



ScienceDirect

Contents lists available at [sciencedirect.com](http://sciencedirect.com)  
Journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

## Patient-Reported Outcomes

# Comparison of the Psychometric Properties of the EQ-5D-3L-Y and EQ-5D-5L-Y Instruments in Spanish Children and Adolescents



Miguel Ángel Pérez-Sousa, PhD, Pedro R. Olivares, PhD, Robinson Ramírez-Vélez, PhD, Narcis Gusi, PhD

## ABSTRACT

**Objectives:** The study aimed to assess the reliability and validity of EQ-5D-5L-Y and to compare the performance of EQ-5D-5L-Y with EQ-5D-3L-Y in children and adolescents.

**Methods:** The Spanish versions of the 3L and 5L of EQ-5D for youths, were administered to children and adolescents from the general population. Feasibility and reliability were determined for the EQ-5D-5L-Y. The EQ-5D-5L-Y and EQ-5D-3L-Y were evaluated in terms of ceiling effects, informativity, and correlations with other generic measurements of health-related quality of life.

**Results:** A total of 714 healthy children and adolescents ( $10.7 \pm 2.1$  years old) from the general population participated in the study. Most of the sample reported full health status. The feasibility and reliability for the EQ-5D-5L-Y were acceptable, but the questionnaire showed a low convergent validity. Absolute informativity (Shannon index) showed a slight increase in all dimensions of the 5L compared with the 3L; nevertheless, there were only statistically significant differences between 5L and 3L in the dimension “feeling worried, sad, or unhappy” and also on the overall system. Relative informativity (Shannon evenness index) showed a decrease in the 5L compared with 3L for all dimensions, except for “looking after myself.” Correlations with other health measurements, in both 3L and 5L, showed similar results to those observed in the international EQ-5D-3L-Y validation study.

**Conclusion:** The results show that EQ-5D-5L-Y is feasible, consistent, and reliable, but there are minor differences in the ceiling effect and informativity between the EQ-5D-5L-Y and EQ-5D-3L-Y versions in the general population.

**Keywords:** ceiling effect, health-related quality of life, informativity, reliability, validity.

VALUE HEALTH. 2021; 24(12):1799–1806

## Introduction

The EQ-5D-3L-Y is a widely used instrument to measure general health-related quality of life (HRQOL) status in children and adolescents. The questionnaire was developed following the standard EQ-5D-3L descriptive system for adults. The EQ-5D-3L-Y contains 5 dimensions of health: “mobility,” “looking after myself,” “doing usual activities,” “having pain or discomfort,” and “feeling worried, sad, or unhappy.” Similar to the adult version, it has 3 levels of response, which indicate the severity of health problems in the participant, containing 243 possible health states. Studies using the EQ-5D-3L-Y have shown a large ceiling effect and a limited ability to detect moderate impairments of HRQOL and also a limited capability to discriminate between respondents of the general population.<sup>1</sup>

The EuroQol Research Foundation, recognizing these issues, developed a new version that retains the same dimensions as the 3-level version, the EQ-5D-3L, but has 5 levels for each dimension, and this is referred to as the EQ-5D-5L.<sup>2</sup> The recent version was

developed in several languages in a multicenter study, including German, Spanish, Swedish, and English.

Previously, EQ-5D-3L and EQ-5D-5L for adults were compared by several studies, which found that the EQ-5D-5L system had more informativity, discriminatory power, and a decreased ceiling effect in both the general and patient populations.<sup>3,4</sup>

Like the EQ-5D-3L-Y, the EQ-5D-5L-Y version needs to be explored in Spanish children and adolescents from the general population. Understanding the differences in psychometric properties between EQ-5D-3L-Y and EQ-5D-5L-Y is key to good clinical use and decision making in health policies. Therefore, it is necessary to examine the psychometric properties of the instrument in terms of feasibility, ceiling effect, redistribution properties, discriminatory power, reliability, and validity.

The exploration of the psychometric properties of the new EQ-5D-5L-Y (published by the EuroQol Research Foundation) and its comparison with the EQ-5D-3L-Y test is essential for the validity of future studies.

This study aimed to examine the validity and reliability of the EQ-5D-5L-Y and perform a comparative evaluation of the psychometric properties of the EQ-5D-3L-Y and EQ-5D-5L-Y instruments, in terms of feasibility, ceiling effect, redistribution properties, inconsistency, and informativity in the general population.

## Methods

### Participants

To select the sample, 6 primary education schools and 6 secondary education schools from the Andalucía community were contacted. The research protocol was first presented to the school's administration and the parents' council. Finally, 4 centers (3 from primary and 1 from secondary school) agreed to participate in the study. Andalucía has a total of 1 400 000 children and adolescents between the ages of 8 and 17 years. Taking into account the universe of the study, a margin of error of 5%, and a confidence interval of 95%, the target population should be 385 subjects. Nevertheless, 950 students were invited to participate to have greater heterogeneity in terms of health conditions. Finally, 714 children and adolescents between the ages of 8 and 17 years completed the survey.

