ORIGINAL RESEARCH ARTICLE

A Comparison of the Psychometric Properties of the EQ-5D-Y-3L and EQ-5D-Y-5L Using Paediatric Multi-Instrument Comparison (P-MIC) Study Data

Mina Bahrampour[1] · Nancy Devlin[2] · Renee Jones[2] · Kim Dalziel[2] · Brendan Mulhern[1] On behalf of the QUOKKA (Quality of Life in Kids: Key Evidence for Decision Makers in Australia) Team

Abstract

Background The EQ-5D-Y is a generic health-related quality of life (HRQoL) measure suitable for children and adolescents. There are two versions of the instrument, which differ in response levels: the three- (Y-3L) and five-level (Y-5L) versions. The Y-5L has been developed to improve the psychometric performance of the Y-3L.

Objective This study aims to assess the psychometric properties of Y-5L and to compare the performance of Y-5L with Y-3L in separate self- and proxy-reported samples.

Methods Both versions of the instrument were administered to children/adolescents (self-report) or caregivers (proxy-report) of Australian children aged 5–18 years. The data were gathered as part of the Australian Paediatric Multi-Instrument Comparison (P-MIC) study. The Y-5L and Y-3L were evaluated in terms of ceiling effects, criterion validity, inconsistency, informativity, and test-retest reliability in both proxy and self-complete populations.

Results Overall, 5945 respondents completed both the Y-3L and Y-5L, with 2083 proxy and 3862 self-completed responses. Ceiling effects were lower in the Y-5L than the Y-3L. Items from the same domains show the strongest correlation for both samples. Regarding informativity, the Y-5L demonstrated more discriminatory power, indicated by having a higher Shannon diversity index in all domains of the Y-5L compared with the Y-3L. Test-retest showed fair agreement between responses at baseline and follow-up.

Conclusion The Y-5L has demonstrated moderate validity and fair reliability for use in measuring HRQoL in children and adolescents, both when self-reported by children or proxy reported. Overall, Y-5L is a dependable and valid extension from the Y-3L.

✉ Mina Bahrampour
  mina.bahrampour@uts.edu.au

  Nancy Devlin
  nancy.devlin@unimelb.edu.au

  Renee Jones
  reneej1@student.unimelb.edu.au

  Kim Dalziel
  kim.dalziel@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

Key Points for Decision Makers

Compared with the EQ-5D-Y-3L, the EQ-5D-Y-5L provides more information about the severity of children and adolescents' health-related quality of life.

The EQ-5D-Y-5L shows moderate validity and fair reliability and can be included in economic evaluation and clinical and quality-of-care studies.

Although the self and proxy data cannot be directly compared, there is evidence that the characteristic of the relationship differs between the instruments across the samples.

1 Introduction

The EQ-5D is one of the most widely used instruments to measure health-related quality of life (HRQoL) in adults [1]. The EQ-5D-Y was developed to measure HRQoL in children and young people aged 4–15 years, and aims to facilitate the economic evaluation of healthcare interventions to inform decision making for children and young people [2]. A self-report version is available for children aged 8–15 years, and a proxy-report version was developed for parents or caregivers of children aged as young as 4 years, which can also be used for older children who are unable to complete the self-report. The EQ-5D-Y was developed based on the adult EQ-5D and defines items for the domains of mobility, 'walking around'; selfcare, 'looking after myself'; usual activities, 'doing usual activities'; pain or discomfort, 'having pain or discomfort'; and anxiety or depression, 'feeling worried, sad or unhappy' [3, 4].

Initially, the EQ-5D-Y was developed with three response levels for each domain (EQ-5D-Y-3L, hereon Y-3L). Prior research has shown the Y-3L to be a reliable instrument [5], however it has been criticised for the extent to which it reflects variability in individual responses due to having only three levels [6] and having high ceiling effects [5, 7] (where a high portion of respondents report 'no problems' to all instrument domains or items). To overcome the limitations and improve the psychometric performance of the Y-3L, a version with five response levels (EQ-5D-Y-5L, hereon Y-5L) has been developed [8]. The feasibility of the Y-5L has been confirmed in previous studies, however these studies have suggested that further testing of the psychometric properties of the Y-5L is required on different samples and conditions [9–12].

The psychometric features of the adult versions (which also has a three- and a five-level version) have been compared previously, and most studies supported the advantages of the EQ-5D-5L descriptive system. By using the 5L version, the ceiling effect decreased and the discriminatory ability and informativity have increased [13, 14]. It is unknown to what extent the advantages of the adult 5L over the adult 3L will also be apparent in the Y-5L compared with the Y-3L. Understanding the differences in psychometric properties between Y-3L and Y-5L can be beneficial in informing valid HRQoL instrument choice and for informing clinical and health policy decision making.

In paediatric research, it is common to use proxy (often caregivers) responses on behalf of the child, especially when the child is younger or does not have the ability to respond. However, studies have demonstrated inconsistencies between child self-reports and proxy-reports [15, 16]. The agreement between the two groups tends to be closely related to physical domains, such as mobility, with divergence in responses when comparing mental domains such as anxiety [17]. Proxies often rate functioning worse than children rate for themselves [18]. Understanding the psychometric properties of the Y-5L compared with the Y-3L in both these groups is therefore important.

This study aimed to compare the distributional and psychometric properties of the Y-3L and Y-5L instruments in separate self- and proxy-report populations using data from the Australian Paediatric Multi-Instrument Comparison (P-MIC) study. The study aimed to assess the following psychometric properties: ceiling effects, construct validity, reliability and informativity. We also explored how the data from each instrument is distributed relative to the other in the P-MIC sample.

2 Methods

2.1 Data

The data for this research were collected as a part of the P-MIC study [19]. The P-MIC study gathered demographic information simultaneously alongside multiple generic paediatric HRQoL instruments from Australian children (as self-report) or caregivers' providing their perception of the child's health (proxy-report). Caregivers of children aged 7–18 years were asked if their child was able to self-report their HRQoL. If the caregiver said yes, only the child was asked to self-complete the HRQoL instruments; if the caregiver said no or the child was younger than 7 years of age, the caregiver was asked to proxy-report the HRQoL instruments. None of the instruments were interviewer-administered [20]. Data included in the analysis are from P-MIC data-cut 2, dated 10 August 2022, and include 94% of the total planned P-MIC participants [21]. The data are of children aged 5–18 years [20] from different health conditions and general population samples recruited via a hospital or online survey panel. The wider study measured HRQoL for children aged 2–4 years using an experimental version of the Y-5L instrument, and this is reported elsewhere [22]. The health condition groups are recurrent abdominal pain, attention-deficit hyperactivity disorder (ADHD), anxiety or depression, autism spectrum disorders (ASD), asthma, dental problems, eating disorder, epilepsy, and sleeping problems. A detailed explanation of the P-MIC study survey

design, data collection samples and data-cuts have been reported by Jones et al. [21].

Data were gathered using online surveys via REDCap. Participants completed both the Y-3L and Y-5L instruments in random order and also completed the EuroQol visual analogue scale (EQ-VAS), which was always presented with the Y-3L. The design of the survey data collection form required a response to each question and hence did not permit missing data. A random subset of participants in the general population sample received the follow-up survey at 2 days to allow for reliability assessment.

