

## Article

# Comparing Measurement Properties of the English EQ-5D-Y Three-Level Version with the Five-Level Version in South Africa

Janine Verstraete <sup>1</sup>, Razia Amien <sup>2</sup> and Des Scott <sup>2\*</sup>

<sup>1</sup> University of Cape Town, Department of Paediatrics and Child Health, Division of Pulmonology Cape Town, South Africa; janine.verstraete@uct.ac.za

<sup>2</sup> University of Cape Town, Faculty of Health Sciences, Department of Health and Rehabilitation Sciences, Division of Physiotherapy, Cape Town, South Africa; amnrz001@myuct.ac.za

\* Correspondence: des.scott@uct.ac.za

**Abstract:** Objectives: The aim of this study was to compare the performance of the EQ-5D-Y-3L (Y-3L) and the expanded five level version, the EQ-5D-Y-5L (Y-5L). Methods: Children/adolescents with an acute or chronic health condition and from the general population completed the Y-5L and Y-3L self-report questionnaires. Performance of the Y-5L and Y-3L was determined by comparing feasibility, redistribution of dimension responses, discriminatory power, validity, and test-retest reliability. Results. Five hundred and fifty children/adolescents completed baseline measures and 173 completed repeat measures. The ceiling effect decreased by 15% from the Y-3L to Y-5L. Informativity of dimensions improved by 0.094 on the Y-5L. There was a range of 4-9% inconsistent responses moving from the Y-3L to Y-5L. Convergent validity of the Visual Analogue Scale (VAS) and Y-3L, Y-5L dimensions was similar, weak to moderate (rs range: 0.18 – 0.38) but similar and strong on paired Y-3L and Y-5L dimensions: Kendall Tau B (range 0.69 – 0.80) and Gamma (range 0.92 – 0.98). The Y-5L and Y-3L showed moderate to substantial agreement for test-retest reliability across dimensions and VAS scores in stable chronic health conditions and fair agreement for the general population. Conclusion: The EQ-5D-Y-5L is a valid, reliable extension of the Y-3L for children/adolescents across health conditions and healthy children/adolescents. The expanded levels reduced the ceiling effect. The relative informativity of report across dimensions increased on the Y-5L compared to the Y-3L with retention of the evenness of reporting. The convergent validity and test-retest reliability of the Y-5L was comparable to the Y-3L..

**Keywords:** Health Related Quality of Life, EQ-5D-Y-3L, EQ-5D-Y-5L, Youth, children, orthopaedic

## 1. Introduction

The EQ-5D is widely used to measure and value health in adults across the world. The EQ-5D-Y, youth version, was developed in 2009 for self-completion in children aged 8-15 years<sup>1</sup>. Therefore the adult measure was modified to include youth friendly wording and examples across dimensions and a general rating of health on a visual analogue scale (VAS) from 0 (worst health) to 100 (best health). The EQ-5D-Y has five dimensions namely *Mobility* (walking about), *Looking After Myself* (washing and dressing), *Doing Usual Activities* (going to school, hobbies, sports, playing, doing things with family or friends), *Having Pain or Discomfort* and *Feeling Worried, Sad or Unhappy*. The original youth version, EQ-5D-Y-3L (Y-3L), describes health on three levels (no problems, some problems and a lot of problems) resulting in 243 (3<sup>5</sup>) health states <sup>1,2</sup>.

Since the development of the Y-3L, the adult version has expanded its levels of report from three to five <sup>3,4</sup>. Expanding the response option on the adult EQ-5D, improved the discriminatory power in general and patient populations with decreased ceiling effect <sup>4-7</sup>. Thus, the response options of the youth version were recently expanded to five levels resulting in the EQ-5D-Y-5L (Y-5L) <sup>8</sup>. The expanded version describes health on five levels

[no/not, a little bit, some/quite, a lot/really, cannot/extreme(ly)] resulting in 3125 (5<sup>5</sup>) health states.

A head-to-head comparison of the Y-5L and Y-3L in paediatric patients with idiopathic scoliosis, 8-17 years, showed a low inconsistency in responses between the two measures, a reduction in the ceiling effect and good test-retest agreement <sup>9</sup>. The Y-5L had similar moderate correlations to the Kidscreen in children with and without short-sightedness and overweight <sup>10</sup>. There is further evidence that the Y-5L is feasible to use with children receiving psychiatric care <sup>11</sup>. These studies however had modest sample sizes with an older mean age of respondent (13-15 years) possibly due to the wider patient age range of 8-17 years, in participants. The targeted age range of children/adolescents reporting on the EQ-5D-Y is 8-15 years <sup>1</sup>.

Including the child's perspective on how a health condition and its management affect their wellbeing, is increasingly used in decision making. Therefore it is important that the measurement properties of new or modified HRQoL instruments are rigorously tested, in different populations of children before the instruments are widely used <sup>12</sup>. This paper presents a head-to-head comparison of the performance of the English Y-3L and Y-5L in a large sample of South African children/adolescents aged 8-15 years. The aim of this study was to investigate the feasibility, redistribution and discriminatory power of dimension responses, convergent validity, known group validity, and the test-retest reliability of the Y-3L and the Y-5L.