Before data collection, the parents were informed of the study's methodology and objectives through an official letter written by the researchers that included an informed consent form. To be included in the study, students had to provide the informed consent form signed by a parent or guardian, attend school on the day the test was administered, and have a good enough level of Spanish to respond in full to the set of questionnaires.

### Instruments

All participants completed a paper-based survey, which included sociodemographic questions and several instruments for the measurement of HRQOL in children and adolescents, including the Spanish version of the EQ-5D-Y (with 5L and 3L levels of severity), the EQ visual analog scale (VAS) for young people, KIDSCREEN-27, and the Cantril's Life Satisfaction Ladder.

#### EQ-5D-5L-Y and EQ-5D-3L-Y

The first version of the self-administered Spanish version of the EQ-5D-Y with 3 levels of response<sup>1</sup> and the new version with 5 levels of response<sup>2</sup> comprised 5 descriptive systems referring to "mobility," "looking after myself," "doing usual activities," "having pain or discomfort," and "feeling worried, sad, or unhappy." The respondent can report 3 or 5 levels of severity for each dimension. A health state is composed by taking 1 level for each dimension; for example, for the EQ-5D-5L-Y, a total of 3125 possible health states are obtained ranging from 11111 (full health) to 55555 (worst health state), and for the EQ-5D-3L-Y, 243 different health state, which range between 11111 (full health) to 33333 (worst health state).

#### EQ-VAS

Both 5L and 3L versions of EQ-5D-Y include a VAS, where the interviewee can report on their health status "today," in a range of scores from 0 to 100, where 0 indicates the worst health status and 100 represents the best health status.

#### KIDSCREEN-27

KIDSCREEN-27 is a generic self-report measure used in both healthy and ill children and adolescents at the age of 8 to 18 years.

It is a reliable, sensitive, and cross-validated measure in 38 languages.<sup>5</sup> The questionnaire contains 5 dimensions: physical well-being (physical), psychological well-being (psychological), autonomy and parent relations (autonomy/parents), social support and peers (friends), and school environment (school). We also included the HRQOL index, which consists of 10 items from all dimensions.

#### Cantril's Life Satisfaction Ladder

This ladder<sup>6</sup> evaluates the general subjective life satisfaction by asking respondents to picture the best and worst possible lives. These 2 extreme values are presented on a ladder at 0 (the worst) and 10 (the best). Respondents have to indicate where they feel they are standing between these 2 extremes. This tool has been used in World Health Organization surveys in children and adolescents.

#### Sociodemographic measures

Information about age, sex, and school grade was collected at the time of data collection.

### Study Design

The set of questionnaires were administered in the classrooms, in small groups of 8 to 10 students, by a technician with experience in these studies. The type of interview was by direct administration of the questionnaires, where the students had to respond individually to the items, after an explanation by the technician of the procedure and steps to be followed. The order of questionnaires in the set was aleatory for both methods of survey administration. The duration of the survey varied depending on the respondents' age, from 30 minutes in children at the age of 8 to 12 years of age to 20 minutes in students at the age of 15 to 17 years.

For the test-retest reliability objective, we selected 100 children (10 by age, ranging from 8 to 17) who should complete the same set of questionnaires between 7 and 10 days after its first completion.

For confidentiality and to facilitate data analysis, each respondent was assigned a code. A phone number and email address were provided to respondents to address any concerns that may arise.

The study was approved by the Bioethics Committee of Universidad de Extremadura and was conducted following the International Ethical Guidelines for Biomedical Research Involving Human Subjects, established in Geneva.

### Analysis

All data are presented as mean and standard deviation for continuous variables and frequencies and percentages for categorical variables. Feasibility was examined by calculating the number of missing values for 3L and 5L versions. The ceiling effect of the EQ-5D was defined as the proportion of "no problem" responses in each dimension and in all dimensions. A reduction in ceiling effect suggests enhanced classification efficiency. We examined the absolute reduction, calculated as the difference between the proportions of ceiling effect in both systems. The relative reduction is calculated with the following formula<sup>7</sup>:

$$\frac{\text{ceiling } 3L - \text{ceiling } 5L}{\text{ceiling } 3L} \times 100.$$