2.2 EQ-5D-Y-3L (Y-3L) and EQ-5D-Y-5L (Y-5L) and EuroQol Visual Analogue Scale (EQ-VAS)

Children and adolescents' HRQoL was measured using the Y-3L and Y-5L descriptive systems and were asked to report the child's health status 'today'. The Y-3L has three response levels (no, some, and a lot/very), which result in 243 unique health states. The Y-5L has five response levels (no, a little bit of, some, a lot of, cannot/extreme), which lead to 3125 possible health states. The Y-5L not only increases the number of levels but also changes the level labels, meaning that level 3 on the Y-3L is descriptively equivalent to level 4 on the Y-5L in four domains, reflecting an additional layer of severity in the Y-5L.

The health states are often described as five-digit vectors by taking one level for each domain, with 11111 representing full health and 33333 and 55555 representing the worst possible health state for the Y-3L and Y-5L respectively.

The EQ-VAS comprises a scale from 0 to 100. Respondents are asked to report their health status 'today', where zero is labelled as the worst imaginable health and 100 indicates the best imaginable health.

2.3 Statistical Analysis

2.3.1 *Ceiling Effects*

An instrument was considered to have a ceiling effect when more than 15% of participants rated their health status at the best level of each domain (i.e. 11111) [23, 24]. In this study, the absolute change in ceiling effect was examined as the difference between the proportions of patients with ceiling effects across the two instruments. The relative reduction was also calculated using Eq. 1:

$$\text{(ceiling Y-3L} - \text{ceiling Y-5L)}/\text{ceiling Y-3L} * 100 \qquad (1)$$

Absolute and relative changes in ceiling effect were calculated in the sample of children with conditions, the general

population sample and an overall combined sample. Results were separated for proxy- and self-report.

Due to the provision of more response level options aside from level 1, the ceiling effect was expected to be lower for Y-5L, as was the case for adult instruments [25, 26].

2.3.2 Inconsistency and Redistribution

Redistribution properties and the level of response consistency were assessed using the criteria applied by Janssen et al. [27]. An inconsistent response in the current study is described as a Y-3L response that is at least two levels away from the Y-5L response, for instance if the respondent chose level one on Y-3L but chose level four on the Y-5L instrument. The size of the inconsistency was measured as |Y-3L − Y-5L|−1. The redistribution properties of the consistent response pairs were described as proportions of the Y-3L–Y-5L response pairs within each Y-3L response level (Y-3L-1, Y-3L-2, and Y-3L-3). Sankey diagrams [28] were used to show the cross-tabulations for each level and domain of the Y-3L with the corresponding level and domain for the Y-5L.

2.3.3 Criterion Validity

Criterion validity indicates how accurately the new measure (Y-5L) assesses the same content as the previously employed measure (Y-3L) and was established by comparing the instrument domain scores using Spearman correlations. Spearman's rank correlation coefficient for skewed data between the Y-3L and Y-5L domains was used. Coefficients from 0.1 to 0.29 were classified as low, 0.3 to 0.49 were classified as moderate, and correlations of ≥ 0.5 were classified as high [29]. The EQ-5D-Y instruments have the same domains, therefore it is hypothesised that these domains will have a high correlation.

The assessment of validity did not include convergence and divergence, which are presented in a separate papers [30, 31].

2.3.4 Discriminatory and Informativity Power

Shannon index (H′) and Shannon evenness index (J′) were applied to assess the informativity and discriminatory power of the instruments. This index shows the distribution of response in each domain. The Shannon index (H′) is used to estimate the discriminatory power for each domain in the Y-3L and Y-5L classification systems. The formula for the Shannon index is shown in Eq. 2:

$$H' = -\sum_{i=1}^{L} p_i log_2 p_i \qquad (2)$$

where H' represents the absolute amount of informativity captured, L is the number of possible levels, and $p_i = n_i/N$, where n_i is the observed number of responses in the ith level (i=1,……,L) and N is the total sample size [32]. The higher the Shannon index, the more information is obtained by the classification system. In the case of an equal (rectangular) distribution, meaning that all levels have the same number of responses, the optimal amount of information is captured and H' has reached its highest limit ($H'max$), which equals to $log_2 L$; therefore, if the number of levels (L) increases, the $H'max$ will rise accordingly. The Shannon evenness index (J') is described as $J' = H'/H'max$ and indicates the evenness of the distribution regardless of the number of levels. The Y-5L is expected to result in more information, thus it is hypothesised to have a higher Shannon index.

2.3.5 Test–Retest Reliability

P-MIC data included a question that asked participants about their health at baseline and at follow-up, with participants selecting between 'the same', 'worse' or 'better'. Among the participants who completed the baseline survey, a subset of 169 were allocated to the 2-day follow-up period and responded. From this group, 115 participants (68%) who reported 'the same' health in response to the question above were included in the test-retest reliability analysis.

The test-retest reliability of the instruments was analysed by comparing the initial survey responses and the 2-day follow-up survey responses using EQ-VAS categories. EQ-VAS scores were used as there are currently no value sets available for the EQ-5D-Y instruments in Australia, and EQ-5D-Y test-retest reliability using level sum scores are presented in a separate paper (28). As EQ-VAS is scaled from 0 to 100, even slight health changes (for example response 70 vs. 72) might have a large impact on test-retest reliability. To circumvent this, the EQ-VAS scores were grouped into 10 categories, and the agreement of these 10 groups were then compared at baseline and follow-up (2 days). The reliability of the Y-3L and Y-5L was also assessed by estimating the weighted kappa coefficients of domains at the 2-day follow-up to estimate concordance. Kappa shows the extent of agreement between two sets of data collected at two different time points [33]. Interpretation of agreement using kappa coefficients was prespecified as follows: kappa < 0.2 indicates poor agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and kappa > 0.81 indicates almost perfect agreement [34]. Test-retest analysis on a domain basis has been published elsewhere [30].

3 Results

3.1 Sample and Demographics

In total, 5945 respondents completed both the Y-3L and Y-5L, with 2083 surveys completed by the proxies of children aged 5–18 years (of which 979 were 5 or 6 years of age); and 3862 self-completed by children aged between 7 and 18 years. In the total sample, 25.7% of the children were part of the general population sample, while the remainder were part of condition groups or were hospital recruits. The mean age of the child in the proxy group was 9 years (± 2.4) and the mean age for the self-report group was 11.9 years (± 3.7). The sample was 44.9% female and 53.8% male in the proxy group, and 46.7% female and 51.7% male in the self-report group. The remainder of the participants were from other gender groups or preferred not to answer the question. Among the proxies, 95.1% were children's parents, 2% were grandparents, 0.4% were their sibling and 2.5% were other types of carers. The proportion of stated problems on the Y-3L and Y-5L are reported in Table 1. More details regarding respondent characteristics, including by age and condition, can be found in the papers by Jones et al. [20, 30, 35].

3.2 Psychometric Results

3.2.1 Ceiling Effect

Table 2 indicates the percentages of the ceiling effect and its change in absolute and relative terms for the Y-5L compared with the Y-3L in the group of children with conditions, the general population group, and the combined overall sample. The proportion of respondents who reported no problems was lower in Y-5L samples in comparison with Y-3L. The reduction of ceiling effect in relative terms was 14.02% in the proxy group and 16.65% in the self-report group. Both versions, Y-3L and Y-5L , of the questionnaire present a similar absolute ceiling effect reduction (Additional file 1: Table S1 shows the proportion of responses at level 1 for each domain).

