The Patient - Patient-Centered Outcomes Research
https://doi.org/10.1007/s40271-025-00743-9

ORIGINAL RESEARCH ARTICLE

Determining the Content Validity of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D Instruments for Assessing Generic Child and Adolescent Health-Related Quality of Life: A Qualitative Study

Jill Carlton[1] · Philip A. Powell[1] · Andrew Kirkcaldy[1] · Donna Rowen[1]

Accepted: 28 April 2025

Abstract

Background Health technology assessment agencies typically recommend generic measures of health to generate quality-adjusted life-years. Most agencies provide recommendations on which measure to use for adults, whereas few make recommendations for children. Two widely used preference-weighted measures of child and adolescent health that have evidence of good psychometric performance are the EQ-5D-Y-3L and the Child Health Utility 9D Index (CHU9D). The EQ-5D-5L has also been used to assess adolescent health. However, evidence on their content validity—a core measurement property—is limited. The objective of this study was to explore the content validity of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D measures, including their relevance, comprehensiveness, and comprehensibility.

Methods We assessed the content validity of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D using online semi-structured cognitive interviews in the UK. Participants were asked to comment on the relevance, comprehensibility, and comprehensiveness of the measures, including response options, recall period, and completion instructions. Interviews were informed by a topic guide. Purposive sampling allowed for appropriate breadth in the sample, with variation in gender, and presence of health conditions, disease, or disability. Interviews were recorded and transcribed verbatim before thematic content analysis.

Results In total, we conducted 49 interviews between August 2022 and June 2023: 21 children/adolescents aged 8–17 years and 28 parents/guardians of children aged 4–17 years. The mean duration of the interviews was 45 min. Relevance was broadly supported, but issues were identified. Comprehensibility was inconsistent on some items, and participants expressed difficulty with grouped items (e.g., 'anxiety/depression'). Participants had difficulty distinguishing qualitatively between some response options (e.g., 'a little bit/a bit'). Some participants noted that instrument comprehensiveness was insufficient.

Conclusions Although the content of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D was broadly supported, potential problems were identified in aspects of comprehensibility, relevance, and comprehensiveness. These present opportunities for future research and refinement to ultimately improve the content validity of these measures for assessing child and adolescent health.

1 Introduction

Although the UK National Institute for Health and Care Excellence (NICE) typically recommends the EQ-5D to generate utilities and quality-adjusted life-years for adults in England and Wales [1], there is no such recommended measure for children. Despite this, evidence shows that the EQ-5D, an adult measure with two versions (the three-level

Key Points for Decision Makers

Evidence for the content validity of three commonly used measures for generating child utilities is limited.

This study found that the content of the EQ-5D-5L, EQ-5D-Y-3L, and Child Health Utility 9D Index (CHU9D) instruments is supported but that the measures are not without issues. Potential problems were identified in aspects of comprehensibility, relevance, and comprehensiveness.

Refinements to the content of the measures may be beneficial to enhance content validity to better reflect child/adolescent health-related quality of life.

✉ Jill Carlton
j.carlton@sheffield.ac.uk

[1] Sheffield Centre for Health and Related Research (SCHARR), School of Medicine and Population Health, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

Published online: 16 May 2025

EQ-5D-3L and the five-level EQ-5D-5L), is the most commonly used to generate utility values for children and adolescents in health technology assessment submissions to NICE [2]. This suggests that the utility values of adult health states are often assumed to be appropriate for estimating the utility values of child and adolescent health states, which may not hold true [2]. Accordingly, several potential candidate generic preference-weighted measures (PWMs) have been developed to generate utilities specifically in children/adolescents. These include the EQ-5D-Y-3L, which was developed as an adaptation of the EQ-5D-3L for children [3], and the Child Health Utility 9D Index (CHU9D), which was developed *de novo* with children aged 7–11 years [4].

