



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

Sarah Derrett^{1,2}  · Mike Herdman^{2,3}  · Lucky G. Ngwira^{4,5}  · Elizabeth Yohe Moore⁶ · Jennifer Jelsma^{2,7}

Accepted: 23 January 2021
© The Author(s) 2021

Abstract

Introduction The beta EQ-5D-Y-5L is a new patient-reported outcome measure (PROM) for children aged 8–15 years that is currently under development by the EuroQol Group. The EQ-5D-Y-5L is similar to the EQ-5D-Y but has five levels of severity per dimension rather than three. The increased number of levels increases the granularity of the responses but possibly has also increased the difficulty of distinguishing between levels. The EuroQol's Version Management Committee (VMC) required a robust method to determine how well children distinguish between the five EQ-5D-Y-5L ordinal severity qualifiers (i.e. 'no problems' through to 'extreme problems'), which are a critical aspect of both health measurement and the valuation of health states.

Objective This paper describes the development, testing, selection, and piloting of such a method.

Methods Following a literature review and consultation with the wider VMC and a Language Support Services agency, a range of exercises were developed to assess the ordering and comprehension of the five severity qualifiers. Three exercises were pre-tested with children in Spain and New Zealand. One exercise, preferred and understood by children, was then piloted.

Results Five children in Spain and 11 in New Zealand tested the three exercises. In both countries, all children found the three exercises easy to understand and complete. Of the 12 children who expressed a preference, nine said they preferred the card ranking. Card ranking also allowed the interviewer to observe difficult choices being made as the children physically rearranged the card order until they settled on their final order. Following rigorous assessment of translatability and cultural portability by an independent Language Support Service, card ranking was piloted in South Africa ($n=9$) and in Indonesia ($n=10$), where it highlighted severity qualifier order inversions that would otherwise not have been detected.

Conclusion The card ranking exercise was found to be a preferred and acceptable means of testing the ordering of translations of severity qualifiers among children. Additional formal testing of the exercise in other countries and languages is now underway. The approach developed and tested by the VMC for cognitive debriefing of beta EQ-5D-Y-5L language/country versions may also be useful in determining the adequacy of translated qualifiers in debriefing of adult EQ-5D-5L versions and other PROMs.

1 Introduction

Patient-reported outcome measures (PROMs) assess the health status or health-related quality of life of patients (or individuals from the general population) at, or over, a specified time [1]. PROMs can be condition specific (assessing particular health conditions, e.g. diabetes) or generic (assessing health across a wide range of health conditions

and in healthy people). PROMs are used in clinical research, policy-making, and population survey contexts [2–4]. For example, PROMs may be used in clinical trials of new interventions and, in particular population subgroups over time to assess group or population-level health status changes. One type of PROM, multi-attribute utility instruments (MAUIs), are often used in cost-utility analyses (CUAs) to inform policy makers' decisions about the costs and benefits associated with, for example, the introduction of new clinical or pharmaceutical treatments [5]. With MAUIs, preference weights (utilities) are estimated for all combinations of health states measured by the instrument. Estimated utilities are derived using valuation studies, usually undertaken with

✉ Sarah Derrett
sarah.derrett@otago.ac.nz

Extended author information available on the last page of the article

Key Points

The EQ-5D-Y-5L is a patient-reported outcome measure (PROM) for children aged 8–15 years that is being developed by the EuroQol Group. It is similar to the EQ-5D-Y but has five severity response options per dimension rather than three.

A novel approach to determine how well children distinguish between the five EQ-5D-Y-5L ordinal severity qualifiers (i.e. ‘no problems’ through to ‘extreme problems’) in new translations was developed and piloted.

A card ranking approach was preferred by children and usefully examined the ordering of translated severity qualifiers within the standard cognitive debriefing process. This approach may also be useful in determining the adequacy of translated qualifiers in debriefing of adult EQ-5D-5L versions and other PROMs.

members of the general population, with techniques such as time trade-off or discrete choice experiments [5]. Estimated utilities can then be used together with survival data to generate quality-adjusted life-years for economic CUAs [6, 7]. Consequently, the appropriateness of severity qualifiers (and their translations) within MAUIs may have implications for country-based resource allocations.

PROMs require careful translation and cultural adaptation for two key reasons. First, it is desirable that individuals self-reporting their health can readily comprehend the severity qualifiers to allow them to describe their health accurately. For example, in a pain dimension, it should be clear that ‘a lot of pain’ represents a greater degree of the attribute than ‘some pain’, which is in turn greater than ‘a little pain’. This will usually be clear, especially when there are few severity levels (e.g. three) in a dimension. However, when the number of levels increases, it can become more difficult to translate severity qualifiers to ensure they are sufficiently distinct from each other and the intended hierarchical ordering is clear. Additionally, when considering MAUIs, qualifier inversion may be particularly problematic for valuation studies (e.g. when deriving utilities for use in CUA), when the qualifiers being valued are ‘decontextualised’ from other severity labels in the same dimension. ‘Preference inversion’ has been reported in some languages, albeit infrequently [8, 9]. Identified inversions may, at least in part, be due to some respondents confusing the hierarchical ordering of severity qualifiers (qualifier inversion) when qualifiers are presented out of the ordered context of the full questionnaire.