Both instruments demonstrated ceiling effect issues, especially for the asthma and dental problem condition groups for both proxy and self-reports. The absolute change in ceiling effect between instruments was highest in 'asthma' for both groups (Table 2). Epilepsy had a negative change in proxy reports; however, the change is only 1% and it is therefore not a pattern that can be interpreted with any confidence.

Table 1 Description of responses for proxy-reports [n = 2083] and self-reports [n = 3862]

	Proxy-reports		Self-reports	
	Y-3L—n (%)	Y-5L—n (%)	Y-3L—n (%)	Y-5L [n (%)]
Mobility (walk around)				
L1	1885 (90.49)	1850 (88.81)	3446 (89.23)	3344 (86.59)
L2	153 (7.35)	123 (5.9)	353 (9.14)	329 (8.52)
L3	45 (2.16)	65 (3.12)	63 (1.63)	123 (3.18)
L4		20 (0.96)		57 (1.48)
L5		25 (1.2)		9 (0.23)
Looking after myself				
L1	1491 (71.58)	1398 (67.11)	3262 (84.46)	3178 (82.29)
L2	447 (21.46)	350 (16.8)	515 (13.34)	425 (11)
L3	145 (6.96)	175 (8.4)	85 (2.20)	164 (4.25)
L4		83 (3.98)		81 (2.10)
L5		77 (3.7)		14 (0.36)
Usual activities				
L1	1488 (71.44)	1414 (67.88)	2939 (76.1)	2780 (71.98)
L2	451 (21.65)	341 (16.37)	796 (20.61)	615 (15.92)
L3	144 (6.91)	177 (8.50)	127 (3.29)	309 (8.00)
L4		101 (4.85)		117 (3.03)
L5		50 (2.40)		41 (1.06)
Pain/discomfort				
L1	1508 (72.4)	1435 (68.89)	2367 (61.29)	2163 (56.01)
L2	548 (26.31)	446 (21.41)	1360 (35.21)	1129 (29.23)
L3	27 (1.30)	147 (7.06)	135 (3.50)	413 (10.69)
L4		41 (1.97)		133 (3.44)
L5		14 (0.67)		24 (0.62)
Feeling worried, sad or unhappy				
L1	1099 (52.76)	992 (47.62)	1835 (47.51)	1598 (41.38)
L2	874 (41.96)	810 (38.89)	1768 (45.78)	1585 (41.04)
L3	110 (5.28)	180 (8.64)	259 (6.71)	391 (10.12)
L4		61 (2.93)		186 (4.82)
L5		40 (1.92)		102 (2.64)

For Y-3L: L1, no; L2, some; L3, a lot of

For Y-5L: L1, no; L2, a little bit of; L3, some; L4, a lot; L5, cannot or extreme

For the Y-5L 'worried sad or unhappy domain': L3, quite; L4, really

Y-3L EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

Table 2 Changes in ceiling effect in different condition groups

Condition	Proxy-reports (%)				Self-reports (%)			
	Ceiling Y-3L	Ceiling Y-5L	Absolute change	Relative change	Ceiling Y-3L	Ceiling Y-5L	Absolute change	Relative change
General population	63.25	57.46	7.94	13.2	60.10	52.16	5.79	9.2
Recurrent abdominal pain	21.77	12.90	8.87	40.74	10.82	8.58	2.24	20.7
ADHD	22.34	18.09	4.25	19.02	22.70	19.08	3.62	15.95
Anxiety or depression	17.89	16.84	1.05	5.87	12.47	9.09	3.38	27.11
ASD	12.57	8.74	3.83	30.47	14.98	11.01	3.97	26.5
Asthma	52.43	42.7	9.73	18.56	50.0	41.06	8.94	17.88
Dental	51.45	44.51	6.94	13.49	43.85	36.91	6.94	15.83
Eating disorder	9.76	7.32	2.44	25.0	8.97	6.21	2.76	30.77
Epilepsy	17.71	18.75	-1.04	-5.87	26.70	24.43	2.27	8.5
Sleep	25.0	22.44	2.56	10.24	20.0	13.16	6.84	34.2
Total sample	36.29	31.20	5.09	14.02	33.58	25.02	5.56	16.56

ADHD attention-deficit hyperactivity disorder, *ASD* autism spectrum disorder, *Y-3L* EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

3.2.2 Inconsistency and Redistribution

Cross-tabulations of responses to the Y-3L and Y-5L showed that participants reported health across the EQ-5D-Y domain levels. Table 3a and b show the proportion of consistent and inconsistent responses. The highest inconsistency was related to 'feeling worried, sad or unhappy' and the lowest was related to 'mobility' in both the self-report and proxy groups. Consistent levels are bolded in Table 3a and b. The inconsistent level ranged from 3.42 to 11.30% for the self-report, and between 2.40 and 10.99% for the proxy-report data. A related table for the proportion of consistent levels can be found in Additional file 1: Table S2. Most of the inconsistent responses were related to participants choosing level 2 on Y-3L and level 1 on Y-5L.

Table 4a and b present responses for Y-3L and Y-5L domains for the condition groups. In general, the Y-5L responses demonstrate a redistribution across most domains and condition groups, underscoring the advantages offered by the additional levels in the Y-5L measure. Mobility has the highest percentage of 'no problem' in both groups.

Results show preliminary evidence supporting the validity of the instrument across various conditions, as distributions align with expectations based on the type of condition. For example, there has been a reduction in the ceiling effect in the 'feeling worried, sad or unhappy' domain for participants reporting mental health conditions such as ADHD, anxiety, depression, ASD, and eating disorders. For children with mental health conditions, higher severity levels were reported for the 'feeling worried, sad or unhappy' domain, leading to a broader distribution of responses.

Figures 1 and 2 demonstrate how the Y-3L levels are distributed on the Y-5L levels. The results indicate that 'feeling worried, sad or unhappy' is more distributed among the response levels, compared with other domains, with more responses in the severe levels. The response pattern is relatively consistent across both groups; for instance, 'mobility' has the highest number of respondents in the 'no problem' response level.

3.2.3 *Criterion Validity*

Table 5 shows the results of Spearman correlations between the Y-3L and Y-5L. In both the proxy and self-report data, the highest correlation is related to the same domain, consistent with expectations. The correlation of the same domains for Y-5L and Y-3L was high (> 0.5). The highest correlation was between the domain 'looking after myself' in both groups (proxy-report = 0.84 and self-report = 0.75), and the lowest correlation was between the domain 'feeling worried, sad or unhappy' in both groups (proxy-report = 0.64 and self-report = 0.65).