## 2. Methods

### *Study design and participants*

An observational, analytical cohort study with repeated measures for test-retest reliability was conducted. Three research settings, each with children/adolescents in different health states, but from similar socio-economic backgrounds (low to middle income) in Cape Town, South Africa, were included in the study. In the first research setting, children/adolescents requiring acute medical/surgical treatment were recruited from an acute tertiary paediatric hospital and a paediatric orthopaedic hospital. These children/adolescents were from medical, surgical or trauma wards. Secondly, children/adolescents with stable chronic health conditions were recruited from schools for learners with special needs. These schools have specialised education services for learners with normal intellect diagnosed with physical disabilities (e.g. cerebral palsy, spina bifida or muscle diseases) and/or learning disabilities (e.g. autism spectrum disorder or attention deficit disorder). Thirdly, children/adolescents attending two similar mainstream schools, with generally healthy learners, were recruited as a general population sample.

All children/adolescents aged 8-15 years, who were able to read and write English, at each facility were eligible for the study. Only those who returned a signed informed consent and assent were included in the study and those who were critically ill or who were medically unstable were excluded as the research may have been too distressing. The sample size of 550 participants was powered at 95% to detect a difference in correlation of dimension scores between the three condition groups with a small effect size 0.4 and a significance of 0.05.

### *Instruments*

#### EQ-5D-Y

The official Y-3L English version for South Africa was used in this study. The experimental Y-5L English version for the United Kingdom was tested for equivalence in English for South Africa by the EuroQol group. The three or five levels of the descriptive system are expressed with a five-digit code. For example 11223 describes someone with no problems with *Mobility*, no problems with *Looking After Myself*, some problems with *Usual Activities*, some *Pain or Discomfort* and very *Worried, Sad or Unhappy*. Although the Y-3L has a preference-based score the Y-5L does not <sup>13,14</sup>. As such a level sum score (LSS) was

used to describe the responses on the descriptive system where the level labels are treated as numeric data with the best possible score  $(1+1+1+1+1)=5$  and the most severe Y-3L score is  $(3+3+3+3+3)=15$ . The other health states will have a LSS between 5-15 on the Y-3L and 5-25 on the Y-5L, with a larger score indicating a worse health state<sup>15</sup>. This is a crude score with limitations<sup>16,17</sup> but given the differences between the Y-3L and adult value sets, the LSS may give a better indication of performance than the adult EQ-5D-3L and EQ-5D-5L value sets<sup>14,15</sup>. Children/adolescents rated their general health on both the Y-3L and Y-5L VAS forms as the latter recorded the answer in an additional box.

Other instruments used in the study (PedsQL Generic and the Self-Rated Health Question) are not explicitly described in this paper.

### *Procedure*

Ethics approval was obtained from the University of Cape Town, Faculty of Health Sciences, Human Research Ethics Committee ([HREC 154\\_2019](#)). The study was carried out in accordance with the declaration of Helsinki involving human participants<sup>18</sup> and the recommended Covid precautions.

Children/adolescents aged 8-15 years admitted to either of the acute inpatient hospital settings were recruited during an onsite visit. For those who were willing and provided consent and assent the parent was asked to complete the socio-demographic information for the child and the children/adolescents self-completed the Y-5L, PedsQL, Self-Rated Health (SRH) Question and Y-3L in that order. Only the results from the Y-5L and Y-3L are reported on in this paper. The Y-5L was presented first as testing of the adult measures found that if the three-level version was presented first the additional levels on the five-level were not considered<sup>4</sup>.

Due to the Covid pandemic children/adolescents at schools were recruited through information leaflets that were sent home. For those who were willing and provided consent and assent the instruments were independently self-completed by the child/adolescent. The participants with a stable chronic condition or from the general population who returned the research packs at baseline were invited to complete a second measure of the Y-5L and Y-3L 7-14 days later to determine test-retest reliability. Acutely ill children were not included as their health condition was expected to change. There are no current guidelines on the time period for retest but Marx et al (2003) found no difference between two days and two weeks<sup>19</sup>.

### *Data Management and Analysis*

#### *General Performance and Feasibility:*

The analysis of the Y-3L and Y-5L followed that of Janssen et al (2008) for comparison of the adult versions EQ-5D-3L and EQ-5D-5L<sup>20</sup>. The EQ-5D-Y responses and descriptive data were summarised in terms of frequency of dimension responses across condition groups. The feasibility was assessed by comparing the number of missing values for each EQ-5D-Y measure. The ceiling of the EQ-5D-Y was defined as the proportion of children/adolescents scoring no problems across all five dimensions (11111). The absolute reduction in proportion scoring 11111 from the 3L to the 5L was calculated and due to the small number of respondents with an acute or chronic health condition reporting 11111 a percentage reduction was also calculated as  $(\text{ceiling}_{Y-3L} - \text{ceiling}_{Y-5L}) / \text{ceiling}_{Y-3L}$ .