To allow for multi-country application, PROMs require robust cultural adaptation to ensure appropriate language versions are available [10]. The standard cognitive debriefing

procedure involves in-depth testing of a preliminary consensus version of a translated questionnaire via face-to-face interviews with a small sample of five to ten native speakers of the target language to assess the acceptability, comprehensibility, and interpretation of translated terms [11]. From the start, the EuroQol Group, responsible for developing and managing the widely used suite of EQ-5D MAUIs, recognised the importance of having a rigorous translation procedure in place for all their instruments to ensure quality and linguistic and semantic equivalence across languages [12].

The EuroQoL’s measures include the standard EQ-5D-3L instrument for adults, which consists of two main elements: a five-dimension descriptive system (with three severity levels per dimension) and a visual analogue scale (VAS), capturing global health on a 0–100 vertical line. Similarly, the EuroQol Group’s EQ-5D-Y is a generic self-report measure of health status designed for use in children and adolescents aged 8–15 years [13]. The EQ-5D-Y’s descriptive system assesses health status across five dimensions (mobility, looking after myself, usual activities, pain/discomfort, and emotional well-being) and the VAS [13]. Each dimension in the EQ-5D-Y has three levels of severity, and respondents choose one level in each dimension to provide a profile of their health status on the day of assessment. No value sets are currently available for the EQ-5D-Y, although a valuation protocol has been developed, and valuation work is underway in some countries [6].

An expanded adult version (EQ-5D-5L), with five severity qualifiers per dimension, was introduced in 2011 [14]. Given the increase in severity qualifiers (from three to five), the Version Management Committee (VMC), the body within the EuroQol Group charged with overseeing translations of the EQ-5D, introduced an additional task to determine whether qualifiers in translated versions represent similar severity to those in the original English language (source) version. Participants in cognitive debriefings for translations of the EQ-5D-5L were asked to provide VAS ratings of the severity qualifiers in addition to standard cognitive debriefing questioning [12].

Currently, an expanded version of the EQ-5D-Y is under development with five levels of severity per dimension [15]. This instrument, the EQ-5D-Y-5L, is considered a beta version, as it is not yet a finalised official EuroQoL product. Early cognitive debriefing exercises with translated or adapted versions of the beta EQ-5D-Y-5L revealed that some children had difficulty distinguishing between severity levels even when they were seen in the context of the descriptive system, where the ordering should be clearer. As the adult VAS ratings of severity had not always been successful in preventing ‘mis-ordering’, the VMC sought development of child-friendly methods to determine whether an appropriate degree of distinguishability and hierarchical ordering is achieved in each new translation of the beta EQ-5D-Y-5L.

Elsewhere, only a small number of studies appear to have investigated strategies to rank preferences or qualifiers with children. Strategies have included scales with ‘cartoon’ faces [16, 17], smiley faces [18], and graduated circles [19].

This paper aims to describe the development, selection, and testing of a novel method for use in future translations of the beta EQ-5D-Y-5L. The objectives were as follows:

1. Develop and test several child-friendly approaches to assessing children’s interpretations of severity qualifiers in translated versions of the EQ-5D-Y-5L.
2. Explore children’s preferences for, and understanding of, alternative approaches to inform the selection of one preferred approach.
3. Determine the translatability, cultural portability, and validity of the selected approach for use in future translations of the EQ-5D-Y-5L.
4. Pilot the selected approach in other countries/languages.

2 Methods

2.1 Iterative Consensus-Based Development

The project used a multi-stage research process and range of methods (Fig. 1). Underpinning all stages was an emphasis on consensus decision making. The key team involved were the EuroQoL Group’s multilingual VMC (eight members from seven countries with disciplinary backgrounds including patient-reported outcome methodologists, health professionals, health services, public health, and/or university academics). The EuroQoL Youth Population Working Group, who developed the new EQ-5D-Y-5L, were also consulted, along with analysts from an independent Language Support

Service (LSS; translation agency) with backgrounds in translation methods, neuroscience, economics, and public health.

2.2 EQ-5D-Y-5L Severity Qualifiers for Assessment

As mentioned, the aim of this paper is to describe the development and testing of an approach for assessing children’s interpretations of severity qualifiers in translated versions of the beta EQ-5D-Y-5L. The EQ-5D-Y-5L asks children about problems across five dimensions: mobility (MO), looking after myself (LAM), usual activities (UA), pain or discomfort (PD), and worried, sad, or unhappy (WSU). Severity qualifiers and numerical codes (for reporting) are presented in Table 1. Unique severity qualifiers are evident for PD level 5 and WSU 1, 3, 4, and 5.