3.2.4 Informativity

The results of the analysis of informativity of the Y-3L and Y-5L are presented in Table 6. The Shannon index (H′), showed a gain in all domains and indicates better

Table 3 Redistribution properties from Y-3L to Y-5L

(a) Proxy-report inconsistency (n = 2083)

Mobility (walk around)

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	1819	54	11	1	0
L2	28	67	50	7	1
L3	3	2	4	12	24

Looking after myself

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	1357	112	16	5	1
L2	41	232	143	29	2
L3	0	6	16	49	74

Usual activities

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	1311	149	19	6	3
L2	99	177	124	46	5
L3	4	15	34	49	42

Pain/discomfort

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	1343	140	17	2	6
L2	90	303	125	29	1
L3	2	3	5	10	7

Feeling worried, sad or unhappy

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	838	236	18	5	2
L2	151	557	135	25	6
L3	3	17	27	31	32

(b) Self-report inconsistency (n = 3862)

Mobility (walk around)

Y-3L	L1	L2	L3	L4	L5
			Y-5L		
L1	3268	140	32	5	1
L2	67	184	78	24	0
L3	9	5	13	28	8

Looking after myself

Y-3L	L1	L2	L3	L4	L5
			5L		
L1	3077	151	25	8	1
L2	93	269	128	23	2
L3	8	5	11	50	11

Table 3 (continued)

Usual activities	Y-5L				
Y-3L	**L1**	**L2**	**L3**	**L4**	**L5**
L1	**2586**	262	77	9	5
L2	187	**338**	202	56	13
L3	7	15	30	**52**	**23**
Pain/discomfort	Y-5L				
Y-3L	**L1**	**L2**	**L3**	**L4**	**L5**
L1	**1946**	364	49	7	1
L2	209	**757**	333	59	2
L3	8	8	31	**67**	**21**
Feeling worried, sad or unhappy	Y-5L				
Y-3L	**L1**	**L2**	**L3**	**L4**	**L5**
L1	**1340**	441	39	11	4
L2	251	**1108**	288	97	24
L3	7	36	64	**78**	**74**

Consistent levels are shown in bold

For Y-3L: L1, no; L2, some; L3, a lot of

For Y-5L: L1, no, L2, a little bit of; L3, some; L4, a lot; L5, cannot or extreme

For worried sad or unhappy: L3, quite; L4, really

Y-3L EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

Table 4 Domain responses for Y-3L and Y-5L in different condition groups

(a) Proxy-reports

Condition	Level	Mobility (walk around)				Looking after myself				Usual activities				Pain				Feeling worried, sad or unhappy			
		Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%
Recurrent abdominal pain (n = 124)	1	118	95.16	114	91.94	100	80.65	93	75.00	89	71.77	89	71.77	46	37.10	34	27.42	58	46.77	42	33.87
	2	6	4.84	10	8.06	23	18.55	23	18.55	31	25.00	23	18.55	77	62.10	60	48.39	62	50.00	70	56.45
	3	0	0	0	0	1	0.81	6	4.84	4	3.23	7	5.65	1	0.81	26	20.97	4	3.23	8	6.45
	4			0	0			2	1.61			5	4.03			3	2.42			2	1.61
	5			0	0			0	0			0	0			1	0.81			2	1.61
ADHD (n = 188)	1	177	94.15	169	89.89	118	62.77	112	59.57	106	56.38	103	54.79	145	77.13	136	72.34	72	38.30	72	38.30
	2	10	5.32	12	6.38	65	34.57	45	23.94	71	37.77	54	28.72	41	21.81	38	20.21	106	56.38	84	44.68
	3	1	0.53	7	3.72	5	2.66	26	13.83	11	5.85	23	12.23	2	1.06	7	3.72	10	5.32	19	10.11
	4			0	0			2	1.06			7	3.72			5	2.66			9	4.79
	5			0	0			3	1.60			1	0.53			2	1.06			4	2.13
Anxiety or depression (n = 95)	1	89	93.68	84	88.42	82	86.32	78	82.11	65	68.42	62	65.26	62	65.26	58	61.05	24	25.26	22	23.16
	2	6	6.32	8	8.42	10	10.53	7	7.37	23	24.21	17	17.89	32	33.68	26	27.37	61	64.21	46	48.42
	3	0	0	3	3.16	3	3.16	5	5.26	7	7.37	7	7.37	1	1.05	8	8.42	10	10.53	17	17.89
	4			0	0			5	5.26			6	6.32			2	2.11			8	8.42
	5			0	0			0	0			3	3.16			1	1.05			2	2.11
ASD (n = 183)	1	159	86.89	158	86.34	70	38.25	62	33.88	72	39.34	60	32.79	130	71.04	122	66.67	56	30.60	51	27.87
	2	21	11.48	16	8.74	85	46.45	40	21.86	86	46.99	56	30.60	49	26.78	33	18.03	107	58.47	82	44.81
	3	3	1.64	7	3.83	28	15.30	49	26.78	25	13.66	42	22.95	4	2.19	22	12.02	20	10.93	31	16.94
	4			1	0.55			19	10.38			19	10.38			2	1.09			10	5.46
	5			1	0.55			13	7.10			6	3.28			4	2.19			9	4.92
Asthma (n = 185)	1	179	96.76	174	94.05	155	83.78	149	80.54	162	87.57	157	84.86	152	82.16	147	79.46	120	64.86	106	57.30
	2	5	2.70	8	4.32	26	14.05	28	15.14	23	12.43	22	11.89	33	17.84	33	17.84	64	34.59	69	37.30
	3	1	0.54	2	1.08	4	2.16	3	1.62	0	0	3	1.62	0	0	4	2.16	1	0.54	8	4.32
	4			1	0.54			5	2.70			3	1.62			1	0.54			1	0.54
	5			1	0.54			0	0			0	0			0	0			1	0.54
Dental (n = 173)	1	162	93.64	160	92.49	140	80.92	123	71.10	153	88.44	141	81.50	135	78.03	124	71.68	128	73.99	115	66.47
	2	9	5.20	6	3.47	30	17.34	38	21.97	17	9.83	18	10.40	35	20.23	38	21.97	41	23.70	47	27.17
	3	2	1.16	3	1.73	3	1.73	7	4.05	3	1.73	12	6.94	3	1.73	6	3.47	4	2.31	7	4.05
	4			3	1.73			5	2.89			2	1.16			4	2.31			4	2.31
	5			1	0.58			0	0			0	0			1	0.58			0	0

Table 4 (continued)

(a) Proxy-reports

Condition	Level	Mobility (walk around)				Looking after myself				Usual activities				Pain				Feeling worried, sad or unhappy			
		Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%
Eating disorder (n = 41)	1	35	85.37	33	80.49	33	80.49	31	75.61	25	60.98	22	53.66	20	48.78	19	46.34	6	14.63	6	14.63
	2	5	12.20	5	12.20	6	14.63	4	9.76	8	19.51	11	26.83	18	43.92	13	31.71	24	58.54	15	36.59
	3	1	2.44	2	4.88	2	4.88	4	9.76	8	19.51	3	7.32	3	7.32	6	14.63	11	26.83	13	31.71
	4			1	2.44			1	2.44			5	12.20			2	4.88			3	7.32
	5			0	0			1	2.44			0	0			1	2.44			4	9.76
Epilepsy (n = 96)	1	61	63.54	63	65.63	41	42.78	35	36.46	40	41.67	35	36.46	49	51.42	51	53.13	39	40.63	37	38.54
	2	19	19.79	6	6.25	21	21.88	13	13.54	30	31.25	17	17.71	44	45.83	28	29.17	47	48.96	42	43.75
	3	16	16.67	10	10.42	34	35.42	10	10.42	26	27.83	17	17.71	3	3.13	12	12.50	10	10.42	7	7.29
	4			6	6.25			12	12.50			14	14.58			4	4.17			4	4.17
	5			11	11.46			26	27.08			13	13.54			1	1.04			6	6.25
Sleep (n = 156)	1	149	95.51	148	94.87	97	62.18	89	57.05	107	68.59	101	64.74	118	75.64	118	75.64	65	41.67	56	35.90
	2	7	4.49	5	3.21	48	3.77	42	26.92	40	25.64	26	16.67	37	23.72	29	18.59	80	51.28	71	45.51
	3	0		2	1.28	11	7.51	16	10.26	9	5.77	15	9.62	1	0.64	7	4.49	11	7.51	18	11.54
	4			1	0.64			4	2.56			10	6.41			2	1.28			9	5.77
	5			0	0			5	3.21			4	2.56			0	0			2	1.28