#### *Redistribution Properties of the EQ-5D-Y-3L to the EQ-5D-Y-5L:*

Paired dimension responses on the Y-3L and Y-5L were assessed for inconsistency using criteria established in previous studies comparing the adult EQ-5D versions<sup>4,21</sup>. A response pair was considered inconsistent if the Y-5L response was at least two levels away from the Y-3L response. One expected that reporting problems on the most severe level on the Y-3L (level 3) would redistribute to "a lot of problems" (level 4 Y-5L) or "cannot" (level 5 Y-5L). "Some problems" (level 2 Y-3L) might redistribute to "a little bit of a problem" (level 2 Y-5L); "some problems" (level 3 Y-5L) or "a lot of problems" (level

4 Y-5L). “No problems” (Level 1 Y-3L) would redistribute to “no problems” (level 1 Y-5L) or “a little bit of a problem” (level 2 Y-5L).

Discriminatory Power:

The Shannon Index (H') and the Shannon Evenness Index (J') were used to evaluate the discriminatory power of the Y-3L and Y-5L dimensions in terms of absolute and relative informativity<sup>4,22</sup>. The Shannon H' and J' indices are defined as follows:

$$H' = \sum_{i=1}^L p_i \log_2 p_i \quad \text{and} \quad J' = \frac{H'}{H'_{max}}$$

Where H' is the absolute amount of informativity, L is the number of dimensions levels and  $p_i$  is the proportion of observations in the  $i^{\text{th}}$  level where Y-3L has three levels and Y-5L has five levels. A higher H' index reflects that the descriptive system has captured more information, the maximum H' index is 1.58 and 2.32 on the Y-3L and Y-5L respectively. The Shannon Evenness index (J') reflects the spread of the responses across levels regardless of the number of levels included in the descriptive system.

Convergent Validity:

The convergent validity between Y-3L/Y-5L and VAS was assessed, using the VAS as reference. The Y-3L/Y-5L dimension scores were treated as cardinal values to calculate mean and standard deviations (SD) for comparison. Spearman rank order coefficients ( $r_s$ ) were calculated by dimension between the Y-3L/Y-5L and VAS. Furthermore, convergent validity between Y-3L and Y-5L was evaluated by dimension response-pairs, using Kendall tau B and Gamma statistic. Spearman's rank order and Pearson's correlations coefficients were interpreted according to Cohen: 0.1–0.29 low association, 0.3–0.49 moderate association and  $\geq 0.5$  high association<sup>23</sup>.

Known-Groups Validity:

Known-groups validity was tested for the dimensions of the Y-3L and Y-5L for health condition (acute, chronic, and general population) and medication (taking medication versus no medication) with Spearman rank order coefficients. The known-group validity across health condition was further assessed for the mean LSS and VAS score across the age groups by computing the Analysis of Variance (ANOVA). It was anticipated that those needing acute medical management would have the lowest VAS and LSS scores followed by those with stable chronic conditions and then the general population group.

Reliability:

Test-retest reliability was assessed using weighted Cohen's kappa (k) for dimension scores and the Intraclass Correlation Coefficient (ICC) for LSS and VAS scores. Kappa values were interpreted according to Landis and Koch's guidelines: <0.2 poor agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement<sup>24</sup>. An ICC of >0.7 was considered reliable<sup>25</sup>.

All data analyses were conducted using SPSS Windows 27.0 (IBM SPSS Inc., Chicago, IL, USA) and Statistica Windows Version 13.0 (TIBCO Software Inc., Palo Alto, CA, USA).

### 3. Results

A total of 146 children/adolescents needing acute medical/surgical management were eligible for recruitment, nine caregivers were uncontactable to obtain consent, one caregiver refused consent and six children refused assent. One of the 130 enrolled participants in the acute setting withdrew from the study and her data was not included, leaving 129 children in the acute health group. Research packs were sent home with 613 eligible participants at the schools for learners with special needs. A total of 249 packs were returned of which 51 were excluded [refused consent/assent (n=15), consented but no measures completed (n=17), measures completed without consent (n=2), only one measure completed (n=17)]. There were 198 children in the chronic health group. Research packs were similarly sent home with 932 children/adolescents attending mainstream schools. A total of 273 packs were returned of which 50 were excluded (refused consent (n=17), consented

but no measures completed (n=23), only one measure completed (n=10)], resulting in 223 children from the general population. The data of 550 children/adolescents has been included for analysis (Appendix Table 1).

The mean age of the children/adolescents was 11.3 years (SD 1.6) (Appendix Table 1). The children/adolescents with chronic conditions were older (mean 12.2 years; SD 1.9) than those in the acute medical setting (mean 11.6 years; SD 1.9) or from the general population (mean 11.3 years; SD 1.6). Gender of participants was equally distributed for the overall sample however those with chronic conditions had a higher proportion of males (n=123; 62%) whereas the general population had a higher proportion of females (133; 60%). Most of the children/adolescents needing acute medical management (n=129) were receiving orthopaedic management (n=83; 64%) whereas those with chronic conditions (n=198) had a physical disability (n=87; 44%) or learning difficulty (n=95; 48%). The children/adolescents in the general population group (n=223) had minor health conditions including asthma, eczema and allergies.