2.3 Five Exercises Initially Developed

Following consultation with the VMC and LSS, and considering the literature, five exercises were initially developed. Instructions were prepared for interviewers administering the exercises to children. ‘Warm-up’ exercises were developed in a colourful format designed to appeal to children to familiarise children with the tasks involved. Exercise 1 was smileys with eyebrows. Exercise 2 was smileys (no eyebrows) with traffic light colour coding ranging from green (representing ‘no problems’) to red (representing ‘extreme/cannot do’). In both exercises 1 and 2, children were instructed to draw lines connecting the severity qualifier text to the smiley they thought best represented the severity described (or to the same smiley if they thought the qualifiers were equivalent). The qualifiers’ text was randomly ordered on the page (i.e. not according to the intended hierarchical order). Exercise 3 was paired choices; pairs of

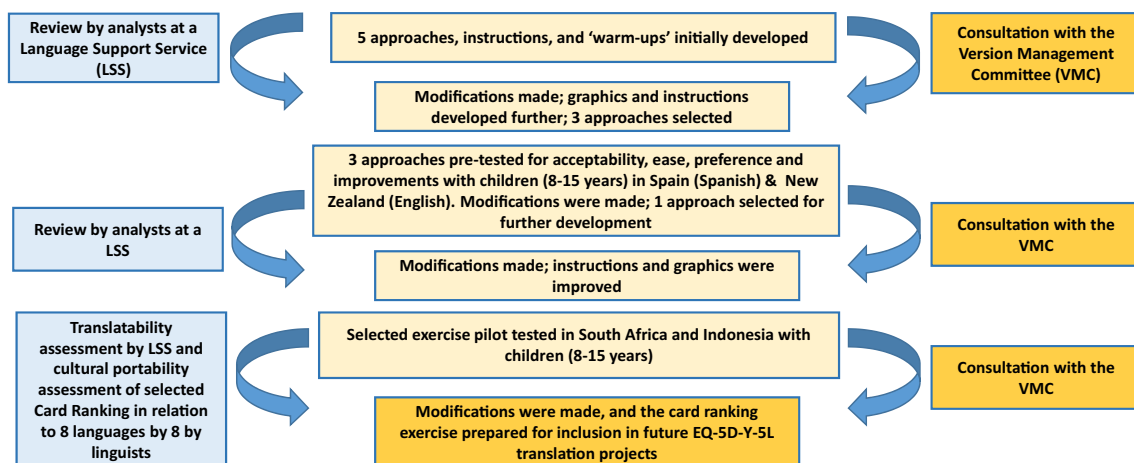


Fig. 1 Overview of project stages.

cards containing severity qualifiers were to be distinguished by placing the cards into three piles labelled ‘This child has the smallest problem’, ‘Equal problems’, and ‘This child has the biggest problem’. Exercise 4 was graduated circles, where five differently sized circles represented the magnitude of qualifier severity. Children were instructed to draw lines between each qualifier (from a randomly ordered list) to the circles. The smallest problem was to be linked to the circle of smallest size, the biggest problem to the biggest circle, etc. Children could draw lines to the same circle if they considered the qualifiers to be equivalent. Exercise 5 was card ranking, where sets of shuffled cards containing the severity qualifiers were to be placed onto a column of five empty boxes indicating their relative severity between anchor points of ‘Child has the smallest problem’ and ‘Child has the biggest problem’. Children were given one set of shuffled cards at a time for each of the EQ-5D-Y-5L dimensions. Children were to read all five cards before placing them into the five boxes ranging from the smallest amount of problem to the biggest. Children were to place two cards into the same box if they regarded the severity qualifiers as equivalent.

2.4 Three Exercises for Pre-Testing in Spain and New Zealand

The three exercises were then pre-tested for acceptability and overall preference with children aged 8–15 years in two countries: Spain (Spanish) and New Zealand (English) (Fig. 2). Parental approval and the child’s assent were obtained before children participated in pre-testing. In Spain, the English wording for the exercise instructions were translated into Spanish by the local investigator. Convenience sampling was used in both countries, with effort made to involve children with a range of characteristics (e.g. different ages, sexes, educational backgrounds, and ethnicities). In Spain, where pre-testing occurred first, five participants were sought. In New Zealand, it was intended to recruit up to ten children to allow for administering the three exercises in different orders. Data were collected about completion times, ease of completion, preference for particular exercise(s),

difficulty with any exercise(s), and ways in which the exercises could be made easier for children to complete in the future.

2.5 Translatability and Cultural Portability Assessment

The card-ranking exercise was the approach selected following pre-testing, the earlier linguistic appraisal, and consensus among the VMC committee. An independent LSS then undertook a translatability and cultural portability assessment of the English source text of the exercise, warm-up, and instructions. Linguists reviewed the instrument in relation to eight languages to identify any concepts, phrases, or components that would be difficult to translate or appeared to be culturally specific. Further changes were made to the selected exercise in response to this review.

2.6 Pilot Testing the Card Ranking Exercise in South Africa and Indonesia

EQ-5D-Y-5L beta versions were being developed and validated in South Africa (English language) and Indonesia (Indonesian). The card ranking exercise was therefore incorporated into existing validation studies. The purpose was to determine the feasibility of administering the game within a real-world context and to describe its potential usefulness in assessing the hierarchical ordering and distinguishability of the severity qualifiers. Children (aged 8–15 years) with and without health problems were recruited from schools and medical institutions in South Africa and Indonesia. Only four dimensions (MO, LAM, PD, and WSU) were included in the card ranking exercise because these four encompass the full range of severity qualifiers from the source English language version of the EQ-5D-Y-5L; the qualifiers for UA are the same as for MO and LAM (see Table 1).