(b) Self-reports

Condition	Level	Mobility (walk around)				Looking after myself				Usual activities				Pain				Feeling worried, sad or unhappy			
		Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%
Recurrent abdominal pain (n = 268)	1	237	88.43	228	85.07	229	85.45	226	84.33	180	67.16	176	65.67	56	20.90	50	18.66	88	32.84	83	30.97
	2	29	10.83	28	10.45	38	14.18	32	11.94	78	29.14	51	19.03	190	70.90	118	44.03	154	57.46	126	47.01
	3	2	0.75	7	2.61	1	0.37	9	3.36	10	3.73	28	10.45	22	8.29	80	29.85	26	9.71	35	13.06
	4			4	1.49			1	0.37			6	2.24			19	7.09			20	7.46
	5			1	0.37			0	0			7	2.61			1	0.37			4	1.49
ADHD (n = 304)	1	272	89.47	258	84.87	236	77.63	231	75.99	202	66.45	197	64.80	189	62.17	175	57.57	112	36.84	91	29.93
	2	29	9.54	31	10.20	60	19.74	46	15.13	97	31.98	66	21.71	108	35.53	90	29.61	178	58.55	150	49.34
	3	3	0.99	12	3.95	8	2.63	17	5.59	5	1.64	30	9.87	7	2.33	28	9.21	14	4.65	38	12.50
	4			3	0.99			9	2.96			10	3.29			8	2.63			20	6.58
	5			0	0			1	0.33			1	0.33			3	0.99			5	1.64

Table 4 (continued)

(b) Self-reports

Condition	Level	Mobility (walk around)				Looking after myself				Usual activities				Pain				Feeling worried, sad or unhappy			
		Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%
Anxiety or depression (n = 385)	1	346	89.87	334	86.75	331	85.97	331	85.97	269	69.87	254	65.97	220	57.14	198	51.43	83	21.56	62	16.10
	2	32	8.31	38	9.87	51	13.25	35	9.09	105	27.27	79	20.52	159	41.30	129	33.51	263	68.31	188	48.83
	3	7	1.82	9	2.34	3	0.78	16	4.16	11	2.86	39	10.13	6	1.56	44	11.43	39	10.13	77	20.00
	4			4	1.04			3	0.78			11	2.86			12	3.12			41	10.65
	5			0	0			0	0			2	0.52			2	0.52			17	4.42
ASD (n = 327)	1	272	83.18	253	77.37	201	61.47	185	56.57	189	57.80	149	45.57	186	56.89	160	48.93	92	28.13	75	22.94
	2	53	16.28	49	14.98	107	32.72	82	25.08	122	37.39	90	27.52	128	39.14	112	34.25	203	62.80	164	50.15
	3	2	0.61	16	4.89	19	5.81	41	12.54	16	4.89	59	18.04	13	3.98	39	11.93	32	9.79	53	16.21
	4			9	2.75			18	5.50			24	7.34			13	3.98			23	7.03
	5			0	0			1	0.31			5	1.53			3	0.92			12	3.67
Asthma (n = 302)	1	275	91.60	269	89.07	276	91.40	270	89.40	264	87.42	257	85.10	209	69.25	194	64.24	188	62.25	169	55.96
	2	17	5.63	20	6.62	21	6.95	40	13.25	35	11.59	26	8.61	82	27.15	78	25.83	104	34.44	111	36.75
	3	10	3.31	8	2.65	5	1.66	4	1.32	3	0.99	10	3.31	11	3.64	22	7.28	10	3.31	14	4.64
	4			5	1.66			8	2.65			8	2.65			5	1.66			5	1.66
	5			0	0			0	0			1	0.33			3	0.99			3	0.99
Dental (n = 317)	1	296	93.38	291	91.80	291	91.80	285	89.91	283	89.27	277	87.38	228	71.92	204	64.35	189	59.62	164	51.74
	2	18	5.68	15	4.73	24	7.58	21	6.62	32	10.95	27	8.52	82	25.87	94	29.65	121	38.17	132	41.64
	3	3	0.95	9	2.84	2	0.64	8	2.52	2	0.64	10	3.15	7	2.28	9	2.84	7	2.28	16	5.05
	4			2	0.63			3	0.95			3	0.95			10	3.15			3	0.95
	5			0	0			0	0			0	0			0	0			2	0.63
Eating disorder (n = 145)	1	108	74.48	104	71.72	107	73.79	112	77.24	73	50.34	57	39.31	46	31.72	37	25.52	24	16.55	16	11.03
	2	33	22.76	24	16.55	29	20.00	19	13.10	58	40.00	42	28.97	86	59.31	56	38.62	72	49.66	47	32.41
	3	4	2.76	14	9.66	9	6.27	4	2.76	14	9.66	27	18.62	13	8.97	34	23.45	49	33.79	36	24.83
	4			2	1.38			8	5.52			13	8.97			15	10.34			22	15.17
	5			1	0.69			2	1.38			6	4.14			3	2.07			24	16.55
Epilepsy (n = 176)	1	153	86.93	152	86.36	136	77.27	122	69.32	114	64.77	112	63.64	102	57.95	96	54.55	83	47.16	75	42.61
	2	16	10.00	8	4.55	31	17.61	34	19.32	47	26.75	38	21.59	65	36.93	48	27.27	84	47.73	73	41.48
	3	7	3.98	8	4.55	9	5.11	10	5.68	15	8.52	16	9.09	9	5.11	21	11.93	9	5.11	14	7.95
	4			8	4.55			9	5.11			8	4.55			11	6.25			4	2.27
	5			0	0			1	0.57			2	1.14			0	0			10	5.68

Table 4 (continued)

(b) Self-reports

Condition	Level	Mobility (walk around)				Looking after myself				Usual activities				Pain				Feeling worried, sad or unhappy			
		Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%	Y-3L	%	Y-5L	%
Sleep (n = 190)	1	168	88.42	160	84.21	148	77.89	138	72.63	135	71.53	126	66.32	104	54.74	87	45.79	70	36.84	51	26.84
	2	21	11.53	20	10.53	39	20.53	32	16.84	48	25.26	35	18.42	81	42.63	69	36.32	105	55.26	96	50.53
	3	1	0.53	8	4.21	3	1.58	15	7.89	7	3.68	21	11.05	5	2.63	25	13.16	15	7.89	26	13.68
	4			1	0.53			4	2.11			8	4.21			7	3.68			9	4.74
	5			1	0.53			1	0.53			0	0			2	1.05			8	4.21

ADHD attention-deficit hyperactivity disorder, *ASD* autistic spectrum disorders, *Y-3L* EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

informativity and discriminatory performance of the Y-5L. Higher Shannon index in the Y-5L can be interpreted as more information being distributed in the levels. The Shannon Evenness index (J′) illustrates that the relative informativity of the change from Y-3L to Y-5L is almost comparable, with only a marginal decrease in all domains.

