1.210		−8.579 to 3.851	−1.142					×¶¶
	3			0.958		−3.905 to 3.063	−0.347					
	Summary									Rejected in 2/3 subsamples		
4. Trouble sleeping	1			**3.043**		−1.804 to 2.208	0.234	1.8	3.6	30.8	0.494	
	2			2.011		−0.381 to 2.889	1.082					
	3			1.386		−1.208 to 1.983	0.421					
	Summary									Rejected in 0/3 subsamples		
5. Worrying about what will happen	1			−0.298		−1.950 to 1.791	−0.106	2.1	2.5	36.6	0.613	
	2			−0.698		−0.956 to 2.300	0.668					
	3			−0.301		−2.272 to 1.699	−0.183					
	Summary									Rejected in 0/3 subsamples		
Social functioning												
1. Trouble getting along with other kids	1			−0.233		−2.190 to 2.786	−0.117	1.3	0.8	45.2	0.670	
	2			−0.637		−1.972 to 2.620	−0.175					
	3			−0.759		−2.550 to 2.698	−0.210					
	Summary									Rejected in 0/3 subsamples		
2. Other kids not friends	1			−1.546		−2.310 to 2.332	−0.123	1.8	1.1	48.6	0.706	
	2			−0.895		−1.034 to 2.308	0.087					
	3			−1.917		−2.156 to 2.482	−0.146					
	Summary									Rejected in 0/3 subsamples		
3. Other kids tease	1			0.700		−1.152 to 1.819	0.040	1.6	1.8	57.5	0.628	
	2			1.160		−1.048 to 1.619	0.121					
	3			0.783		−0.763 to 1.593	0.208					
	Summary									Rejected in 0/3 subsamples		
4. Cannot do things other kids can do	1	Yes¶¶		2.153		−1.297 to 2.408	0.267	1.7	1.1	46.1	0.581	×¶¶
	2			1.555		−1.383 to 2.443	0.163					
	3	Yes¶¶		1.213		−1.126 to 2.478	0.350					×¶¶
	Summary									Rejected in 2/3 subsamples		
5. Keeping up when playing	1			−0.768		−1.531 to 1.450	−0.426	1.7	0.7	65.1	0.639	
	2	Yes¶¶		−1.562		−2.316 to 1.566	−0.516					×¶¶
	3			−0.091		−1.492 to 1.206	−0.505					
	Summary									Rejected in 1/3 subsamples		
School functioning												
1. Pay attention	1			−1.299		−1.549 to 2.134	−0.058	1.6	1.3	34.9	0.577	
	2			−0.950		−1.814 to 2.290	0.054					
	3***			−1.586		−1.969 to 1.900	−0.251					
	Summary									Rejected in 0/3 subsamples		
2. Forget things	1			−0.287		−2.155 to 3.504	0.526	1.8	2.1	14.9	0.564	
	2			0.322		−1.804 to 3.384	0.585					
	3***			−0.782		−2.005 to 3.254	0.442					
	Summary									Rejected in 0/3 subsamples		

continued on next page

Table 4. Continued

Item*,†	Subsample	Disordered threshold?	DIF‡	Item fit residual§	Item χ^2 test significant‖?	Threshold range	Item mean location	Missing data (%)**	Floor effect (%)**,††	Ceiling effect (%)**,‡‡	Corrected item-total correlation**	Outcome (× = reject)§§
3. Keeping up with school work	1			**−2.926**		−2.196 to 1.900	−0.239	1.9	1.3	37.1	0.631	
	2			−1.816		−2.552 to 1.951	−0.226					
	3***			**−3.288**		−1.643 to 1.833	−0.151					
	Summary									Rejected in 0/3 subsamples		
4. Away from school, feel sick	1	Yes‖‖		1.532		−1.336 to 2.144	−0.209	1.8	1.4	39.1	0.488	×‖‖
	2			0.918		−1.693 to 2.076	−0.363					
	3***	Yes‖‖		1.072		−2.548 to 1.746	−0.077					×‖‖
	Summary									Rejected in 2/3 subsamples		
5. Away from school, to doctor	1	Yes‖‖		0.965		−1.541 to 1.327	−0.422	1.8	1.0	54.0	0.436	×‖‖
	2	Yes‖‖		0.699		−1.030 to 1.063	−0.221					×‖‖
	3***	Yes‖‖		1.153		−1.646 to 1.145	−0.226					×‖‖
	Summary									Rejected in 3/3 subsamples		