When deciding on which PWM to use to generate child utilities, one key consideration is evidence of good measurement properties, which apply to all outcome measures [5]. These include quantitative indicators of psychometric performance, including construct validity, reliability, and responsiveness as well as robust qualitative evidence of content validity (i.e., "the degree to which the content of an HR-PRO [health-related patient-reported outcome] instrument is an adequate reflection of the construct to be measured") [6]. There is emerging evidence of good performance of child- and adolescent-specific PWMs that can be used to inform UK policy, including the EQ-5D-Y-3L and CHU9D, across several psychometric properties [7–9]. Existing content validity evidence for the use of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D in children and adolescents is limited across their entire intended age range. It consists of some initial evidence of selected aspects of content validity during the development of the measures [3, 4] and subsequent isolated studies in limited populations, such as children with specific health conditions (e.g., Blackmore et al. [10]). This is a notable omission, given that content validity is considered the most important and foundational measurement property of an outcome measure [11].

Content validity can be evidenced during the development of outcome measures by ensuring concept elicitation and subsequent cognitive interviewing is undertaken with participants representative of the target population. Further, in best practice, it is assumed that this evidence of content validity should be supplemented (and confirmed) with additional, independent studies undertaken in the population of interest using robust cognitive interviewing methods [11]. Appropriate qualitative techniques are used to assess the relevance (i.e., whether the items, recall period, and response options are relevant to the intended population), comprehensibility (i.e., whether all aspects of the measure including instructions, items, and response options are understood as intended), and comprehensiveness (i.e., whether all important aspects of the target construct are being assessed) of measures [11]. Guidelines on assessing and documenting evidence of the content validity of outcome measures – of which PWMs are a special case – have been developed [12, 13].

As highlighted in a recent systematic review [14], existing content validity evidence for the use of the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D in children/adolescents is limited. It consists of some initial evidence of selected aspects of content validity during the development of the measures [3, 4] and subsequent isolated studies in limited populations, such as children with specific health conditions [10]. To the authors' knowledge, no study has independently and comprehensively assessed the content validity of the above-mentioned PWMs as a generic measure across a range of children with and without common health complaints.

The aim of this project was to assess the content validity of three PWMs available for use in child/adolescent populations (EQ-5D-Y-3L, EQ-5D-5L, and CHU9D), via self-report and proxy report, across their intended age range (i.e., self-report covering 8–17 years for the CHU9D and EQ-5D-Y-3L and covering 12–17 years for the EQ-5D-5L; and proxy-report covering 4–17 years for the CHU9D and EQ-5D-Y-3L and covering 12–17 years for the EQ-5D-5L). Comprehensive and robust semi-structured qualitative methods were employed that conform to best practice standards for cognitive interviewing [12]. Outcome measurement is never definitive, but rather an iterative process of gradual refinement. Therefore, it was expected that the evidence from this study would identify aspects of good content validity of the measures and yield recommendations for future improvement and research.

2 Methods

We conducted semi-structured interviews with participants to assess the content validity of the three measures. Content validity requires subjective assessment from the target population, including understanding of patient-reported outcome measure (PROM) content, assessing its relevance, and identifying whether important content is missing, so qualitative methods are particularly suitable for this purpose [11]. We followed the Standards for Reporting Qualitative Research (SRQR) checklist in the production of this manuscript.

2.1 Description of Measures

2.1.1 CHU9D

The CHU9D is a generic PWM developed in the UK. It contains nine dimensions: worry, sadness, pain, tiredness, annoyance, school, sleep, daily routine, and activities. Each item is assessed across five severity levels. The instrument was developed using 'bottom-up' methodology, involving

qualitative research with children aged 7–11 years. More than 70 children with a range of health problems were interviewed to identify potential dimensions [15]. Further ranking work with children was undertaken to develop the response option scales and produce a draft descriptive system [16]. The final descriptive system was derived following a psychometric survey with children recruited from schools and hospital settings [17]. Although it was originally designed for use in children aged 7–11 years, subsequent studies have demonstrated validity in adolescents aged 12–18 years [18] and for completion via parent/guardian proxy for children aged 4–7 years [19]. The instrument has been translated into several languages, and value sets exist for multiple countries (e.g., Stevens [20] and Rowen et al. [21]).