3.2.5 Test–Retest Reliability

The Kappa between the EQ-VAS scores categorised in 10 groups was 0.74 ($p < 0.001$) for the proxy-reported group, indicating substantial agreement, while the corresponding Kappa for the self-reported group was 0.51 ($p < 0.001$), indicating moderate agreement. The Kappa coefficient for the domains are presented in Table 7; the *p*-value for all domains was < 0.001. Table 7 shows that the reliability pattern is not consistent for either of the instruments and that it is domain-driven. Mobility had the highest agreement in both groups.

4 Discussion

The aim of this study was to assess the psychometric performance of the Y-5L in comparison with the Y-3L, in both self-reported and proxy-report responses. The psychometric performance of Y-5L compared with Y-3L was reported in terms of ceiling effects, criterion validity, inconsistency and redistribution properties, informativity (discriminatory power), and test-retest reliability. Overall, Y-5L is a valid and reliable extension of the Y-3L. The domains have a high correlation, suggesting evidence of criterion validity. In addition, the Y-5L showed superior discriminatory power, an improved distribution, and slightly reduced ceiling effects compared with the Y-3L (some conditions showed a substantial reduction in ceiling effect). Ceiling effects were still present for the Y-5L, however they were lower compared with the Y-3L. The lower ceiling effect for Y-5L has also been seen in other studies for children [9, 12], young adult and adult [13, 36] populations. The ceiling effect was high in the proxy-reported data, which might be caused by different reporting patterns, with children potentially more likely to report a problem. For both groups, the highest proportion of problems was reported for 'feeling worried, sad or unhappy', while the percentage of reported problems for 'mobility' was small. These results are consistent with the results of another study comparing the youth instruments [9]. The Y-5L has a lower ceiling effect compared with the Y-3L, which demonstrates that the increase in levels leads to more granular responses when reporting mild health problems. A reduction in ceiling effect also suggests that improvement (e.g. from Level 2 to Level 1) over time may be captured on the Y-5L which would not be detected by the Y-3L, although responsiveness testing is needed to test this.
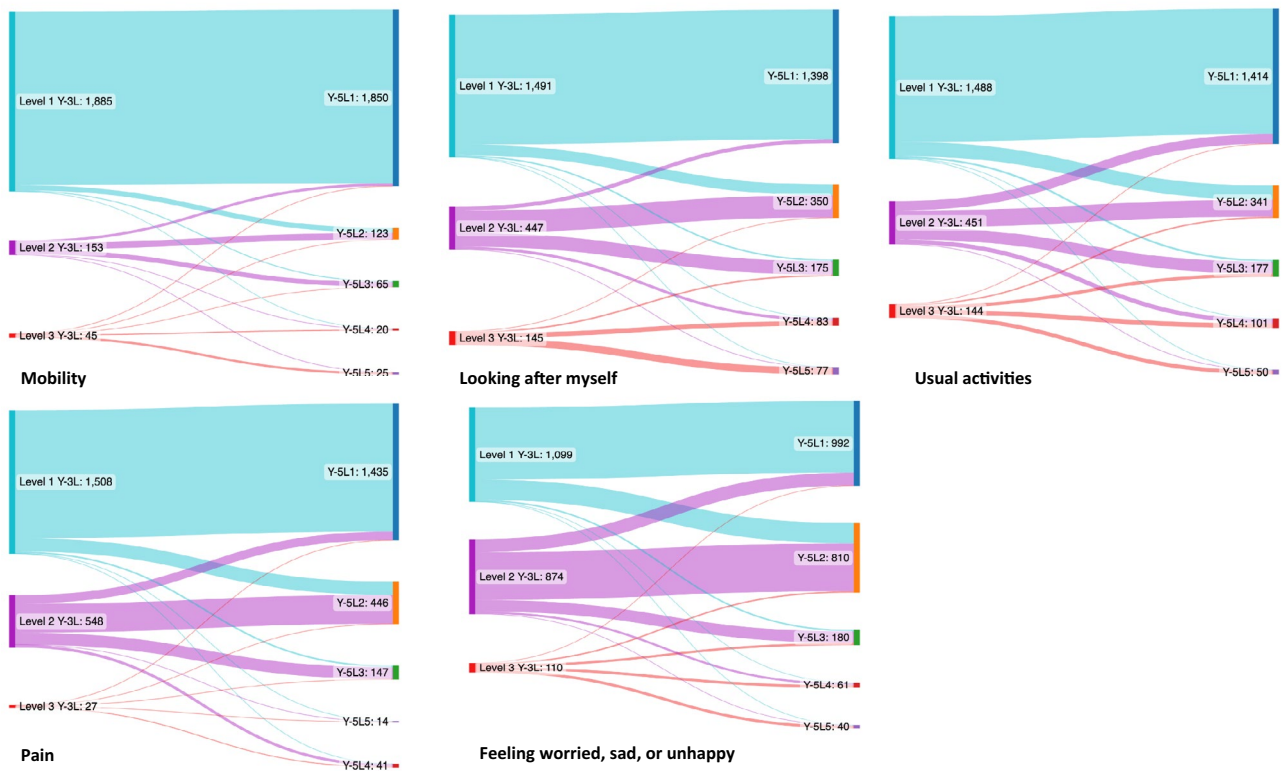
Fig. 1 Sankey diagrams for level proportions for proxy-reports at baseline

All the condition groups showed a lower ceiling effect for the Y-5L compared with the Y-3L, except for the epilepsy proxy-reported data.

The low percentage of inconsistencies found when comparing the two instruments shows that the Y-3L data redistribute logically to the Y-5L, and that the Y-5L and Y-3L descriptive systems are comparable. This could in part be due to a high ceiling effect, where both instruments tend to show that children, especially those from the general population, do not have any problems in a certain area, even with a refined response scale. This has been previously observed in studies of both adults and young adults in general populations [36–38].

The two instruments showed a high degree of correlation based on the criteria. The remaining variation is driven by the level differences, which could be due to the inconsistencies in wording across the levels, and the reason for the differences could be qualitatively further investigated. The highest inconsistency was related to 'feeling worried, sad or unhappy', which also had the greatest difference in the Shannon index, i.e. gained more discriminatory power when changing from the 3L to 5L. Hence 'feeling worried, sad or unhappy' is the domain that is mostly changing when adding levels to the Y-3L in our data. The greater number of differences observed in the 'feeling worried, sad or unhappy' domain is likely due to the larger number of responses that are not level 1, meaning greater potential for inconsistency.

The inclusion of additional levels in the instrument appears to have had an impact on the distribution of responses within the condition groups, particularly within the 'pain' and 'feeling worried, sad or unhappy' domains. These two domains also showed more variability in another youth comparison study (9). It is conceivable that changing the level 2 description in the Y-5L from 'some problem' to 'little problem' may have influenced participants to select level 2 over level 1, or having a more extreme level in Y-5L has caused the distribution from level 3 to levels 4 and 5. The spread of responses in different levels, especially from level 3 of Y-3L to levels 4 and 5 of Y-5L, has also been seen in adult condition groups (14). This might be helpful when used by clinicians, alongside clinical trials, and other surveys to obtain more detailed data when using the Y-5L. This could be further tested in studies that include assessment of responsiveness when comparing the two instruments across different condition groups.