#### Redistribution Properties From 3L to 5L

Redistribution properties and (in)consistency of responses were evaluated using the method applied in previous studies,<sup>3,7</sup>

which were described as proportions of the 3L-5L response pairs within each 3L response level (ie, 3L-1, 3L-2, and 3L-3). An inconsistent response pair was defined as a 3L response that is at least 2 levels away from the 5L response (eg, a child chose level 1 [no problem] in the 3L but responded 3 [moderate problems] in the 5L); the other pairs were regarded as consistent. The size of inconsistency was calculated as  $|3L-5L| - 1$ , after recoding the EQ-5D-3L responses on the EQ-5D-5L scale (1 = 1; 2 = 3, 3 = 5). We calculated the proportion of each consistent pair in each 3L response level and the percentage of inconsistent pairs in each dimension, in addition to their corresponding mean and median VAS values. Our hypothesis was to find the decrease of mean and median VAS values, when moving to lower status pairs in each dimension<sup>8</sup>; the linear trend was examined through the nonparametric Jonckheere trend test.

### Informativity

Informativity power determines the degree of uniform distribution of responses in each dimension. The more evenly the answers are distributed, the more useful the questionnaire is. We used the Shannon index ( $H'$ ) and the Shannon evenness index ( $J'$ ) of informativity to compare the discriminatory power of the 3L and 5L versions, by dimension and overall system. Shannon's methodology and indices, originally from the information theory, have been applied to classification and health state mainly for the EQ-5D.<sup>4,7,9</sup> The Shannon index is defined as follows:

$$H' = - \sum_{i=1}^L p_i \log_2 p_i$$

where  $H'$  represents the absolute amount of informativity captured,  $L$  is the number of possible levels, and  $p_i$  is the level of responses in the level  $i$ th. The higher the  $H'$  is, the more the information is captured by the system. Informativity is dependent on the number of response options and the distribution of the observations across levels. In the case of an even (or rectangular) distribution, that is, if all levels are equally filled, the optimal amount of information is captured, and the Shannon index has reached its upper limit ( $H'_{max}$ ), which is presented by the following formula:

$$H'_{max} = \log_2 C.$$

For example,  $H'_{max}$  for the 5L is  $\log_2 5$  or 2.32 and for the 3L is  $\log_2 3$  or 1.58. If the number of levels is increased,  $H'_{max}$  increases accordingly.  $J'$  is constrained between 0 and 1. The less evenness in the responses, the lower  $J'$  is, and vice versa.  $J'$  is calculated as  $J' = H'/H'_{max}$ , indicating the usage of the system ( $H'$ ), given its inherent capacity ( $H'_{max}$ ). The 95% confidence intervals for  $H'$  were estimated using a nonparametric bootstrap method. Our hypothesis was that the 5L version has more discriminatory power (larger  $H'$  values) than the 3L version. In contrast, Shannon Evenness index  $J'$ , reflecting that populations need a larger spread to cover 5 levels than for 3. Therefore, we expected the  $H'$  to increase (higher absolute levels of information) and  $J'$  to stay equal or marginally decrease in the 5L version.

### Convergent Validity

The convergent validity was analyzed through the Spearman rank correlation coefficient between the dimension of 3L and 5L versions and previously validated measures of the child HRQOL. Following the guidelines provided by Cohen et al,<sup>10</sup> coefficients from 0.1 to 0.29 were deemed to be low, 0.3 to 0.49 as moderate, and correlations of 0.5 or above as high. In line with previous

studies,<sup>1,11</sup> it was hypothesized that "mobility" and "pain/discomfort" dimensions would show a moderate correlation with the dimensions of physical well-being of the KIDSCREEN-27 and a significant association with the rest of the HRQOL measures used. We also hypothesized that the anxiety or depression dimension of the EQ-5D-5L-Y would show a moderate to high correlation with the psychological well-being dimension of the KIDSCREEN-27 and a moderate to high correlation with the VAS and the global HRQOL index of the KIDSCREEN-10, the general health item of the KIDSCREEN-27, and the Cantril's Life Satisfaction Ladder.

### Reliability

We assessed the reliability of the EQ-5D-5L-Y by calculating percentages of agreement and kappa coefficients<sup>12</sup> to estimate test-retest concordance. We interpreted the kappa values according to Landis and Koch's guidelines<sup>13</sup> with kappa < 0.2 indicating poor agreement, 0.21-0.40 indicating fair agreement, 0.41-0.60 indicating moderate agreement, 0.61-0.80 indicating substantial agreement, and kappa > 0.81 indicating almost perfect agreement. We calculated the intraclass correlation coefficient<sup>14</sup> for the VAS, and an intraclass correlation coefficient of >0.7 was considered acceptable.