DIF indicates differential item functioning.
*See Appendix Table 1 in Supplemental Materials found at https://dx.doi.org/10.1016/j.jval.2024.08.005 for the full item wordings.
†Where items were adjusted for disordered threshold and/or DIF, their Rasch statistics before the adjustment(s) are shown; for other items, their statistics after all adjustments are shown.
‡Based on item-specific χ^2 test with threshold P value of $< .05$ and Bonferroni correction.
§Fit residuals outside the range of −2.5 and 2.5 are highlighted in **bold**.
‖Using threshold P value of $< .01$ and Bonferroni correction.
**Assessed on the full analytic sample rather than the subsample.
††Percentage responding "Almost always," (ie, the lowest health-related quality of life, excluding missing responses).
‡‡Percentage responding "Never," (ie, the highest health-related quality of life, excluding missing responses).
§§Rejection for disordered threshold, DIF, or poor item fit (fit residual outside range of ±2.5 and significant statistic for item χ^2 test).
‖‖The criteria that led to rejection.
***Subsample failed to show unidimensionality.

measures of childhood HRQoL,[55] such applications should significantly expand the evidence base for cost-effectiveness-based decision making for childhood populations.

The proposed health-state classification system needs to undergo further psychometric validation before use in clinical research and practice. This should cover multiple criteria, including reliability, construct validity, responsiveness, and empirical validity (once a value set is developed).[76,77] These validations should be comparative, ie, assess the performance of the system relative to existing childhood MAUIs (eg, CHU9D).[43,78] The psychometric performance of the derived utilities could likewise be compared with the total score of the full version of PedsQL GCS, which has already shown acceptable psychometric performance in diverse settings.[46-53] These steps would confirm that the new system fills a methodological gap, namely, a health utility instrument that can be applied across a wide range of childhood age groups and has strong psychometric properties.

Several caveats apply to this study's methodological approach. First, there was some difficulty in empirically identifying the structure of PedsQL GCS for psychometric evaluation. None of the structures examined in CFA, including the original structure and those suggested by EFA, produced statistics indicating good fit. The EFA results themselves were not wholly consistent: for example, physical item 7 (Get aches and pain) in the CheckPoint sample loaded with emotional items in unconstrained EFA (Table 2) but not so under different EFA conditions (Appendix Tables 2 and 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.005). Overall, there was no evidence of a robust alternative to the original structure.

Second, the study faced data constraints. The accessed data exclusively concerned Australian children, and further work remains to verify generalizability to other populations. Assessment of cross-cultural validity is required concerning both the item selection and the psychometric performance of the selected items. This process could involve reassessing the content validity of the selected items such as "Feel sad/blue," in which the word "blue" is no longer a commonly used expression. Moreover, the self-reported PedsQL GCS data in CheckPoint were only available from children aged 11 to 12 years, and we cannot say whether the same items would be selected from other childhood age groups. Likewise, data for the Toddlers sample were available only from 4 year olds, which potentially restricts the generalizability of findings to the age range of 2 to 4 years targeted by the Toddlers version.

Third, the study utilized proxy-reported PedsQL GCS data in the Mallinson and Toddlers samples and self-reported data in CheckPoint but was unable to assess the correspondence between self and proxy reports without the data from both reports for the same children obtained at the same time. Disagreements between child and proxy reports have been documented,[22,23] and further work is required to verify whether respondent type alters the item selection. That said, the consideration of proxy-reported data from LSAC enhances the applicability of the proposed system for self- and proxy reports and across childhood age groups. Further research should explore the potential variation in the psychometric properties of the PedsQL GCS items across the different age versions, which in this study were grouped into the Mallinson and Toddlers samples.