Extending the levels of the EQ-5D-Y descriptive system from three to five levels resulted in higher absolute,
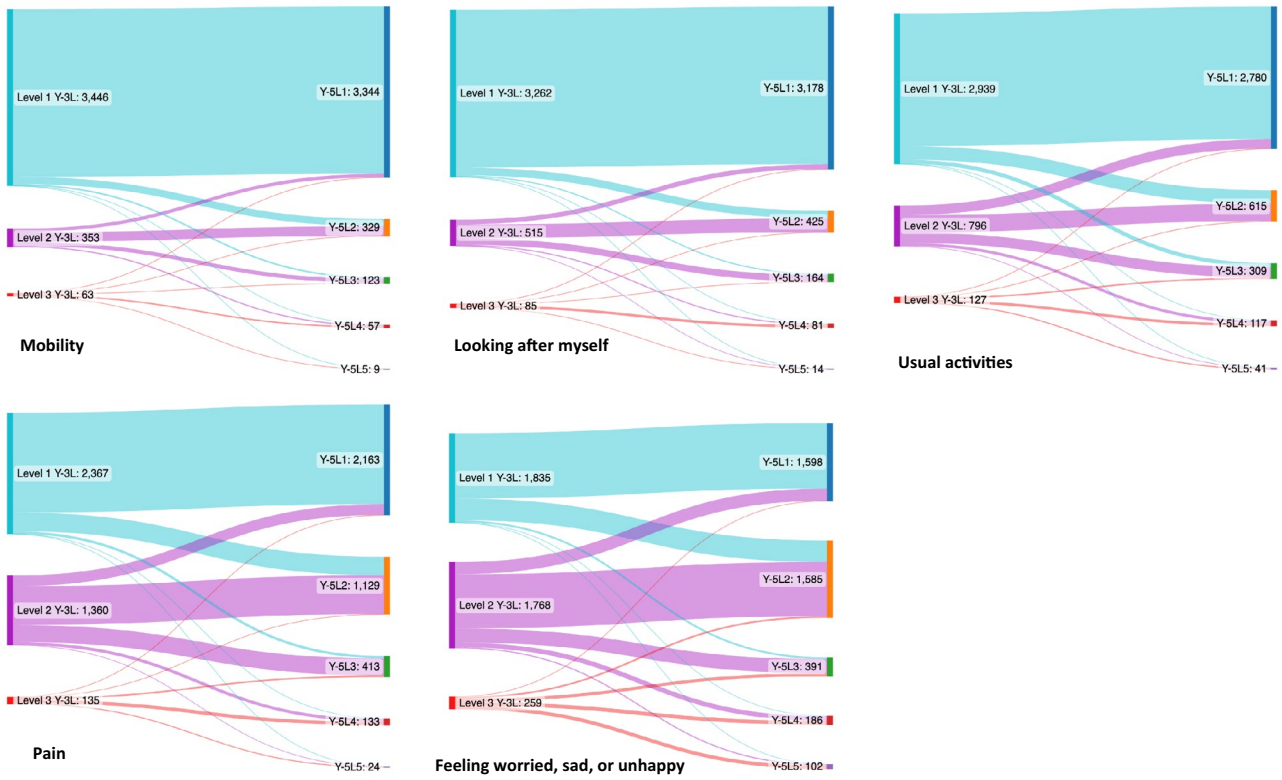
Fig. 2 Sankey diagrams for level proportions for self-reports at baseline

Table 5 Criterion validity of Y-3L and Y-5L in proxy-reports (*n* = 2083) and for self-reports (*n* = 3862) using correlation coefficients

	Y-5L				
	Mobility (walk around)	Looking after myself	Usual activities	Pain	Feeling worried, sad or unhappy
Proxy-reports					
Y-3L					
Mobility (walk around)	**0.77**	0.40	0.40	0.34	0.16
Looking after myself	0.39	**0.84**	0.55	0.24	0.23
Usual activities	0.42	0.59	**0.73**	0.37	0.37
Pain	0.37	0.26	0.36	**0.72**	0.31
Feeling worried, sad or unhappy	0.12	0.22	0.37	0.27	**0.64**
Self-reports	**Y-5L**				
Y-3L					
Mobility (walk around)	**0.71**	0.33	0.40	0.39	0.21
Looking after myself	0.35	**0.75**	0.47	0.25	0.27
Usual activities	0.41	0.44	**0.66**	0.36	0.39
Pain	0.38	0.23	0.37	**0.69**	0.31
Worried, sad or unhappy	0.20	0.24	0.39	0.32	**0.65**

The highest correlation for each domain is shown in bold

Y-3L EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

Table 6 Discriminatory and informativity power between Y-3L and Y-5L

	Proxy-reports				Self-reports			
	Y-3L		Y-5L		Y-3L		Y-5L	
	H'	J'	H'	J'	H'	J'	H'	J'
Mobility (walk around)	0.36	0.33	0.48	0.30	0.39	0.35	0.52	0.32
Looking after myself	0.75	0.69	1.03	0.64	0.49	0.45	0.64	0.40
Usual activities	0.76	0.69	1.00	0.62	0.64	0.58	0.88	0.55
Pain	0.64	0.58	0.88	0.55	0.75	0.71	1.07	0.66
Feeling worried, sad or unhappy	0.86	0.79	1.1	0.69	0.89	0.81	1.20	0.75

H' Shannon index, *J'* Shannon evenness index, *Y-3L* EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

Table 7 Kappa coefficients for test-retest for Y-3L and Y-5L (*n* = 115)

	Proxy-reports (*n* = 55)		Self-reports (*n* = 60)	
	Y-3L	Y-5L	Y-3L	Y-5L
Mobility (walk around)	1	0.56	0.88	0.85
Looking after myself	0.70	0.54	0.85	0.90
Usual activities	0.66	0.27	0.58	0.76
Pain	0.38	0.36	0.59	0.56
Feeling worried, sad or unhappy	0.51	0.35	0.57	0.67

Y-3L EQ-5D-Y-3L, *Y-5L* EQ-5D-Y-5L

discriminatory power. Furthermore, the extension to five levels resulted in a diverse range of responses across severity levels for most condition groups, aligning consistently with findings from an adult comparison study [13]. The relative evenness of the Y-5L was comparable or slightly lower than that of the Y-3L. This trend aligns with observations from previous comparative studies involving adults [13, 27, 39].

In summary, the Y-5L provides more nuanced information about the severity of young people's HRQoL by adding response levels; however, the reliability and responsiveness of Y-5L still needs to be determined for different condition groups. Both versions of EQ-5D-Y are useful tools for measuring HRQoL in young people, and the choice of which one to use will depend on the specific research question, condition group, study design, and available resources.

This study is not without limitations, one of which is that our analysis relied on using level sum scores due to the current lack of Australian value sets for either the Y-3L or Y-5L. These value sets are however currently in development. Future research can extend the analysis by preference weighting the descriptive data. Another limitation was a relatively small sample size for the 2 day interval for evaluating test-retest reliability, compared with the sample used for the other analyses conducted in this study. The study does not include a dyad sample to facilitate a direct comparison between self and proxy reports. Consequently, we assessed self-reported and proxy-reported data separately. Future research using dyad samples would allow for formal comparison of self- and proxy-reported data psychometric characteristics.

Overall, the results confirm the psychometric performance of Y-5L in the Australian paediatric population. The Y-5L is considered a valid and reliable instrument for measuring HRQoL in children and adolescents. The Y-5L may provide slightly more detail compared with the Y-3L and could be considered for use in clinical practice/clinical trials or other evaluations.