### Results

A total of 714 children ( $10.7 \pm 2.1$  years old) from the general population participated in the study. By sexes, 354 (49.6%) were females and 360 (50.4%) were males. The frequencies and percentages of reported problems by the sample on the EQ-5D-5L-Y and EQ-5D-3L-Y are presented in Table 1. The dimensions "mobility" (93.6%), "looking after myself" (95.2%), and "doing usual activities" (92.4%) concentrated most percentages in the first level of response; "no problems," "having pain or discomfort," and "feeling worried, sad, or unhappy" showed a greater dispersion of the responses, with 15.5% and 16.0% of problems presenting in the second level "A little bit of..." The missing data in the new EQ-5D-5L-Y ranged between 0.1% and 0.4%, so the feasibility was acceptable.

The number of reported problems in all dimensions from the third to the fifth level of response was negligible. It can be observed that the percentage in the first and second level of response is very similar to that reported in the EQ-5D-5L-Y.

In Table 2, frequencies and percentages of ceiling effect are shown in absolute and relative terms. Both 3L and 5L versions of the questionnaire present a similar ceiling effect, although it is slightly lower in the 5L. The reduction of a ceiling effect in relative terms is between 4.8% for "feeling worried, sad, or unhappy" and 0% for "doing usual activities"; thus, we cannot observe a plausible reduction from 3L to 5L.

The proportion of consistent and inconsistent responses for each dimension is presented in Table 3. In the redistribution from 3L to 5L, the response of most participants was consistent, because no values differed from 3L to 5L in more than 2 levels of response. The sum of inconsistency by dimensions, in number and percentage, was highest for "having pain or discomfort" (7.0%) and lowest for "mobility" (1.5%). In addition, we have shown the mean and median VAS values for each of the 3L-5L response pairs. All mean and median VAS scores show reduced values when comparing the healthiest subgroup (g.1.1) with the most disabled subgroup (g.3.5). To check the linear trend, we used the nonparametric Jonckheere trend test. The results in all dimensions were statistically significant ( $P < .001$ ).

The informativity results of 3L and 5L are presented in Table 4. In terms of absolute informativity (Shannon  $H'$ ), the 5L presented

**Table 1.** Percentages of reported problems in the EQ-5D-5L-Y and EQ-5D-3L-Y.

EQ-5D-5L-Y		EQ-5D-3L-Y	
<b>Mobility (walking about)</b>			
No problems	668 (93.6)	No problems	678 (95.0)
A little bit of a problem	44 (6.1)	Some problems	34 (4.8)
Some problems	1 (0.1)	A lot of problems	0 (0.0)
A lot of problems	0 (0.0)	Missing	2 (0.3)
Cannot	0 (0.0)		
Missing	1 (0.1)		
<b>Looking after myself</b>			
No problems	680 (95.2)	No problems	695 (97.3)
A little bit of a problem	32 (4.5)	Some problems	15 (2.1)
Some problems	0 (0.0)	A lot of problems	3 (0.4)
A lot of problems	0 (0.0)	Missing	1 (0.1)
Cannot	0 (0.0)		
Missing	2 (0.3)		
<b>Doing usual activities</b>			
No	660 (92.4)	No problems	660 (92.4)
A little bit of a problem	43 (6.0)	Some problems	52 (7.3)
Some problems	10 (1.4)	A lot of problems	1 (0.1)
A lot of problems	1 (0.1)	Missing	1 (0.1)
Cannot	0 (0.0)		
Missing	0 (0.0)		
<b>Having pain or discomfort</b>			
No pain or discomfort	660 (81.0)	No pain or discomfort	588 (82.4)
A little bit of pain or discomfort	107 (15.0)	Some pain or discomfort	106 (14.8)
Some pain or discomfort	24 (3.4)	A lot of pain or discomfort	17 (2.4)
A lot of pain or discomfort	0 (0.0)	Missing	3 (0.4)
Extreme pain or discomfort	3 (0.4)		
Missing	2 (0.3)		
<b>Feeling worried, sad, or unhappy</b>			
Not worried, sad, or unhappy	580 (81.2)	Not worried, sad, or unhappy	609 (85.3)
A little bit worried, sad, or unhappy	114 (16.0)	A bit worried, sad, or unhappy	94 (13.2)
Quite worried, sad, or unhappy	5 (0.7)	Very worried, sad, or unhappy	10 (1.4)
Really worried, sad, or unhappy	11 (1.5)	Missing	1 (0.1)
Extremely worried, sad, or unhappy	1 (0.1)		
Missing	3 (0.4)		

Note. Data are expressed as frequencies and percentages, n (%).

more informativity than 3L with an increase from 3L to 5L in all dimensions. Nevertheless, for “having pain or discomfort,” this increase was only 3.8%. Relative informativity in terms of Shannon evenness index ( $J'$ ) was similar to absolute informativity. Nevertheless, the change from 3L to 5L is reduced in “mobility,” “doing usual activities,” “having pain or discomfort,” and “feeling worried, sad, or unhappy.” In contrast, there was an increase of informativity in “looking after myself.”