This study nevertheless has several strengths. First, it closely followed recommendations[11,14] for the development of health-state classification systems amenable to health utility derivation. Moreover, the item selection process was transparent in how it considered the psychometric performance and conceptual validity of items. Second, there were nontrivial variations in Rasch

performance between subsamples and analyzing multiple subsamples improved the robustness of item selection. Third, all analyses were conducted on data obtained from general population samples, which included both healthy children and those with ongoing clinical conditions. This has the advantage of ensuring that the results are truly generic and applicable across diverse pediatric populations. Although further work may be necessary to investigate applicability for health technology assessment decision making in tightly defined clinical groups, the generic system is highly relevant for wider health technology assessment decision making.[56]

Conclusion

A reduced health-state classification system was generated from the PedsQL GCS. The item set contains 4 items, 1 from each of the original subscales of the PedsQL GCS, which, after further psychometric validation, should be amenable to preference-based valuation. This will enable the prospective and retrospective derivation of preference-based health utilities from PedsQL GCS applications, which, in turn, should support cost-utility analyses and inform decision making regarding interventions for children.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Dr Petrou is an editor for *Value in Health* and had no role in the peer-review process of this article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.08.005.

Article and Author Information

Accepted for Publication: August 27, 2024

Published Online: October 27, 2024

doi: https://doi.org/10.1016/j.jval.2024.08.005

Author Affiliations: Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, England, UK (Kwon, Petrou); Menzies Centre for Health Policy and Economics, University of Sydney, Sydney, NSW, Australia (Raghunandan, Howard, Howell); Department of Health Economics, Wellbeing and Society, Australian National University, Canberra, NSW, Australia (Nghiem, Lancsar, Huynh); Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London. England, UK (Smith).

Correspondence: Stavros Petrou, PhD, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, England, United Kingdom. Email: stavros.petrou@phc.ox.ac.uk

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This research has been funded by the Australian Government's Medical Research Future Fund under Grants MRF1200816 and MRF1199902. S.P. receives support as a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0616-10103) and from the NIHR Applied Research Collaboration Oxford and Thames Valley. The views expressed are those of the authors and not necessarily those of the Australian Government, the NIHR or the Department of Health and Social Care in England.

Role of the Funders/Sponsors: Funders were not involved in the conceptualization and conduct of this study.

Acknowledgment: The authors would like to acknowledge the contribution of Dr Marcus Sellars during the conceptualization stage of the study.

Data Availability: The Excel files which were inputted to RUMM2030 are available from the corresponding author upon reasonable request.

Ethical Approval: This project has ethics approval from the University of Sydney Human Research Ethics Committee (project no. 2023/042).

REFERENCES

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, England: Oxford University Press; 2015.
2. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. London, England: National Institute for Health and Care Excellence (NICE); 2013.
3. Canadian Agency for Drugs and Technologies in Health. *Guidelines for the Economic Evaluation of Health Technologies: Canada*. 4th ed. Ottawa, Canada; 2017.
4. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee, version 5. Australian Government Department of Health. https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf; Published 2016. Accessed October 15, 2022.
5. Healthcare Improvement Scotland, Scottish Medicines Consortium. Working with SMC - a Guide for Manufacturers. https://scottishmedicines.org.uk/media/8208/working-with-smc.pdf; Published 2017. Accessed October 15, 2022.
6. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
7. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford, England: Oxford University Press; 2017.
8. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ*. 1986;5(1):1–30.
9. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *Pharmacoeconomics*. 2019;37(1):29–43.
10. Krucien N, Watson V, Ryan M. Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Econ*. 2017;26(12):e1–e16.
11. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Qual Life Res*. 2009;18(2):253–265.
12. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. *Med Decis Mak*. 2008;28(1):113–126.
13. Brazier J, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess (Winchester, England)*. 2012;16(32):1–114.
14. Brazier JE, Mulhern BJ, Bjorner JB, et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care*. 2020;58(6):557–565.
15. Costa DS, Aaronson NK, Fayers PM, et al. Deriving a preference-based utility measure for cancer patients from the European Organisation for the Research and Treatment of Cancer's Quality of Life Questionnaire C30: a confirmatory versus exploratory approach. *Patient Relat Outcome Meas*. 2014;5:119.
16. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a preference-based measure for myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846–855.
17. Rentz AM, Kowalski JW, Walt JG, et al. Development of a preference-based index from the National Eye Institute visual function questionnaire-25. *JAMA Ophthalmol*. 2014;132(3):310–318.
18. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health*. 2011;14(5):721–731.
19. Young TA, Yang Y, Brazier JE, Tsuchiya A. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Med Decis Mak*. 2011;31(1):195–210.
20. Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health*. 2013;16(4):461–479.
21. Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *Pharmacoeconomics*. 2011;29(8):641–652.
22. Khadka J, Kwon J, Petrou S, Lancsar E, Ratcliffe J. Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: a systematic review. *Soc Sci Med*. 2019;240:112543.
23. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res*. 2001;10(4):347–357.

24. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ*. 2003;12(8):697–702.
25. Verstraete J, Ramma L, Jelsma J. Item generation for a proxy health related quality of life measure in very young children. *Health Qual Life Outcomes*. 2020;18(1):1–15.
26. Kwon J, Freijser L, Huynh E, 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.
27. Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). *Qual Life Res*. 1996;5(2):205–211.
28. Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res*. 1996;5(6):532–538.
29. Beusterien KM, Yeung J-E, Pang F, Brazier J. Development of the multi-attribute adolescent health utility measure (AHUM). *Health Qual Life Outcomes*. 2012;10(1):1–9.
30. Moodie M, Richardson J, Rankin B, Iezzi A, Sinha K. Predicting time trade-off health state valuations of adolescents in four Pacific countries using the Assessment of Quality-of-Life (AQoL-6D) instrument. *Value Health*. 2010;13(8):1014–1027.
31. Kang E. Validity of child health-6 dimension (CH-6D) for adolescents. *Value in Health*. 2016;19(7):A854.
32. Saigal S, Rosenbaum P, Stoskopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. *Qual Life Res*. 2005;14(1):243–252.
33. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res*. 2009;18(8):1105–1113.
34. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. *Appl Health Econ Health Policy*. 2011;9(3):157–169.
35. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*. 2010;19(6):875–886.
36. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res*. 2019;28(7):1951–1961.
37. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multi-attribute utility function for a comprehensive health status classification system: health Utilities Index Mark 2. *Med Care*. 1996:702–722.
38. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. *Ann Med*. 2001;33(5):375–384.
39. Jabrayilov R, van Asselt AD, Vermeulen KM, et al. A descriptive system for the Infant health-related quality of life Instrument (IQI): measuring health with a mobile app. *PLoS One*. 2018;13(8):e0203276.
40. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res*. 1976;11(4):478–507.
41. Kaplan RM, Sieber WJ, Ganiats TG. The quality of well-being scale: comparison of the interviewer-administered version with a self-administered questionnaire. *Psychol Health*. 1997;12(6):783–791.
42. Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. *J Patient Rep Outcomes*. 2020;4(1):1–14.
43. Kwon J, Smith S, Raghunandan R, et al. Systematic review of the psychometric performance of generic childhood multi-attribute utility instruments. *Appl Health Econ Health Policy*. 2023:1–26.
44. 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):800–812.
45. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. *Med Care*. 1999;37(2):126–139.
46. 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):705–719.
47. Amedro P, Huguet H, Macioce V, et al. Psychometric validation of the French self and proxy versions of the PedsQL 4.0 generic health-related quality of life questionnaire for 8-12 year-old children. *Health Qual Life Outcomes*. 2021;19(1):1–14.
48. Limbers CA, Larson M. A systematic review of psychometric properties of the Pediatric Quality of Life Inventory 4.0 generic core scales: in pediatric cancer patients and survivors. *Expert Rev Qual Life Cancer Care*. 2016;1(2):145–152.
49. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: an analysis of 8591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;5(1):1–13.
50. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;5(1):1–10.
51. Hartman EE, Pawaskar M, Williams V, et al. Psychometric properties of PedsQL generic core scales for children with functional constipation in the Netherlands. *J Pediatr Gastroenterol Nutr*. 2014;59(6):739–747.
52. Hao Y, Tian Q, Lu Y, Chai Y, Rao S. Psychometric properties of the Chinese version of the Pediatric Quality of Life Inventory™ 4.0 generic core scales. *Qual Life Res*. 2010;19(8):1229–1233.