References

1. Gusi N, P Olivares, R Rajendram. The EQ-5D Health-Related Quality of Life Questionnaire. Springer; 2010. p. 87–99.
2. 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 Health. 2019;22(1):31–7.
3. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, Egmar AC, Greiner W, Gusi N, Herdman M, Jelsma J, Kind P, Scalone L, Ravens-Sieberer U. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res. 2010;19(6):875–86.
4. Burström K, Egmar A-C, Lugnér A, Eriksson M, Svartengren M. A Swedish child-friendly pilot version of the EQ-5D instrument—the development process. Eur J Pub Health. 2011;21(2):171–7.
5. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, Egmar AC, Gusi N, Herdman M, Jelsma J, Kind P, Olivares PR, Scalone L, Greiner W. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res. 2010;19(6):887–97.
6. Otto C, Barthel D, Klasen F, Nolte S, Rose M, Meyrose A-K, Klein M, Thyen U, Ravens-Sieberer U. Predictors of self-reported health-related quality of life according to the EQ-5D-Y in chronically ill children and adolescents with asthma, diabetes, and juvenile arthritis: longitudinal results. Qual Life Res. 2018;27:879–90.
7. Eidt-Koch D, Mittendorf T, Greiner W. Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. BMC Pediatr. 2009;9(1):1–8.
8. Kreimeier S, Åström M, Burström K, Egmar A-C, Gusi N, Herdman M, Kind P, Perez-Sousa MA, Greiner W. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res. 2019;28(7):1951–61.
9. 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 Health. 2021;24(12):1799–806.
10. Verstraete J, Scott D. Comparison of the EQ-5D-Y-5L, EQ-5D-Y-3L and PedsQL in children and adolescents. J Patient Reported Outcomes. 2022;6(1):67.
11. Wong CKH, Cheung PWH, Luo N, Cheung JPY. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. Eur J Health Econ. 2019;20(5):647–56.
12. Fitriana TS, Purba FD, Rahmatika R, Muhaimin R, Sari NM, Bonsel G, Stolk E, Busschbach JJ. Comparing measurement properties of EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients. Health Qual Life Outcomes. 2021;19(1):1–12.
13. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22(7):1717–27.
14. Kim SH, Kim HJ, Lee SI, Jo MW. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. Qual Life Res. 2012;21(6):1065–73.
15. Chang PC, Yeh CH. Agreement between child self-report and parent proxy-report to evaluate quality of life in children with cancer. Psychooncology. 2005;14(2):125–34.
16. Jiang M, Ma Y, Li M, Meng R, Ma A, Chen P. A comparison of self-reported and proxy-reported health utilities in children: a systematic review and meta-analysis. Health Qual Life Outcomes. 2021;19(1):45.
17. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care. 2005;43(5):493.
18. Mack JW, McFatrich M, Withycombe JS, Maurer SH, Jacobs SS, Lin L, et al. Agreement between child self-report and caregiver-proxy report for symptoms and functioning of children undergoing cancer treatment. JAMA Pediatr. 2020;174(11):e202861.
19. Jones R, Mulhern B, McGregor K, Yip S, O'Loughlin R, Devlin N, Hiscock H, Dalziel K, T 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 (Basel), 2021. 8(8).
20. Jones R, O'Loughlin R, Xiong X, Bahrampour M, McGregor K, Yip S, Devlin N, Hiscock H, Mulhern B, Dalziel K, o.b.o.t.Q.o. L.i.K.K.E.t.S.D.i.A.P. Team, Collecting paediatric health-related quality of life data: assessing the feasibility and acceptability of the australian paediatric multi-instrument comparison (P-MIC) study. Children, 2023. 10(10):1604.
21. Jones R, Mulhern B, Devlin N, Hiscock H, O'Loughlin R, McGregor K, Shilana Y, Dalziel K, on behalf of the 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. Version 2 [online]; 2022.
22. van Heusden A, Rivero-Arias O, Herdman M, Hiscock H, Devlin N, Dalziel K, in collaboration with the Quality Of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team. Psychometric performance comparison of the adapted versus original versions of the EQ-5D-Y-3L and -Y-5L in proxy respondents for 2- to 4-year-olds. PharmacoEconomics, Epub 18 Jan 2024. https://doi.org/10.1007/s40273-024-01350-6.
23. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293–307.

25. Wang P, Luo N, Tai ES, Thumboo J. The EQ-5D-5L is more discriminative than the EQ-5D-3L in patients with diabetes in Singapore. Value Health Reg Issues. 2016;9:57–62.

26. Jia YX, Cui FQ, Li L, Zhang DL, Zhang GM, Wang FZ, Gong XH, Zheng H, Wu ZH, Miao N, Sun XJ, Zhang L, Lv JJ, Yang F. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual Life Res. 2014;23(8):2355–63.

27. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. Value Health. 2008;11(2):275–84.

28. Kennedy ABW, Sankey HR. The Thermal Efficiency of Steam Engines. Report of the Committee Appointed to the Council Upon the Subject of the Definition of a Standard or Standards of Thermal Efficiency for Steam Engines: With an Introductory Note. (Including Appendixes and Plate at Back of Volume). Minutes of the Proceedings of the Institution of Civil Engineers, 1898. 134(1898):278–312.

29. Cohen J. A power primer. Psychol Bull. 1992;112(1):155–9.

30. Jones R, O'Loughlin R, Xiong X, Bahrampour M, Devlin N, Hiscock H, Chen G, Mulhern B, Dalziel K, T. on behalf of the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia Project, 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.

31. Jones R, Xiong OLRX, Bahrampour M, Devlin N, Hiscock H, Chen G, Mulhern B*, Dalziel K*, on behalf of the QUality Of life in Kids: Key evidence to strengthen decisions in Australia (QUOKKA) project team. Comparative psychometric performance of common generic paediatric health related quality of life instruments: Results from the Australian Paediatric Multi-Instrument Comparison Study. . [Submitted]. PharmacoEconomics 12 May 2023., 2023.

32. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27(3):379–423.

33. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull. 1968;70(4):213.

34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 1977:159–174.

35. Jones R, Mulhern B, Devlin N, O'Loughlin R, Xiong X, Bahrampour M, McGregor K, Yip S, Hiscock H, Dalziel K, The Australian paediatric multi-instrument comparison (P-MIC) study: data quality, feasibility, acceptability, and descriptive comparisons of the EQ-5D-Y-3L, EQ-5D-Y-5L, TANDI and PedsQL, in EuroQol Plenary 2022. 2022: Chicago.

36. Martí-Pastor M, Pont A, Ávila M, Garin O, Vilagut G, Forero CG, Pardo Y, Tresserras R, Medina-Bustos A, Garcia-Codina O, Cabasés J, Rajmil L, Alonso J, Ferrer M. Head-to-head comparison between the EQ-5D-5L and the EQ-5D-3L in general population health surveys. Popul Health Metr. 2018;16(1):14.

37. Kim TH, Jo MW, Lee SI, Kim SH, Chung SM. Psychometric properties of the EQ-5D-5L in the general population of South Korea. Qual Life Res. 2013;22(8):2245–53.

38. Ferreira LN, Ferreira PL, Ribeiro FP, Pereira LN. Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. Health Qual Life Outcomes. 2016;14:89.

39. Pattanaphesaj J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. Health Qual Life Outcomes. 2015;13:14.