The convergent validity in 3L and 5L is presented in Table 5. Most dimensions had a statistically significant correlation with KIDSCREEN-27 and Cantril’s Life Satisfaction Ladder; nevertheless, the magnitude of this correlation was low. In contrast, the results presented a moderate or high agreement between the 3L and 5L versions.

Finally, of the 100 children and adolescents allocated in the test-retest reliability group, 89 participated ( $11.6 \pm 3.1$ ). The kappa coefficient,  $P$  value, and agreement percentage were as follows: “mobility” (kappa = 1, agreement = 100%), “looking after myself” (kappa = 1, agreement = 100.0%), “doing usual activities” (kappa = 0.66,  $P < .001$ , agreement = 98.5%), “having pain or discomfort” (kappa = 0.42,  $P < .001$ , agreement = 86.7%), and “feeling worried, sad, or unhappy” (kappa = 0.58,  $P < .001$ , agreement = 86.7%).

## Discussion

This study aimed to examine the validity and reliability of the EQ-5D-5L-Y and compare its psychometric properties in terms of

**Table 2.** Proportions of “no problems” responses for the EQ-5D-3L-Y and EQ-5D-5L-Y systems and ceiling effect change.

Dimension	3L*, n (%)	5L <sup>†</sup> , n (%)	Ceiling effect reduction	
			Absolute <sup>‡</sup> (%)	Relative <sup>‡</sup> (%)
Mobility (walking about)	678 (95.0)	668 (93.6)	−1.4	−1.5
Looking after myself	695 (97.3)	680 (95.2)	−2.1	−2.2
Doing usual activities	660 (92.4)	660 (92.4)	0.0	0.0
Having pain or discomfort	588 (82.4)	660 (81.0)	−1.4	−1.6
Feeling worried, sad, or unhappy	609 (85.3)	580 (81.2)	−4.1	−4.8
Full health (11111)	491 (68.8)	472 (66.1)	−2.7	−3.9

\*EQ-5D-3L-Y.

†EQ-5D-5L-Y.

‡Ceiling effect reduction is expressed in absolute (ceiling 3L – ceiling 5L) and relative terms (ceiling 3L – ceiling 5L / ceiling 3L × 100).

feasibility, ceiling effect, redistribution properties, inconsistency, and informativity.

The feasibility of 3L and 5L was very high; the results showed a lower percentage of missing or inappropriate responses than previous studies.<sup>1,15</sup> The prevalence of severe problems reported in 3L and 5L versions was very low, which is common for general population samples. The highest proportion of problems was reported on the “having pain or discomfort” and “feeling worried, sad, or unhappy” dimensions. For “mobility,” “looking after myself,” and “doing usual activities,” the percentage of reported problems was small. Therefore, both 3L and 5L versions presented very high ceiling effects, especially for these dimensions. In general, all dimensions in 3L and 5L showed a higher ceiling effect; nevertheless, a slight reduction is appreciable from 3L to 5L for all dimensions and for full health, with the exception of “doing usual activities.” To our knowledge, this is the first study examining the validity and reliability of the EQ-5D-5L-Y, so our results are incomparable with similar studies performed on young people. Nevertheless, in studies with adults from the general population, the ceiling effect reduction was slightly higher.<sup>16,17</sup> The high ceiling effect of the EQ-5D for young people has already been shown in previous studies in general population<sup>1,11</sup> and all brought attention to this fact. Nevertheless, the high ceiling effect should be contextualized because children and adolescents from the general population participated in our study. Therefore, future studies that aimed to compare the reported problems among unhealthy groups are recommended. It has not yet been possible to eliminate this error from the youth version of EQ-5D; nevertheless, it is key to note the population studied. In our work, the subjects participating were children and adolescents from the general population. Nevertheless, the findings show that in the general population, the ability of EQ-5D-5L-Y to detect moderate impairments of HRQOL and to detect the discriminative power between different health statuses is limited. Consequently, it would be beneficial to conduct similar studies in individuals with illness diversity.

In contrast, the inconsistency presented by the new EQ-5D-5L-Y was lower, ranging between 7% and 1.5%. “Having pain or discomfort” and “feeling worried, sad, or unhappy” showed more variability when the responses were redistributed from 3L to 5L. This could be explained by the nature of these dimensions. Concretely, these dimensions address psychosocial traits, whereas “mobility,” “looking after myself,” and “doing usual activities” refer to physical aspects.<sup>3</sup> It is necessary to note that the high ceiling effect influences this low inconsistency because most of the respondents rated their health on the positive end—“no problems”—which means the answer in both versions was unchanged. Therefore, the low inconsistency in all dimensions could

indicate that participants from the general population can consistently use the 2 versions. In addition, the mean and median VAS trend was statistically significant, decreasing in less healthy individuals. This also indicates the consistency between the 3L and 5L systems. Our findings are similar to previous studies performed on adults from the general population.<sup>16,17</sup>