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

The EQ-5D-5L instrument consists of five dimensions with five response levels and is intended for use with adults. However, there is recognition that the measure can be used with participants aged ≥12 years. The measure has five dimensions covering mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each with five severity levels. The EQ-5D-Y-3L version was generated by adapting the adult version of the EQ-5D-3L [3], with minor changes to wording to ensure both relevance and clarity for younger respondents [22]. The instrument has five dimensions and three severity levels and has been translated into several languages. Both measures have 100-point visual analogue scales (VAS) measuring overall health (EQ VAS). Although no current value sets for the Y-version (either 3-level or 5-level) exist in the UK, there are published value sets for the 3-level Y-version in multiple countries (e.g., Prevolnik Rupel et al. [23], Rencz et al. [24]) (see File A in the electronic supplementary material [ESM]).

2.2 Screening Survey

A short screening survey (hosted on Qualtrics) was designed to facilitate purposive sampling to cover a range of predetermined characteristics for the study. It included fixed-response questions covering the child's gender, ethnicity, age, and presence of a long-standing physical or mental impairment, illness, or disability self-reported as diagnosed by a doctor. 'Long-standing' was defined as anything that had troubled or was likely to trouble them over a period of at least 12 months. The survey also included questions about the parents'/caregivers' genders, ethnicity, age, and presence of a long-standing physical or mental impairment, illness, or disability. Sampling was determined based on the characteristics of the child/adolescent rather than of the parent/caregiver.

2.3 Sample Size

The study was designed to include 49 online interviews, 21 with children/adolescents and 28 with parents/caregivers. This sample size was based on COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidance that each PROM is assessed in qualitative content validity interviews with a minimum of seven participants [13] after stratifying for age of child and measures to be used in the interview (Table 1). All the measures assessed are recommended as appropriate by the instrument developers for the age group of the children concerned. This meant that the EQ-5D-5L was included for children aged 12–17 years and their parents/guardians. The EQ-5D-Y and CHU9D were used for children aged 8–17 years and all parents/guardians of children aged 4–17 years.

2.4 Recruitment and Consent

Participants were recruited from multiple sources, which included a range of patient organisations across different conditions (File B in the ESM). Patient organisations advertised the study via social media, newsletters, and/or emails. Study advertisements included a link to a short screening survey. The study was also advertised on the university staff volunteers email distribution list. Towards the end of the study, a recruitment company was contracted to obtain the final sample.

A sampling frame was used to ensure an appropriate breadth in the child and adolescent samples across age groups. Effort was made to ensure the sample included a range of genders and underlying child health conditions, including both physical and mental health problems (File C in the ESM). The research team purposively sampled participants from the information provided via the short screening survey. The research team contacted potential participants to schedule a mutually convenient interview date/time. A confirmation email was sent to the participant, along with copies of the appropriate PWMs, which included a link to an online consent form. Online consent was completed before any interview took place. Participants were encouraged to look at the PWMs before the interview.

2.5 Interview Process

The interviews were conducted online via Google Meet. Online interviews were chosen to help enable inclusivity across geographical locations, increasing the validity and value of the research. Two qualitative researchers with experience assessing content validity in outcome measures conducted the interviews. The researchers had no prior relationship with the participants. Parents/caregivers were given the option of being present during any interview with