#### *General Instrument Performance and Feasibility:*

The Y-3L and Y-5L responses across condition groups and for the total sample are presented in Appendix Table 2. Across all dimensions the Y-5L had 4% missing responses and the Y-3L had 1%. The number of missing responses per dimensions was low ranging from 0-2% on the Y-5L and 0-1% on the Y-3L. The highest missing responses was for the dimension of *Usual Activities* on the Y-5L for the general population. There were however only three children/adolescents with missing responses on the Y-3L dimensions compared to 21 on the Y-5L. There was reduction in the reporting of no problems when moving from the Y-3L to the Y-5L across all dimensions. The greatest reduction was for *Mobility* with a 6% absolute reduction in total and 9% reduction for those with chronic illness. *Usual Activities* showed a 3% absolute total reduction with 5% and 4% in the chronic and general population groups respectively. *Pain or Discomfort* and *Worried, Sad or Unhappy* both showed an absolute reduction in the total sample of 4%. For *Pain or Discomfort* the greatest reduction was in the acute group (9%) whereas for *Worried, Sad or Unhappy* this was spread across groups (acute 4%, chronic 3% and general population 2%).

The proportion of participants reporting a ceiling effect with no problems in each dimension (11111) showed a 15% percentage reduction from the Y-3L to the Y-5L (Table 1). The ceiling effect was higher on the Y-3L and Y-5L for the 8-11-year group, however the percentage of reduction in ceiling effect moving from the Y-3L to the Y-5L was similar between age groups. There was a higher reduction in ceiling effect for those who were receiving acute medical management (30%) or chronic conditions (22%) compared to the general population (11%).

The report of floor effect with participants reporting the most severe problems in all dimensions (33333/55555) was <2% for the Y-3L and Y-5L. However, when considering the reporting of the most severe problems per dimensions a reduction in reporting when moving from the Y-3L to Y-5L in the groups with a health condition, is evident. For those receiving acute care the absolute reduction in reporting the most severe problems was modest but notable for *Pain and Discomfort* (6%), *Worried, Sad or Unhappy* (4%) and *Usual Activities* (4%). Although the absolute percentage reduction was lower for those with acute illness the reporting of the most severe problems reduced across all dimensions by 1-3%.



**Table 1.** Ceiling effect for the EQ-5D-Y-3L and EQ-5D-Y-5L across condition groups and for the total sample n=550.

		Ceiling Y-3L	Ceiling Y-5L	Absolute Reduction	Percentage reduction
	<b>n</b>	<b>11111</b>	<b>11111</b>		
Total	550	40%	34%	6%	15%
Condition group					
Acute	129	16%	11%	5%	30%
Chronic	198	29%	23%	7%	22%
General population	223	64%	57%	7%	11%
Age group					
8_11 years	305	44%	38%	6%	14%
12-15 years	245	36%	30%	6%	15%

*Redistribution Properties of the EQ-5D-Y-3L to the EQ-5D-Y-5L*

The dimension of *Usual Activities* and *Pain or Discomfort*, closely followed by *Worried, Sad or Unhappy*, had the highest number of inconsistent responses when moving from the Y-3L to the Y-5L. *Mobility* had the smallest number of inconsistent responses (Table 2).

**Table 2.** Cross tabulation for the EQ-5D-Y-3L and EQ-5D-Y-5L dimension scores.

Y-3L	Y-5L					Total inconsistent responses n (%)
MOB	no	a little bit	some	a lot	cannot	
No	346	34	<b>6</b>	<b>0</b>	<b>2</b>	23 (4%)
Some	<b>12</b>	47	17	1	<b>5</b>	
A lot	<b>1</b>	3	1	4	66	
LAM	no	a little bit	some	a lot	cannot	
No	409	18	<b>3</b>	<b>0</b>	<b>3</b>	35 (6%)
Some	<b>19</b>	44	15	2	<b>5</b>	
A lot	<b>0</b>	<b>3</b>	1	6	19	
UA	no	a little bit	some	a lot of	cannot	
No	334	24	<b>5</b>	<b>4</b>	<b>1</b>	43 (8%)
Some	<b>14</b>	48	25	8	<b>11</b>	
A lot	<b>3</b>	5	4	11	47	
P/D	no	a little bit	some	a lot	extreme	
No	315	40	<b>5</b>	<b>1</b>	<b>0</b>	42 (8%)
Some	27	81	48	8	<b>3</b>	
A lot	<b>1</b>	<b>5</b>	3	8	2	
WSU	no	a little bit	quite	really	extreme	
Not	318	36	<b>2</b>	<b>0</b>	<b>3</b>	41 (7%)
A bit	<b>29</b>	96	26	6	<b>2</b>	
Very	<b>1</b>	<b>4</b>	3	6	8	

n=550, bold and italicised indicates inconsistent responses. MOB: Mobility; LAM: Looking After Myself; UA: Doing Usual Activities; P/D: Having Pain or Discomfort; WSU: Feeling Worried, Sad or Unhappy

*Discriminatory Power*

Informativity of all dimensions improves on the Y-5L compared to the Y-3L with an average improvement of 0.094 with similar evenness (Table 3). *Mobility* and *Usual Activities* showed the greatest difference in spread of information between the Y-3L and Y-5L.