Eligible children needed to be able to read and write in English or Indonesian (in South Africa and Indonesia, respectively). It was planned that at least eight children, with a range of ages, would be recruited in each country as that was the number involved in the standard cognitive debriefing

Table 1 Beta EQ-5D-Y-5L (UK English) dimensions and levels shaded to illustrate between-dimension qualifier similarities and differences

Beta EQ-5D-Y-5L dimensions (and abbreviation)	Severity qualifiers (and numerical code)				
	First level [1]	Second level [2]	Third level [3]	Fourth level [4]	Fifth level [5]
Mobility (MO)	No [problems walking about]	A little bit	Some	A lot	Cannot do
Looking after myself (LAM)	No [problems washing or dressing myself]	A little bit	Some	A lot	Cannot do
Usual activities (UA)	No [problems doing my usual activities]	A little bit	Some	A lot	Cannot do
Pain or discomfort (PD)	No [pain or discomfort]	A little bit	Some	A lot	Extreme
Worried, sad or unhappy (WSU)	Not [worried, sad or unhappy]	A little bit	Quite	Really	Extremely

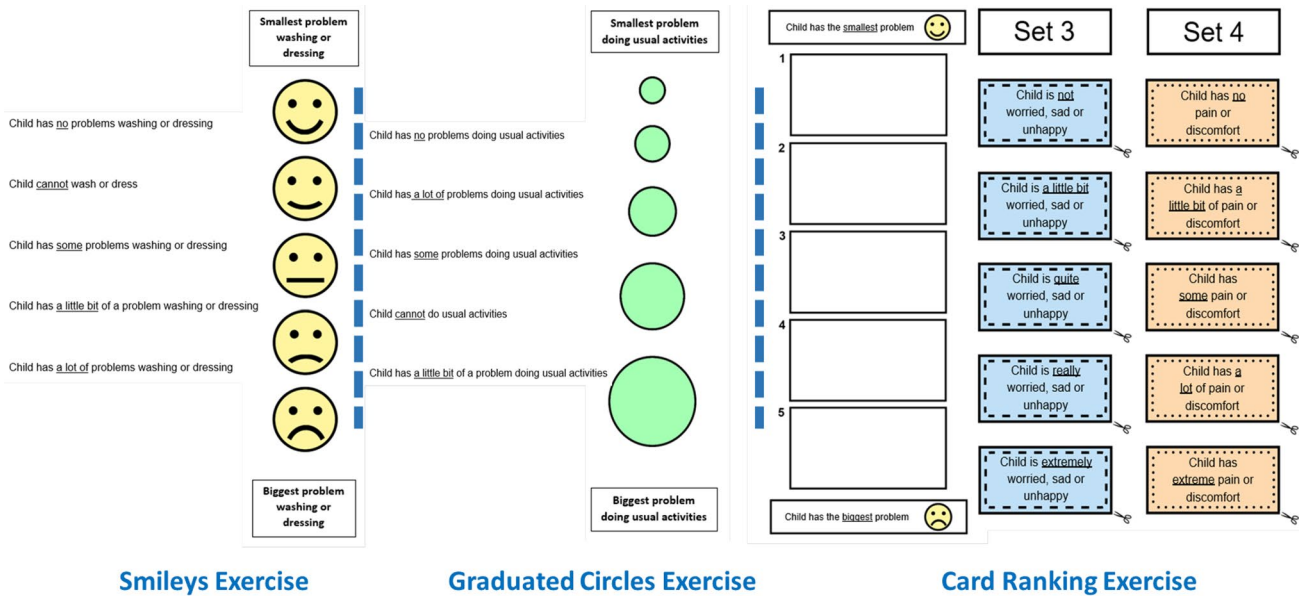


Fig. 2 Overview of the three games selected for pre-testing in Spain and New Zealand

process used by the VMC. Children who were critically ill and admitted to the intensive care unit were excluded. Informed consent was obtained from the caregivers and assent from the children. The interviewers were asked to comment on the exercise and the results obtained, including the ease with which the children had performed the exercise.

2.7 Data Collection

Each child's qualifier ordering, for the four dimensions assessed, was to be entered onto a VMC data collection sheet by the interviewer using the relevant EQ-5D-Y-5L number code. This was to provide a tabulated summary of the order of qualifiers to compare with the developers' intended hierarchical order. Interviewers in South Africa and Indonesia were also asked to report on their overall impressions of administering the card ranking exercise: its ease of completion by children and usefulness in identifying qualifier inversions.

3 Results

3.1 Five Exercises Initially Developed

Feedback from the LSS about the five proposed exercises initially proposed, and further consultation with the VMC, led us to make some amendments. The LSS had previously investigated the cross-cultural use of smileys and found that some populations prefer more simplistic 'emoji' faces and others more 'cartoon-style' faces. Exercise 1 (smileys with

eyebrow expressions) was therefore rejected. Additionally, feedback indicated that exercise 2 (smileys with traffic light colours) was founded on a premise that green equates to 'good' and red to 'bad', which is not the case in all cultures or countries, where red can represent 'happiness' or 'good luck'. The traffic light colours were therefore replaced with a uniform pale yellow in exercise 2 for pre-testing. Exercise 3 (paired choices) was considered cognitively difficult for children as young as 8 years and was rejected. Because graduated circles (exercise 4) had been used successfully by others [19], albeit in a different context, the VMC decided to test graduated circles. The LSS considered exercise 5 (card ranking) to have the best face validity, so this also moved to pre-testing.