Our results showed an increase of absolute informativity ( $H'$ ) in all dimensions and in full health in the newer 5L. In contrast, the relative informativity decreased in all dimensions except for “looking after myself” and for full health. In general, this indicates that the 5L produced higher informativity, even though our results present lower informativity than others carried out in adults.<sup>3,8,17</sup> Increasing the levels of response per dimension will probably produce a higher absolute informativity, but a decrease in relative informativity.<sup>9</sup> Nevertheless, these results should be interpreted with caution because Shannon indexes are new in the field of HRQOL and more studies are needed using the EQ-5D in children and adolescents to perform comparisons.

Regarding convergent validity, all dimensions of the 3L and 5L were significantly associated with previously validated HRQOL measures. Nevertheless, the dimensions of the 2 questionnaires had a low correlation magnitude. In this way, our hypothesis is partially fulfilled because even though this correlation is statistically significant, its magnitude is low, which is why the convergent validity of both the 3L and 5L versions remains in doubt. In contrast, we found a moderate or high correlation between the 3L and 5L versions. This discrepancy in the correlation between versions of EQ-5D, Youth 3L and 5L versions and other questionnaires could be due to the dimension scoring system. The KIDSCREEN system is based on a score from 0 to 100 and the Cantril's Life Satisfaction Ladder from 0 to 10. Therefore, its comparison with the EuroQol system with a score of 1 to 3 for the version with 3 response levels and from 1 to 5 for the version with 5 response levels can lead to misunderstandings. Thus, we urge you to be cautious with these results, stressing the need for more studies analyzing whether the convergent analysis is entirely correct when we compare questionnaires with very different scale ranges. The findings were consistent with previous research in young people, where low correlations between EQ-5D-Y dimensions and KIDSCREEN were found.<sup>1,11</sup> Nevertheless, the small magnitude of the associations could be affected by the high ceiling effect, especially in the dimensions “mobility,” “looking after myself,” and “doing usual activities.” This fact has also been highlighted in similar studies.<sup>1,11</sup> The results also showed a very high reliability for the percentage of agreement between test-retest and a significant kappa value. Nevertheless, the high ceiling effect had a negative impact on the reliability, given that the kappa coefficient could not be used to calculate the mean



**Table 3.** Redistribution properties from EQ-5D-3L-Y to EQ-5D-5L-Y.

Dimension	3L <sup>*</sup>	5L <sup>†</sup>	Subgroup	n	Proportions (%)	VAS mean	VAS median
Mobility (walking about)	1	1	g1.1	679	95.1	91.2	96
		2	g1.2	16	2.2	87.7	91
	2	2	g2.2	10	1.4	71.0	73
		3	g2.3	1	0.1	70	70
	3	4	g2.4	0	0.0		
		5	g3.5	0	0.0		
Missing				3	0.4		
Inconsistency				5	0.7	84.5	87
Looking after myself	1	1	g1.1	673	94.2	86.5	95
		2	g1.2	20	2.8	85.5	90
	2	2	g2.2	12	1.7	82.7	82.5
		3	g2.3	0	0.0		
	3	4	g2.4	0	0.0		
		5	g3.5	0	0.0		
Missing				3	0.4		
Inconsistency				6	0.8	81.2	80
Doing usual activities	1	1	g1.1	629	88.1	87.5	95
		2	g1.2	25	3.5	82.9	90
	2	2	g2.2	17	2.3	80.8	80
		3	g2.3	5	0.7	74	80
	3	4	g2.4	0	0.0		
		5	g3.5	0	0.0		
Missing				1	0.1		
Inconsistency				37	5.2	80.8	85
Having pain or discomfort	1	1	g1.1	535	74.9	88.4	95
		2	g1.2	44	6.1	82.2	87.5
	2	2	g2.2	61	8.5	81.9	85
		3	g2.3	16	2.2	64.6	70
	3	4	g2.4	0	0.0		
		5	g3.5	0	0.0		
Missing				2	0.3	55.0	55
Inconsistency				5	0.7		
Feeling worried, sad, or unhappy	1	1	g1.1	553	77.5	89.1	95
		2	g1.2	51	7.1	75.5	80
	2	2	g2.2	60	8.4	78.7	80
		3	g2.3	2	0.3	77.5	77.5
	3	4	g2.4	5	0.7	64	65
		5	g3.5	1	0.1	60	60
Missing				4	0.6		
Inconsistency				32	4.5	79.4	80

VAS indicates visual analog scale.

\*EQ-5D-3L-Y.

†EQ-5D-5L-Y.