53. Petersen S, Hägglöf B, Stenlund H, Bergström E. Psychometric properties of the Swedish PedsQL, Pediatric Quality of Life Inventory 4.0 generic core scales. *Acta Paediatr*. 2009;98(9):1504–1512.
54. Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. *Arch Dis Child*. 2001;84(3):205–211.
55. Varni JW. The PedsQL measurement model for the pediatric quality of life Inventory. PedsQL. https://www.pedsql.org/index.html; Published 2023. Accessed November 1, 2023.
56. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. Patterns, trends and methodological associations in the measurement and valuation of childhood health utilities. *Qual Life Res*. 2019;28(7):1705–1724.
57. Raghunandan R, Howard K, Smith S, et al. Psychometric evaluation of the PedsQL GCS and CHU9D in Australian children and adolescents with common chronic health conditions. *Appl Health Econ Health Policy*. 2023;21(6):949–965.
58. Australian Institute of Family Studies. Accessing LSAC data. https://growingupinaustralia.gov.au/data-and-documentation/accessing-lsac-data; Published 2022. Accessed November 10, 2022.
59. Sanson A, Johnstone R, The LSAC Research Consortium, FaCS LSAC Project Team. Growing up in Australia takes its first steps. *Family Matters*. 2004;67:46–53.
60. Edwards B. Growing Up in Australia: the Longitudinal Study of Australian Children: entering adolescence and becoming a young adult. *Family Matters*. 2014;95:5–14.
61. Growing up in Australia: the longitudinal study of Australian children (LSAC). Department of Social Services. https://www.dss.gov.au/about-the-department/longitudinal-studies/growing-up-in-australia-lsac-longitudinal-study-of-australian-children-overview; Published 2022. Accessed November 10, 2022.
62. Wake M, Clifford S, York E, et al. Introducing growing up in Australia's child health check point: A physical health and biomarkers module for the longitudinal study of Australian children. *Family Matters*. 2014;95:15–23.
63. Clifford SA, Davies S, Wake M, Child Health CheckPoint Team. Child Health CheckPoint: cohort summary and methodology of a physical health and biospecimen module for the Longitudinal Study of Australian Children. *BMJ Open*. 2019;9(suppl 3):3–22.
64. Mallinson T. Rasch analysis of repeated measures. *Rasch Meas Trans*. 2011;25(1):1315–1320.
65. Rummlab Pty Ltd. Rumm 2030+ Rasch Measurement Tools for Research and Education; 1975-2024 Copyright. https://www.rummlab.com.au/. Accessed January 5, 2024.
66. StataCorp. *Stata Statistical Software. Release 17*. College Station, TX: StataCorp LLC; 2021.
67. Varni JW, Limbers CA, Newman DA. Using factor analysis to confirm the validity of children's self-reported health-related quality of life across different modes of administration. *Clin Trials*. 2009;6(2):185–195.
68. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago; 1960.
69. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43(4):561–573.
70. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res*. 2007;57(8):1358–1362.
71. Smith Jr EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3(2):205–231.
72. Smith S, Lamping D, Banerjee S, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess (Winchester, England)*. 2005;9(10):1–93. iii-iv.
73. Lohr KN, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193–205.
74. Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (Rumm) program in health outcome measurement. *J Appl Meas*. 2016;17(4):416–431.
75. Hendriks A, Smith SC, Chrysanthaki T, Cano SJ, Black N. DEMQOL and DEMQOL-Proxy: a Rasch analysis. *Health Qual Life Outcomes*. 2017;15(1):164.
76. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ*. 1999;8(1):41–51.
77. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF−6D based on hypothetical preferences in a general population. *Health Econ*. 2005;14(11):1169–1189.
78. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A review of the psychometric performance of selected child and adolescent preference-based measures used to produce utilities for child and adolescent health. *Value Health*. 2020;24(3):443–460.