3.2 Pre-Testing in Spain and New Zealand

Graphics for the three exercises were further refined, along with appropriate warm-up exercises and instructions for the interviewers. The LSS advised that the use of different colours in the graphics for each exercise should be minimised because, in some countries, interviewers administering the exercises may not have access to colour printers. Consequently, designs were developed that would work well in a black and white format. The smileys were pale yellow, the graduated circles were pale green, and the card ranking card sets had unique black and white borders to distinguish between the sets (see Fig. 2).

The three exercises were tested in children aged 8–15 years, first in Spain with five children and then in New Zealand with 11 children. Following the Spanish pre-testing,

the exercise instructions were shortened to make them easier to administer, and the number of warm-up exercises was reduced. The order of exercise administration between participants in New Zealand was varied to reduce the likelihood of order influencing children's preferences. Demographic characteristics of the children, and an overview of the pre-testing from both countries, are presented in Table 2.

Overall, the interviewers' reading of instructions to the children and the children's completion of all three exercises took less than 20 minutes; completion times did not appear related to age. Of those expressing a preference, nine of 12 children preferred the card ranking exercise. Children of all ages considered all exercises to be easy to understand, and age did not appear related to preference for the different approaches. Because of this, and the previous suggestion from the LSS that this exercise had strong face validity, card ranking was selected for pilot testing. Additionally, pre-testing found the smiley and graduated circles (involving drawing connecting lines) awkward to correct if children changed their ordering decisions; sometimes differently coloured pens were required to show the final preferred order. In contrast, the card ranking exercise simply required children to reposition the cards on the page. Finally, card ranking permitted interviewers to observe which qualifiers constituted 'tough choices'. The children tended to place qualifiers '1' and '5' into boxes 1 and 5 first, and the interviewer could observe the decision making as children swapped the order of the three remaining cards until they achieved a final card order that was satisfactory to them.

3.3 Further Consensus, Translatability and Cultural Portability Assessment

Following the pre-testing, subsequent consultation with the VMC resulted in the use of first-person severity qualifiers, which had already been translated during from the forward testing and backward testing stages of the standard translation process, rather than the third-person wording used in the pre-testing. For example, in the MO dimension, the first severity qualifier wording would use the exact translated wording ('I have no problems walking about') rather than the third-person wording ('Child has no problems walking about'). An additional instruction for the card ranking exercise was added, explaining that (e.g. for MO) 'The next cards are about problems a child has walking about. The cards are not about your own walking; you can think of (imagine) any child.'

The card ranking exercise and instructions then underwent significant translatability and cultural portability assessment by the LSS in eight languages: English (Australia), French (France), Norwegian (Norway), Czech (Czech

Republic), Japanese (Japan), Arabic (Egypt), Hindi (India), and Zulu (South Africa). Although the linguists found the card ranking and instructions to be generally translatable, a number of issues were identified. Amendments were then made, e.g. words were deleted (such as 'teenager'), and more appropriate vocabulary was used (e.g. 'digital voice recorder' instead of 'tape recorder'). Colloquial words in the instructions (in italics), such as 'Did anything sound strange or *funny*?', '*Sitting in* on an interview', and a duplication of concepts in the warm-up exercise 'Attending (*going to*) school' were altered. The assessors were concerned that, in some languages (e.g. Arabic), younger children would be unable to complete the exercise without adult guidance and that this should be permitted. However, the VMC team did not support this suggestion because other EQ-5D instruments are available for interviewer administration and/or proxy completion. Instead, the instructions were amended to explicitly note that the interviewer should only assist the child with the card ranking task in the first introductory 'warm-up' set of cards and not with placing the four sets of EQ-5D-Y-5L cards being assessed.

3.4 Pilot Testing Card Ranking in South Africa and Indonesia

In both South Africa and Indonesia, the interviewers reported that the children completing the card ranking understood the task. The warm-up introductory (attending school) lead-in exercise was helpful. In Indonesia, the interviewer reported that the warm-up helped children to feel more relaxed. However, in South Africa, two 8-year-old children were unable to progress further than the introductory task and were excluded from the full exercise.

Results of the pilot testing in South Africa ($n=9$) and Indonesia ($n=10$) are presented in Table 3 where any instances of qualifier inversions occurred; there were no inversions for the PD dimension in South Africa or for the MO or WSU dimensions in Indonesia. Four of the 19 had inversions. The interviewers reported that these were either due to the child's age or the severity qualifier wording on the cards rather than children misunderstanding the task. The most disrupted qualifier ranking, compared with the intended hierarchical order, was among the youngest age group.

In South Africa, no child experienced problems when answering about their own health on the EQ-5D-Y-5L as the qualifiers are presented in order in the questionnaire. Interviewers in both countries reported that the problem of qualifier inversion would not have been identified if the card ranking had not been conducted.