**Table 3.** Shannon Index (H') and Shannon Evenness Index (J') for the EQ-5D-Y-3L and EQ-5D-Y-5L dimensions.

	Y-5L		Y-3L		Difference in H'	Difference in J'
	H'	J'	H'	J'		
MOB	0.785	0.136	0.680	0.202	0.105	-0.066
LAM	0.722	0.171	0.646	0.236	0.076	-0.065
UA	0.771	0.155	0.652	0.221	0.119	-0.066
P/D	0.747	0.154	0.653	0.212	0.094	-0.058
WSU	0.747	0.109	0.669	0.077	0.078	0.032
Average difference					0.094	-0.045

n=550 LAM: Looking After Myself; UA: Doing Usual Activities; P/D: Having Pain or Discomfort; WSU: Feeling Worried, Sad or Unhappy

### Convergent Validity

The mean dimension scores for the Y-3L and Y-5L are presented in Table 5. The Spearman rank coefficients were similar across the Y-3L and Y-5L for condition groups and the whole sample, ranging from 0.18 (*Looking After Myself*) to 0.39 (*Pain or Discomfort*) (Table 4). Furthermore, the association between the Y-5L and Y-3L dimension scores for all respondents was high and significant with Spearman rank coefficients ranging between 0.83 (*Mobility*) and 0.73 (*Worried, Sad or Unhappy*) (not shown in Table 4).

The Y-3L Convergent validity between the paired Y-3L-Y-5L responses for each dimension was estimated with Gamma averaging 0.92, ranging from 0.92 (*Worried, Sad or Unhappy* and *Pain or Discomfort*) to 0.96 (*Mobility* and *Looking After Myself*). Kendall's tau B was similarly high with an average of 0.70 ranging from 0.69 (*Pain or Discomfort*) to 0.80 (*Mobility*).

**Table 4.** Convergent validity of the Y-3L, Y-5L and corresponding VAS mean values by dimension with Spearman rank order coefficients for Y-3L-VAS and Y-5L-VAS for each dimension, by condition group and for the sample.

		MOB	LAM	UA	P/D	WSU	VAS
Acute (n=129)	Y-5L	2.96 (1.79)	1.84 (1.34)	3.29 (1.74)	1.88 (1.03)	1.67 (0.978)	67.71 (28.25)
	r	-0.29*	-0.18*	-0.18*	-0.32*	-0.33*	
	Y-3L	2.02 (0.88)	1.47 (0.70)	2.18 (0.85)	1.53 (0.64)	1.45 (0.61)	71.24 (28.51)
	r	-0.17	-0.19*	-0.21*	-0.44*	-0.41*	
Chronic (n=198)	Y-5L	1.87 (1.30)	1.61 (1.14)	1.68 (0.97)	1.62 (0.87)	1.65 (0.99)	84.10 (19.75)
	r	-0.10	-0.07	-0.16*	-0.30*	-0.29*	
	Y-3L	1.47 (0.71)	1.37 (0.61)	1.41 (0.58)	1.42 (0.56)	1.45 (0.60)	83.91 (20.53)
	r	-0.12	-0.07	-0.19*	-0.35*	-0.33*	
General Population (n=223)	Y-5L	1.09 (0.31)	1.05 (0.21)	1.10 (0.34)	1.33 (0.67)	1.34 (0.68)	89.12 (14.53)
	r	-0.21*	-0.18*	-0.22*	-0.40*	-0.34*	
	Y-3L	1.04 (0.20)	1.04 (0.20)	1.07 (0.27)	1.23 (0.44)	1.26 (0.48)	89.77 (14.13)
	r	-0.18*	-0.18*	-0.26*	-0.32*	-0.38*	
Total (n=550)	Y-5L	1.81 (1.39)	1.01 (1.44)	1.83 (1.35)	1.56 (0.87)	1.53 (0.89)	82.24 (21.99)
	r	-0.29*	-0.18*	-0.34*	-0.39*	-0.32*	
	Y-3L	1.43 (0.72)	1.26 (0.55)	1.45 (0.71)	1.37 (0.55)	1.38 (0.57)	83.15 (21.91)
	r	-0.26*	-0.18*	-0.32*	-0.39*	-0.37*	

n=550. Dimension scores treated as numeric data and presented as mean (SD).  $r_s$  denotes Spearman rank correlation co-efficient, negative values indicate that as problems in a dimension increases in severity it decreases the VAS score, \* $p < 0.05$  (2-tailed). MOB: Mobility; LAM: Looking After Myself; UA: Doing Usual Activities; P/D: Having Pain or Discomfort; WSU: Feeling Worried, Sad or Unhappy, VAS: Visual Analogue Scale