Table 1 Participant demographics

Population	Age of child (years)	Instruments	Male	Female	Physical[a]	Mental[a]	N
Children/adolescents	8–11	EQ-5D-Y, CHU9D	4	2	1	4	1
	12–14	EQ-5D-Y, CHU9D, EQ-5D-5L	4	4	5	3	0
	15–17	EQ-5D-Y, EQ-5D-5L, CHU9D	4	3	6	0	1
Total			**12**	**9**	**12**	**7**	**2**
Parent/guardian	4–7	EQ-5D-Y, CHU9D	4	2	3	1	2
	8–11	EQ-5D-Y, CHU9D	2	6	4	1	2
	12–14	EQ-5D-Y, EQ-5D-5L, CHU9D	2	5	0	5	2
	15–17	EQ-5D-Y, EQ-5D-5L, CHU9D	2	5	6	2	0
Total			**10**	**18**	**13**	**9**	**6**

CHU9D=Child Health Utility 9D Index; Physical=physical impairment, illness, or disability; Mental=mental impairment, illness, or disability; No=no physical or mental impairment, illness, or disability

[a]Non-mutually exclusive

their child (i.e., <16 years). If parents/caregivers chose not to be present (i.e., in the same room), they were encouraged to be close in the event of any problem. A topic guide (File D in the ESM), based on current best practice guidance for assessing content validity (i.e., assessing relevance, comprehensibility, and comprehensiveness) [13] was used to inform the content of the interviews. Participants were also asked which measure they preferred. Interviews were recorded, transcribed verbatim, and checked for accuracy before anonymisation. The sequence in which the measures were discussed was randomised for each interview to mitigate for any order effects related to interviewee fatigue. All participants were reimbursed for their time with an online Amazon voucher worth £25.

2.6 Analysis

Transcripts were subjected to thematic content analysis using an a priori framework, following an approach developed by the National Centre for Social Research for the analysis of qualitative data [25]. The framework approach combines multiple aspects of different paradigms in qualitative research and has been considered a 'whole paradigm approach' [26]. The initial framework applied was informed by existing standards of assessment of content validity [13] and included the main themes 'relevance', 'comprehensiveness', and 'comprehensibility'. Two experienced qualitative researchers, one of whom had conducted the interviews, coded the transcripts using NVivo. To enhance trustworthiness, the first five transcripts (10%) were independently dual coded, after which coding and refinement of the analytic framework was discussed. The final framework was applied across all transcripts before analysis. Ten transcripts were dual-coded (20%).

2.7 Ethical Approval

Ethical approval was obtained from the University of Sheffield (ref: 045650). All participants gave informed consent before interviews were undertaken.

3 Results

3.1 Participant Sample

Interviews (n=49) were conducted between 15 August 2022 and 28 June 2023. Participant details are shown in Table 1. The mean duration of interviews (not including consent and background procedures) was 44 min 43 secs (range 22 mins 54 secs to 16 mins 42 secs).

3.2 CHU9D

Table 2 provides a summary of the comprehensibility and relevance results for the CHU9D, including illustrative supporting quotes. Further quotations supporting the findings are provided in File E in the ESM. Most adults and children adequately understood the questionnaire instructions. Participants had mixed views on the usefulness of the example question in the instructions, with most of those who commented expressing that the example was redundant and/or patronising for adults or older children. The recall period of 'today' appeared to be understood as the day on which participants were completing the questionnaire. Responses on the relevance, or appropriateness, of using 'today' as a recall period were mixed, with a slight majority of both adults and children who identified problems. These included a perceived difficulty

in completing the questionnaire if it was administered earlier (vs. later) in the day and the timeframe being too short to get an accurate impression of health-related quality of life (HRQoL). The most common alternative suggestion was a week. Proponents of 'today' as a recall period noted that it made things easier to remember. Problems understanding and distinguishing between response options in the CHU9D were identified for both adults and children. Specifically, participants struggled to distinguish between 'a little bit' and 'a bit', with a few also noting problems distinguishing between these and 'quite' or 'some' and 'a few'. Although people had different views on the relevance and number of response options, of those who commented, most adults preferred five options, whereas most children stated that five was too many and preferred fewer.