Table 2 Children's demographic characteristics, and findings, from pre-testing in two countries

Characteristics and findings	Spain (<i>n</i> = 5)	New Zealand (<i>n</i> = 11)
Participants	3 girls, 2 boys	5 girls, 6 boys
Age in years (<i>n</i>)		
8–9	2	3
10–11	2	4
12–13	0	2
14–15	1	2
Minutes to complete	15–20	9–17
Ease and comprehension of task	All children grasped what they had to do very quickly. The interview was not too long, and they seemed to enjoy it. All tasks were equally easy to complete, and no children reported inversion of severity qualifiers	All exercises and instructions were reported as very easy to understand and complete. Three children inverted qualifiers from the intended hierarchical ordering, e.g. placing 'some problems' at a lesser severity than 'a little bit of a problem'
Exercise preference (<i>n</i>)		
No preference	2	2
Graduated circles	0	0
Smileys	1 ^a	3
Card ranking	3 ^a	6

^aOne Spanish participant preferred the smileys and card ranking equally, so the numbers of Spanish children do not equal 5 in this category

4 Discussion

The VMC recognised the need to explicitly test that children would be able to discriminate between and correctly rank the five levels of the newly developed EQ-5D-Y-5L beta version. Translation of PROMs is time consuming; the standard process requires forward and backward translations and cognitive testing [10, 11]. Consequently, a new language version can take 3 months or more to develop, and additional processes should only be added after due deliberation. Through the use of an iterative staged methods project, which combined consensus building among experts and analyses of collected data, we developed a method that appears fit for purpose, i.e. ensuring that severity qualifiers are hierarchically ordered by children as intended by the EuroQoL Group [14]. We also established that this method could be embedded in the cognitive debriefing exercise within the standard translation protocol of the EuroQoL products. Card ranking appears to be translatable and culturally portable. It may be that, if qualifier inversion is addressed adequately during the translation process, the incidence of preference inversion will be further minimised when utility weights are later developed.

The literature search did not return many examples of ranking exercises suitable for children, so it was fruitful to draw on the international translation expertise and experience of the wider VMC and LSS. This helped us to identify three candidate exercises for testing (smiley faces, graduated circles, and card ranking). It is important to reduce child respondent burden in cognitive debriefing as far as possible; on pre-testing, the child respondents completed all

three exercises in less than 20 minutes, and all respondents found the three prototypes easy to complete and acceptable. There was a clear preference for card ranking. An advantage of this approach, from our point of view, was that the interviewer doing the cognitive debriefing could easily see the child's decision making through the visible swapping of cards until the child was satisfied with their final placement.

One interpretation of reported qualifier inversions in the adult EQ-5D-5L, despite rigorous translation procedures, is that the inversion was not apparent during translation. Although the magnitude of the difference between qualifiers in each dimension of the EQ-5D-5L was assessed using a VAS, this was undertaken in the context of the hierarchically presented descriptors being seen prior to the VAS assessment [12]. This may have contributed to counter-intuitive EQ-5D-5L preference weights being developed in a few languages [8, 9]. The card ranking exercise was found to be fit for purpose in South Africa and Indonesia. Not only was it acceptable to the respondents, but it was also reported as critical for the identification of inversions and resulted in a hierarchically ordered set of qualifiers consistent with the source (English language) version. The VMC requires the card ranking to be completed before the standard cognitive debriefing exercise to avoid prior learning of the intended hierarchical order. For language versions that have already undergone translation, it may be profitable to undertake a similar ranking exercise to ensure there is no inversion of qualifiers before embarking on a valuation exercise in which the dimension qualifiers being assessed are usually presented out of the hierarchical context of the questionnaire. Results from a Malawi (Chichewa) project investigating

Table 3 Data collection from South African ($n=9$) and Indonesian ($n=10$) card ranking

EQ-5D-Y-5L Dimension	Children's Age Group													
	SOUTH AFRICA (n=9) - 4 girls - 5 boys			8-9 years (n=3)			10-11 years (n=3)			12-15 years (n=3)				
Mobility (Walking about)														
No problems [1]	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Little bit of a problem [2]	3	2	2	2	2	2	2	3	2	2	2	2	2	2
Some problem [3]	1	3	3	3	3	3	3	2	3	3	3	3	3	3
A lot of problems [4]	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Cannot [5]	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Looking after myself														
No problems [1]	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Little bit of a problem [2]	1	2	2	2	2	2	3	2	2	2	2	2	2	2
Some problem [3]	5	3	3	3	3	3	2	3	3	3	3	3	3	3
A lot of problems [4]	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Cannot [5]	3	5	5	5	5	5	5	5	5	5	5	5	5	5
Worried, sad or unhappy														
Not worried, sad or unhappy [1]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A little bit [2]	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Quite [3]	4	3	3	3	3	3	4	3	3	3	3	3	3	3
Really [4]	5	4	4	4	4	4	3	4	4	4	4	4	4	4
Extremely [5]	3	5	5	5	5	5	5	5	5	5	5	5	5	5
INDONESIA (n=10) - 4 girls - 6 boys														
Looking after myself														
No problems [1]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Little bit of a problem [2]	2	3	2	2	2	2	2	2	2	2	2	2	2	2
Some problem [3]	3	2	3	3	3	3	3	3	3	3	3	3	3	3
A lot of problems [4]	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Cannot [5]	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Pain or discomfort														
No pain or discomfort [1]	2	1	1	1	1	1	1	1	1	1	1	1	1	1
A little bit [2]	1	2	2	2	2	2	2	2	2	2	2	2	2	2
Some [3]	3	3	3	3	3	3	3	3	3	3	3	3	3	3
A lot of [4]	5	4	4	4	4	4	4	4	4	4	4	4	4	4
Extreme [5]	4	5	5	5	5	5	5	5	5	5	5	5	5	5