### Known-Groups Validity

The hypotheses for known-group validity were confirmed and presented in Table 5. The Y-3L and Y-5L correlations with health condition are all significant and show an increase of reported problems with health condition (acute condition > chronic condition > general population). This was stronger for dimensions of *Mobility*, *Looking After Myself* and

*Usual Activities*. A similar pattern was found in that those who reported taking medication showed an increase in reporting of problems. This was moderate for the dimension of *Usual Activities* and weak and significant for the other dimensions except for *Worried, Sad or Unhappy*. Furthermore, ANOVA with Tukey's post hoc analysis shows that VAS scores and LSS are significantly different between acute, chronic and general population for the Y-5L and Y-3L ( $p < 0.001$ ) (Appendix Table 2).

**Table 5.** Spearman rank correlation of EQ-5D-Y-3L and EQ-5D-Y-5L dimension scores with known health condition and medication.

	Health Condition		Medication	
	Y-5L	Y-3L	Y-5L	Y-3L
MOB	-0.53**	-0.52**	-0.22**	-0.25**
LAM	-0.35**	-0.34**	-0.22**	-0.21**
UA	-0.57**	-0.57**	-0.30**	-0.27**
P/D	-0.26**	-0.22**	-0.29**	-0.19**
WSU	-0.17**	-0.14**	-0.14**	-0.08

$n=550$ , \* $p < 0.05$  \*\* $p < 0.001$  (2-tailed); MOB: Mobility; LAM: Looking After Myself; UA: Doing Usual Activities; P/D: Having Pain or Discomfort; WSU: Feeling Worried, Sad or Unhappy

#### Reliability

Second measures were completed for 79/198 with a stable chronic condition and 94/223 from the general population. The median time between baseline and retest was 9 days (IQR 7-13). The Y-5L and Y-3L showed moderate to substantial agreement across dimensions, VAS and LSS scores for those with a stable chronic condition (Table 6). For the general population group the Y-5L and Y-3L showed fair agreement for dimensions of *Looking After Myself* and *Usual Activities*, the other dimensions showed moderate agreement.

**Table 6.** Test-retest reliability for EQ-5D-Y dimension scores and summary scores for those with a stable chronic condition and the general population.

		Y-5L			Y-3L		
		Total (n=173)	Chronic (n=79)	GP (n=94)	Total (n=173)	Chronic (n=79)	GP (n=94)
MOB	k	0.70**	0.67**	0.54**	0.54**	0.47**	0.55**
LAM	k	0.71**	0.73**	0.32**	0.47**	0.43**	0.38**
UA	k	0.45**	0.45**	0.29**	0.44**	0.42**	0.21**
P/D	k	0.53**	0.54**	0.43**	0.51**	0.42**	0.59**
WSU	k	0.46**	0.40**	0.53**	0.45**	0.41**	0.48**
VAS	ICC <sup>1</sup>	0.64**	0.66**	0.59**	0.82**	0.82**	0.77**
LSS	ICC <sup>1</sup>	0.78**	0.67**	0.75**	0.58**	0.55**	0.34**

\*\* $p < 0.001$ , <sup>1</sup>All Intraclass Class correlation coefficients were two-way random, absolute agreement, k weighted Cohen's Cappa single measures, MOB: Mobility; LAM: Looking After Myself; UA: Doing Usual Activities; P/D: Having Pain or Discomfort; WSU: Feeling Worried, Sad or Unhappy; GP: general population; VAS: Visual Analogue Scale; LSS: Level Sum Score

#### 4. Discussion

This was the first study to compare the English version of the Y-3L and Y-5L in a large sample of children/adolescents with different health conditions and healthy children, in the recommended age range of 8-15 years <sup>26</sup>. Children/adolescents receiving acute medical/surgical management and with a stable chronic condition were considered a suitable population for comparison of the Y-3L and Y-5L as there was a spread of disease severity, and consequently a spread of scores. Furthermore, the Y-3L was shown to have good psychometric properties in a similar group of respondents <sup>27</sup>.



The number of missing responses was low and acceptable for the EQ-5D-Y dimensions albeit lower for the Y-3L (1%) than the Y-5L (4%). The mean VAS scores were not significantly different between the Y-3L and Y-5L VAS however, there was a higher number of missing responses on the Y-3L VAS (5%) versus Y-5L VAS (1%). It is postulated that this is due to the order effect as children/adolescents found the second VAS redundant. However, this did not seem to contribute to the missing responses on the Y-3L dimensions as it had fewer missing responses than the Y-5L.