**Table 4.** Informativity of the EQ-5D-3L-Y and EQ-5D-5L-Y systems.

Dimension	EQ-5D-3L-Y			EQ-5D-5L-Y			% change from 3L to 5L	
	H'	H' 95% CI	J'	H'	H' 95% CI	J'	H' (%)	J' (%)
Mobility (walking about)	0.19	(0.14-0.23)	0.27	0.24	(0.19-0.29)	0.22	26.3	-18.5
Looking after myself	0.12	(0.08-0.17)	0.11	0.18	(0.13-0.22)	0.26	50.0	136.3
Doing usual activities	0.27	(0.22-0.32)	0.25	0.31	(0.25-0.37)	0.22	14.8	-12.0
Having pain or discomfort	0.53	(0.47-0.58)	0.48	0.55	(0.48-0.60)	0.40	3.8	-16.6
Feeling worried, sad, or unhappy	0.46	(0.40-0.51)	0.42	0.56	(0.50-0.63)	0.35	21.7	-16.6
Full health (11111)	1.42	(1.45-1.73)	0.40	1.61	(1.28-1.53)	0.41	13.4	2.5

CI indicates confidence interval; H', Shannon index; J', Shannon evenness index.

**Table 5.** Comparison of correlation coefficients for the EQ-5D-3L-Y and EQ-5D-5L-Y.

KIDSCREEN-10 HRQoL index		
Mobility (walking about)	-0.04	-0.14*
Looking after myself	-0.02	-0.09 <sup>†</sup>
Doing usual activities	-0.22*	-0.09 <sup>†</sup>
Having pain or discomfort	-0.17*	-0.23*
Feeling worried, sad, or unhappy	-0.27*	-0.29*
KIDSCREEN-27 physical well-being		
Mobility (walking about)	-0.07 <sup>†</sup>	-0.22*
Looking after myself	-0.05	-0.14*
Doing usual activities	-0.13*	-0.11*
Having pain or discomfort	-0.09*	-0.21*
Feeling worried, sad, or unhappy	-0.19*	-0.21*
KIDSCREEN-27 psychological well-being		
Mobility (walking about)	-0.06	-0.09 <sup>†</sup>
Looking after myself	-0.01	-0.09 <sup>†</sup>
Doing usual activities	-0.18*	-0.12*
Having pain or discomfort	-0.19*	-0.17*
Feeling worried, sad, or unhappy	-0.25*	-0.29*
Life satisfaction ladder		
Mobility (walking about)	-0.08 <sup>†</sup>	-0.17*
Looking after myself	-0.00	-0.01
Doing usual activities	-0.09*	-0.13*
Having pain or discomfort	-0.10*	-0.27*
Feeling worried, sad, or unhappy	-0.17*	-0.36*
Convergent validity between 3L and 5L		
Mobility (walking about)	0.55*	
Looking after myself	0.46*	
Doing usual activities	0.32*	
Having pain or discomfort	0.48*	
Feeling worried, sad, or unhappy	0.51*	

3L indicates EQ-5D-3L-Y; 5L, EQ-5D-5L-Y; HRQoL, health-related quality of life.

\*P-value <.05.

<sup>†</sup>P-value <.01.

reliability for the “mobility” and “looking after myself” dimensions. These results are similar to those obtained in the international EQ-5D-3L-Y validation study.<sup>1</sup>

This study has some limitations. First, the questionnaires were completed in small student groups, so there may be some response/social desirability bias. All the participants were children and adolescents from the general population, and none presented with any health conditions. Therefore, most people reported being in full health. This fact negatively influences the high ceiling effect that was presented, indicating that the newer 5L version has low sensitivity in the general population. Therefore, additional research is required to determine whether 5L is more efficient than 3L in detecting health variations and to examine behavior in specific populations and the responsiveness to change in sub-groups with heterogeneous health conditions. In addition, the ceiling effect of the EQ-5D-5L-Y is still high compared with the EQ-5D-3L-Y version. Hence, the objective of EuroQol foundation in reducing the ceiling effect is not fully accomplished. Nevertheless, it should be taken into account that the studied population does not have special health characteristics. Therefore, we recommend that this test be used in combination with other more sensitive instruments in the healthy population. Further research is needed to determine the psychometric properties of this instrument in the clinical population and to explore its use in health economic assessments.

On the basis of the results, this study has implications for researchers and clinicians. Until now, the EQ-5D-3L-Y version was used to assess HRQOL in children, often presenting results with a high ceiling effect and little discriminant power. When the new version EQ-5D-5L-Y appears, the informative and discriminating capacity is greater. A version with 5 response levels allows capturing a greater number of respondents who are between levels 1 and 3, when they were at level 1 with the 3-level version. In addition, the 5-level version allows placing 3 anchors, for one part, the best and worst levels (5 and 3, respectively) and on the other, an intermediate anchor (level 3 response), through which to discern between how much below I am in bad or good with respect to this. Thus, we recommend the use of the version with 5 response levels in future studies and assessments of the HRQOL.