*Colouring highlights the qualifier inversions where the child has ordered the cards in a manner not intended by the translators

the application of this new card ranking process within a complete translation are now being prepared for publication [20]. Provisional findings are that the cognitive debriefing involved four rounds of card ranking with revision of wording for severity qualifiers, and re-testing with different children, in each round. Improved wording led to the proportion of qualifier inversions reducing from over 40% in the first round to 2% in the fourth. As in South Africa and Indonesia, the card ranking was reported (by LGN) to be acceptable and understandable and led to improvements that otherwise may not have occurred.

Generally, the source version of a PROM is distributed in a single language for use in different cultural and linguistic settings. The translatability and cultural portability assessments provided us with essential feedback to prevent the use of inappropriate colours and words that were too colloquial. We recommend that translations of new PROMs, and related instructions, undergo similar assessments.

Strengths of the project include the input received from experts in a variety of disciplines, testing in different cultural and linguistic contexts, and the iterative nature of

the process. We suggest that card ranking helps identify potential problems with translations of severity qualifiers of a MAUI, such as the EQ-5D-Y-5L, and provides remedies for those problems through repeated rounds of the exercise. However, the type of problems we experienced in understanding the comprehensibility and hierarchical ordering of the severity qualifiers may not apply to all MAUIs. For example, ordering problems may be more prevalent in instruments such as the EQ-5D that use ordinal or Likert-type scales. It is possible that the card ranking process developed here for children may offer some advantages over the VAS approach that has been used for some time by the VMC to ascertain whether adults 'rank' severity qualifiers in the anticipated hierarchical order [12]. Anecdotally, the VMC know from many years of assessing translations that some adults (e.g. those that are not familiar with mathematical scales) can find the VAS method of ranking severity challenging. We acknowledge that rigorously translating MAUIs, as undertaken by the VMC, takes time. Indeed, time penalties incurred by introducing a card ranking exercise to the cognitive debriefing process for adults should be considered

in relation to the current VAS process. However, card ranking was quick for children to complete and is likely to be more efficient for adults than the VAS severity ranking process currently undertaken. Regardless, time penalties ought to be considered alongside the possibility of MAUIs being translated with severity qualifiers possibly unsuited to deriving utility weights. Further research is now underway by the VMC to modify the child card ranking process for testing with adults in future VMC translations of the adult EQ-5D-5L.

A limitation is that only one translation (Chichewa—Malawi) has piloted the method within a full translation process; a paper describing this translation is now being prepared. However, the card ranking approach is now being evaluated in other languages and countries, such as Ethiopia and Singapore. It will be important for the VMC to continue evaluating the suitability of card ranking within the cognitive debriefing process to ensure it is, indeed, appropriate for other language/culture groups. Another limitation may be that, based on pre-testing and advice from the LSS about the greatest face validity and cultural portability, only card ranking was selected for pilot testing. It is possible that other approaches for ranking severity qualifiers may be preferred by children in some countries or populations. However, for this project, a single approach, acceptable to children in a range of countries, was sought by the VMC for use in future EQ-5D-Y-5L translations internationally. Thus far, card ranking seems acceptable, translatable, and culturally portable.

5 Conclusion

To our knowledge, this is the first time that such an approach has been developed for possible inclusion within translation and cultural adaptation processes. The card ranking method was developed for a children's MAUI but could potentially be applied to the development or translation of other types of PROMs, and perhaps also to translations of adult's MAUIs. Further research is now underway to investigate the appropriateness and timeliness of the card ranking exercise among adults. We recommend assessment of qualifiers be completed prior to the other parts of the standard cognitive debriefing exercise to avoid prior learning of the intended hierarchical order. The VMC has now incorporated the ranking exercise into the translation protocol for the EQ-5D-Y-5L beta versions, and the approach is being adapted for pre-testing within the adult EQ-5D-5L cognitive debriefing process.

Acknowledgements The authors are grateful to all participants and also to Rosalind de Charro, Claire Gudex, Erica Lubetkin, Nalinie Banarsi, and Bianca Smit who, together with SD, MH, and JJ, are members of the EuroQoL Group's VMC. They were consulted throughout the various stages of the project, and their input was invaluable to

improving the project. We also thank Des Scott and Janine Verstraete for testing the card ranking in South Africa and Titi S. Fitriana for testing in Indonesia, Joel Lockley for designing the graphics, Bernhard Slaap for his support, and the EuroQoL Research Foundation for funding the project.