Although the adult five level version, EQ-5D-5L, showed reduced ceiling effects<sup>20,21,28</sup> it was not clear that this would be the case in children as a higher number of items can potentially increase the difficulty of completion with younger children being more likely to select extreme scores on a Likert scale<sup>29</sup>. Although the ceiling effect was higher for children 8-11 years than those 12-15 years this can be attributed to the higher number of general population children in the younger age-group (65% vs 35%) as there was marginal difference in the relative reduction in ceiling effect between the age groups. This likely indicates the Y-5L levels show clear distinction between the levels and can be attributed to the rigorous methodology in the development of the labels<sup>8</sup> ensuring younger children are able to interpret the severity of the qualifiers<sup>30</sup>.

Wong et al (2019) reported a relative reduction in ceiling effect (11111) of 4.2% which is considerably lower than the 15% found in this study. The change in reporting of no problems per dimension was similar in those with a stable chronic illness in this study (1% for *Looking After Myself* and 9% for *Mobility*) to those with idiopathic scoliosis in China (0% *Looking After Myself* and 7.8% for *Worried, Sad or Unhappy*)<sup>9</sup>. However, the relative reduction for these two groups was higher in the South African sample (22%) and may be due to the homogenous selection of respondents in this sample. The greatest reduction was however in those requiring admission to an acute hospital with an absolute decrease of between 0-9% (*Usual Activities* and *Pain or Discomfort*). The reduction in ceiling effect and report of no problems in those with a health condition indicates that the Y-5L may be more accurate in describing health and have greater sensitivity in detecting change in health in patient groups.

The redistribution of scores from the Y-3L to the Y-5L showed few inconsistencies across dimensions (4-8%) and was similar to the comparison of versions in China<sup>9,10</sup>. The discriminatory power of the dimensions improved when moving from the Y-3L to the Y-5L with retention of evenness of the distribution of the responses. The distribution of responses is however not as high as that observed in the adult 3L and 5L measures<sup>4,21,28</sup> with a lower selection of some problems on the Y-3L and some/a lot of problems on the Y-5L. The redistribution scores in this study does however support the validity of the new Y-5L instrument. The convergent validity was not systematically different for Y-3L and Y-5L dimensions scores, using the VAS as a reference, across dimensions or condition groups. The correlation of the paired Y-3L and Y-5L dimension scores was high further supporting convergent validity. Convergent validity of the Y-3L and Y-5L with the PedsQL is not within the scope of this paper and will be described separately in a different paper.

The validity of the Y-5L is further supported by the ability to detect differences between known-groups with differing health conditions, despite these known-groups having heterogeneous conditions.

Despite the Y-5L being able to capture smaller changes in health status with the increased number of dimensions compared to the Y-3L the test-retest reliability was higher across dimensions and for the LSS for the Y-5L. As anticipated the reliability was higher for both measures for dimensions of physical health (*Mobility* and *Looking After Myself*) than *Pain or Discomfort* and *Worried, Sad or Unhappy*. Surprisingly, the reliability of the *Usual Activities* dimension was lower on both the Y-3L and Y-5L than reported in the study in children/adolescents with haematological malignancies<sup>31</sup>. It is postulated that this may have been affected by the covid pandemic and subsequent impact on *Usual Activities*<sup>32</sup>.

Due to the limitations of the Covid pandemic on recruitment of children/adolescents at schools there may be non-response bias. Although children/adolescents were explicitly

instructed to complete the measures on their own without influence from others there was no way to ensure this in the sample with a chronic condition and general population. Another limitation was a potential ordering effect as the order of versions was not randomized, it was always Y-5L first and then Y-3L.

## 5. Conclusion

The English version of the EQ-5D-Y-5L is a valid and reliable extension of the Y-3L for children/adolescents across health conditions and healthy general population children. The expanded levels reduce the ceiling effect, most notably for those with a health condition. The relative informativity of report across the dimensions increases on the Y-5L compared to the Y-3L with retention of evenness of reporting. The convergent validity and test-retest reliability of the Y-5L was comparable to the Y-3L. The reduced ceiling effect and increased discriminatory power of the Y-5L compared to the Y-3L may result in greater sensitivity and responsiveness in measuring health status in specific disease groups and warrants further testing.

**Funding:** Janine Verstraete, Razia Amien and Des Scott reported receiving funding from the EuroQoL Research Foundation (project EQ20180730) to conduct the study. Janine Verstraete and Des Scott are both members of the EuroQoL Research Group.

**Precis:** The newly developed EQ-5D-Y Five-Level Version is a valid and reliable HRQoL outcome in children/adolescents with a variety of health conditions and from the general population.