### Conclusions

To our knowledge, this is the first study to assess the validity and reliability of the newer EQ-5D-5L-Y and to compare it with the

oldest version containing 3 levels of response in general population. Our results show that the 5L version is feasible, consistent, and reliable in healthy youth. The findings also show limited differences in the informativity and ceiling effect of EQ-5D-5L-Y compared with the EQ-5D-3L-Y.

## Article and Author Information

**Accepted for Publication:** July 7, 2021

**Published Online:** September 1, 2021

doi: <https://doi.org/10.1016/j.jval.2021.07.006>

**Author Affiliations:** Faculty of Sport Sciences, University of Extremadura, Cáceres, Spain (Pérez-Sousa, Olivares, Gusi); Epidemiology of Physical Activity and Fitness Across Lifespan research group, University of Seville, Seville, Spain (Pérez-Sousa, Olivares); Faculty of Education, Psychology and Sport Sciences, University of Huelva, Huelva, Spain (Olivares); Instituto de Actividad Física y Salud, Universidad Autónoma de Chile, Talca, Chile (Olivares); Navarrabiomed-Universidad Pública de Navarra, Complejo Hospitalario de Navarra, Instituto de Investigación Sanitaria de Navarra, Pamplona, Navarra, Spain (Ramírez-Vélez); CIBER of Frailty and Healthy Aging, Instituto de Salud Carlos III, Pamplona, Navarra, Spain (Ramírez-Vélez).

**Correspondence:** Miguel Ángel Pérez-Sousa, PhD, Faculty of Sport Sciences, University of Extremadura, Cáceres, Spain, Ave de la Universidad, s/n, Cáceres 10003, Spain. Email: [perezsousa@gmail.com](mailto:perezsousa@gmail.com)

**Author Contributions:** *Concept and design:* Pérez-Sousa

*Acquisition of data:* Pérez-Sousa

*Analysis and interpretation of data:* Pérez-Sousa, Olivares, Ramírez-Vélez

*Drafting of the manuscript:* Pérez-Sousa

*Critical revision of the paper for important intellectual content:* Olivares, Ramírez-Vélez

*Statistical analysis:* Pérez-Sousa, Ramírez-Vélez

*Obtaining funding:* Gusi

*Administrative, technical, or logistic support:* Gusi

*Supervision:* Olivares, Gusi

**Conflict of Interest Disclosures:** The authors reported no conflicts of interest.

**Funding/Support:** This work was supported by grants 20170290 from the EuroQol Research Foundation, Netherlands.

**Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Acknowledgment:** The authors express their gratitude to the respondents who completed the interviews, teachers, and schools.

## REFERENCES

1. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19(6):887–897.
2. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res*. 2019;28(7):1951–1961.
3. 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):275–284.
4. Rencz F, Lakatos PL, Gulácsi L, et al. Validity of the EQ-5D-5L and EQ-5D-3L in patients with Crohn's disease. *Qual Life Res*. 2019;28(1):141–152.
5. Ravens-Sieberer U, Herdman M, Devine J, et al. The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. *Qual Life Res*. 2014;23(3):791–803.
6. Cantril H. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press; 1966.
7. 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. *Qual Life Res*. 2013;22(7):1717–1727.
8. Yfantopoulos JN, Chantzaras AE. Validation and comparison of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in Greece. *Eur J Heal Econ*. 2017;18(4):519–531.
9. 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. *Qual Life Res*. 2007;16(5):895–904.
10. Cohen J. *Statistical Power Analysis for the Behavioral Science*. Lawrence Erlbaum Associates; 1988. <https://www.taylorfrancis.com/books/mono/10.4324/9780203771587/statistical-power-analysis-behavioral-sciences-jacob-cohen>.
11. Gusi N, Perez-Sousa MA, Gozalo-Delgado M, Olivares PR. Validez y fiabilidad de la versión proxy del EQ-5D-Y en español [Validity and reliability of the Spanish EQ-5D-Y proxy version]. *Ann Pediatr (Barc)*. 2014;81(4):212–219.
12. Cohen J. Weighted kappa - nominal scale agreement with provision for degrees of disagreement. *Am Psychol*. 1967;22:544.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
14. Fleiss IJ, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613–619.
15. 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;9:55.
16. 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.
17. Kim TH, Jo MW, Lee Sil, Kim SH, Chung SM. Psychometric properties of the EQ-5D-5L in the general population of South Korea. *Qual Life Res*. 2013;22(8):2245–2253.