Author contributions SD, MH and JJ led all stages of the development of the card ranking process presented in this paper. SD and MH were involved in the acquisition of data; JJ and SD synthesised findings from South African and Indonesian reports; EYM was involved in all LSS reviews and feedback. LGN is leading the translation of the beta EQ-5D-Y-5L in Malawi (Chichewa) where the card ranking process has been included. All authors were involved in the interpretation of findings. SD, MH, and JJ prepared the initial draft manuscript. All authors reviewed and contributed to subsequent drafts and approved the final manuscript.

Data availability The data informing the findings of this project are available from the authors, but restrictions apply to the availability of some data, e.g. the cognitive debriefing report for the pilot testing is not publicly available. However, data are available from the authors upon reasonable request and permission from the EuroQoL Research Foundation. The authors confirm that relevant data are included in the manuscript.

Declarations

Funding This project was funded by the EuroQoL Research Foundation.

Conflicts of interest SD, MH, and JJ are members of the EuroQoL Group's VMC (Netherlands). LGN is a doctoral candidate with the Liverpool School of Tropical Medicine (UK) and the Malawi-Liverpool-Wellcome Trust Clinical Research Programme (Malawi). EYM is an employee of RWS Life Sciences.

Ethics approval Ethical approval for the pilot testing in South Africa came from the University of Cape Town's Human Research Ethics Committee (HREC154/2019) and that for the Malawi translation came from the Malawi College of Medicine Research Ethics Committee and the Liverpool School of Tropical Medicine Research Ethics Committee.

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

References

1. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Educ.* 2017;17(4):137–44. <https://doi.org/10.1093/bjaed/mkw060>.

2. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ*. 2019;364:k5267. <https://doi.org/10.1136/bmj.k5267>.
3. Devlin NJ, Appleby J. Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making. London: The King's Fund; 2010.
4. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *J Patient Rep Outcomes*. 2018;2:42. <https://doi.org/10.1186/s41687-018-0061-6>.
5. Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford: Oxford University Press; 2015.
6. Ramos-Goñi JM, Oppe M, Stolk EA, Shah K, Kreimeier S, Rivero-Arias O, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38(7):653–63.
7. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull*. 2010;96(1):5–21. <https://doi.org/10.1093/bmb/ldq033>.
8. Craig BM, Monteiro AL, Herdman M, Santos M. Further evidence on EQ-5D-5L preference inversion: a Brazil/US collaboration. *Qual Life Res*. 2017;26(9):2489–96. <https://doi.org/10.1007/s11136-017-1591-8>.
9. Craig BM, Pickard AS, Rand-Hendriksen K. Do health preferences contradict ordering of EQ-5D labels? *Qual Life Res*. 2015;24(7):1759–65. <https://doi.org/10.1007/s11136-014-0897-z>.
10. Talley C, McKown S. Patient-reported outcome measure translation: an overview. *Med Writing*. 2018;27(4):26–9.
11. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health*. 2005;8(2):94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>.
12. Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. *Value Health*. 2014;17(1):70–6. <https://doi.org/10.1016/j.jval.2013.10.006>.
13. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*. 2010;19(6):875–86.
14. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36. <https://doi.org/10.1007/s11136-011-9903-x>.
15. Kreimeier S, Åström M, Burström K, Egmar AC, Gusi N, Herdman M, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res*. 2019;28(7):1951–61. <https://doi.org/10.1007/s11136-019-02115-x>.
16. Tomlinson D, von Baeyer CL, Stinson JN, Sung L. A systematic review of faces scales for the self-report of pain intensity in children. *Pediatrics*. 2010;126(5):e1168. <https://doi.org/10.1542/peds.2010-1609>.
17. von Baeyer CL. Children's self-reports of pain intensity: scale selection, limitations and interpretation. *Pain Res Manag*. 2006;11:197616. <https://doi.org/10.1155/2006/197616>.
18. Hall L, Hume C, Tazzyman S. Five degrees of happiness: effective smiley face likert scales for evaluating with children. 2016;311–21.
19. Rebok G, Riley A, Forrest C, Starfield B, Green B, Robertson J, et al. Elementary school-aged children's reports of their health: a cognitive interviewing study. *Qual Life Res*. 2001;10(1):59–70. <https://doi.org/10.1023/A:1016693417166>.
20. Ngwira LG et al, editor. Development of the new EQ-5D-Y-5L using the ranking order exercises. 1st EuroQol African Regional Meeting 2020; Cape Town, South Africa.

Authors and Affiliations

Sarah Derrett^{1,2}  · Mike Herdman^{2,3}  · Lucky G. Ngwira^{4,5}  · Elizabeth Yohe Moore⁶ · Jennifer Jelsma^{2,7}

¹ Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand

² The EuroQol Group, Version Management Committee, Rotterdam, Netherlands

³ Office of Health Economics, London, UK

⁴ Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

⁵ Malawi-Liverpool-Wellcome Trust, Blantyre, Malawi

⁶ RWS Life Sciences, Chicago, IL, USA

⁷ Department of Health and Rehabilitation Sciences, University of Cape Town, Cape Town, South Africa