## References

1. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. *Quality of Life Research*. 2010;19(6):875-886. doi:10.1007/s11136-010-9648-y
2. EuroQol Research Foundation. EQ-5D-Y User Guide. *EuroQol Research Foundation* 2020. 2020;(September):1-20.
3. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x
4. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*. 2008;11(2):275-284. doi:10.1111/j.1524-4733.2007.00230.x
5. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research*. 2013;22(7):1717-1727. doi:10.1007/s11136-012-0322-4
6. Agborsangaya CB, Lahtinen M, Cooke T, Johnson JA. Comparing the EQ-5D 3L and 5L: Measurement properties and association with chronic conditions and multimorbidity in the general population. *Health and Quality of Life Outcomes*. 2014;12(1):1-7. doi:10.1186/1477-7525-12-74
7. Craig BM, Pickard AS, Lubetkin EI. Health problems are more common, but less severe when measured using newer EQ-5D versions. *Journal of Clinical Epidemiology*. 2014;67(1):93-99. doi:10.1016/j.jclinepi.2013.07.011
8. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Quality of Life Research*. 2019;0(0):0. doi:10.1007/s11136-019-02115-x
9. 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. *European Journal of Health Economics*. 2019;20(5):647-656. doi:10.1007/s10198-018-1026-7
10. Pei W, Yue S, Zhi-Hao Y, Ruo-Yu Z, Bin W, Nan L. Testing measurement properties of two EQ-5D youth versions and KIDSCREEN-10 in China. *European Journal of Health Economics*. 2021;(0123456789). doi:10.1007/s10198-021-01307-y
11. Åström M, Åström M, Åström M, et al. EQ-5D-Y-5L as a patient-reported outcome measure in psychiatric inpatient care for children and adolescents - A cross-sectional study. *Health and Quality of Life Outcomes*. 2020;18(1):1-14. doi:10.1186/s12955-020-01366-4
12. Jones R, Mulhern B, McGregor K, et al. Psychometric Performance of HRQoL Measures: An. *Children*. 2021;8(714).
13. Prevotnik Rupel V, Ogorevc M, Greiner W, Kreimeier S, Ludwig K, Ramos-Goni JM. EQ-5D-Y Value Set for Slovenia. *PharmacoEconomics*. 2021;39(4):463-471. doi:10.1007/s40273-020-00994-4
14. Shiroya T, Ikeda S, Noto S, Fukuda T, Stolk E. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. *Medical Decision Making*. 2021;41(5):597-606. doi:10.1177/0272989X211001859
15. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics (United Kingdom)*. 2018;27(1):7-22. doi:10.1002/hec.3564
16. Lamers L., McDonnell J, Stalmeier PF., Krabbe PF., Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economic*. 2006;15:1121-1132. doi:10.1002/hec

17. Parkin D, Rice N, Devlin N. Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? *Medical decision making : an international journal of the Society for Medical Decision Making*. 2010;30(5):556-565. doi:10.1177/0272989X09357473
18. World Medical Association. World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. *Journal of the American Medical Association*. 2013;310(29):2191-2194. doi:10.1001/jama.292.11.1359
19. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology*. 2003;56(8):730-735. doi:10.1016/S0895-4356(03)00084-2
20. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*. 2008;11(2):275-284. doi:10.1111/j.1524-4733.2007.00230.x
21. Pickard AS, Leon MC De, Kohlmann T, et al. Psychometric Comparison of the Standard EQ-5D to a 5 Level Version in Cancer Patients Linked references are available on JSTOR for this article : Psychom 5 Level Version in Cancer Patients. *Medical Care*. 2007;45(3):259-263.
22. Bas Janssen MF, Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*. 2007;16(5):895-904. doi:10.1007/s11136-006-9160-6
23. Cohen S, Percival A. Prolonged Peritoneal Dialysis in Patients Awaiting Renal Transplantation. *British Medical Journal*. 1968;1:409-413.
24. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
25. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
26. Reenen van M, Janssen B, Oppe M, Kreimeier S, Greiner W. EQ-5D-Y User Guide, basic information on how to use the EQ-5D-Y instrument. 2014;(August):0889-8553.
27. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape , South Africa : psychometric properties , feasibility and usefulness - a longitudinal , analytical study. *Health and Quality of Life Outcomes*. 2017;1-14. doi:10.1186/s12955-017-0590-3
28. Janssen M, Pickard A, Golicki A, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*. 2013;22(7):1717-1727.
29. De Civita M, Regier D, Alamgir AH, Anis AH, Fitzgerald MJ, Marra CA. Evaluating Health-Related Quality-of-Life Studies in Paediatric Populations Some Conceptual , Methodological and Developmental Considerations and Recent Applications. *Pharma*. 2005;23(7):659-685.
30. Derrett S, Herdman M, Ngwira LG, Moore EY, Jelsma J. A New Approach to Assessing Children's Interpretation of Severity Qualifiers in a Multi-Attribute Utility Instrument-The EQ-5D-Y-5L: Development and Testing. *Patient*. 2021;(0123456789). doi:10.1007/s40271-021-00496-1
31. Zhou W, Shen A, Yang Z, et al. Patient-caregiver agreement and test-retest reliability of the EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients with haematological malignancies. *European Journal of Health Economics*. 2021;(0123456789). doi:10.1007/s10198-021-01309-w
32. Holt L, Murray L. Children and Covid 19 in the UK. *Children's Geographies*. 2021;0(0):1-8. doi:10.1080/14733285.2021.1921699