Question one ('worried') and question two ('sad') were well understood by almost all adults and children, who identified that they were asking about how 'worried' or 'sad' they (or their child) felt (today). For many, worry was distinguished from sadness as being typically about a future event (rather than something that had happened). However, a few children and one adult expressed how sadness and worry overlapped conceptually. Synonyms provided by some participants for worry included 'anxiety', 'fear', 'uneasiness', and 'scared'. Synonyms for sadness included 'upset', 'unhappy', and 'feeling down'. Some participants noted that 'worry' was too general. Most adults and children who commented stated that the questions were relevant. However, a couple of adults noted that worry may be less relevant for younger (vs. older) children.

Question three ('pain') was understood by children and adults as referring to physical and/or mental (or emotional) pain. Most participants interpreted the item as referring to physical pain. However, a non-trivial minority thought of emotional pain first, or in combination with physical pain and suggested a need for clarification. These interpretations were guided by the position of the item in the questionnaire (after 'worried' and 'sad'). Synonyms for 'pain' included 'hurt' and 'discomfort'. Most children and adults who commented regarded this as a relevant question.

Question four ('tired') was well understood by both adults and children. Participants referred to both physical and mental tiredness, and synonyms included 'lethargic', 'worn out', 'low in energy'. Overlap was noted with 'sleep', with a few participants, and particularly children, referring to sleep (or being sleepy) in their interpretation of the question. Most participants thought the question was relevant, but a few adults questioned the relevance of tiredness to HRQoL, given that it is normal in certain situations (e.g., being physically tired after being physically active, or being a teenager).

Almost all participants demonstrated a conceptually appropriate understanding of question five ('annoyed'), with common synonyms including 'angry' and 'frustrated'. However, a proportion of adults questioned the appropriateness of the word, particularly for younger children, suggesting that 'angry' may be more appropriate. Most adults and children identified this as a relevant question, but approximately one-third expressed concern about its relevance to health. This was partly due to annoyed being understood as a situational emotion, triggered by a multitude of non-health related events, and that did not always persist over time.

Most adults and children demonstrated an appropriate level of understanding for question six ('schoolwork/homework'), with most participants interpreting it as problems doing schoolwork or homework. However, different people interpreted the reason(s) underlying any problems differently, including academic ability to do the work; situational problems concentrating or focusing; motivation; or physical health problems preventing people from doing the work. One adult questioned how they would answer if it was not a school day (e.g., at the weekend). Despite differences in interpretation, the question was almost universally identified as relevant.

Question seven ('sleep') was clear to most adults and children and was understood by most to cover problems with sleep in general (i.e., sleep quality), although interpretations of the focus varied (e.g., on getting to sleep, amount of sleep, etc.). The conceptual overlap with 'tired' was noted, and a few adults/children suggested having one question that combined these items. However, most participants viewed them both as relevant questions to include.

Question eight ('daily routine') was understood consistently by most adults and children as being about daily tasks and activities. Ambiguity was identified in whether the item was referring to completing daily tasks independently or with assistance, and parents noted that the meaning of the question would vary according to the child's age. Some adults suggested separating 'eating' from 'self-care' as distinct activities. A few children suggested adding 'school' and related activities to the examples provided. Most participants identified this question as relevant.

Most adults identified question nine ('able to join in activities') as being about participation and social activities, but some noted that the question fuses two concepts: being able to do sports (and related activities) and joining in or being social. This mixed interpretation was also evident in child interviewees, a few of whom focused on doing activities/sports rather than identifying the social element of the question. Nevertheless, many children identified the question as being about joining in. Most participants thought this question was relevant.

Approximately half of participants found the questionnaire comprehensive. Of the remaining participants, two adults and two children recommended including positive emotion content (i.e., 'happy'); two adults and three children

Table 2 Summary of comprehensibility and relevance results: CHU9D

Component	Comprehensibility	Relevance	Illustrative quote
Instructions	Most adults and children understood instructions as intended. Mixed views were expressed on the appropriateness of the worked example, with most comments noting it was redundant and even patronising for adults/older children	NA	"It's asking me how I am today. Or yeah, just how you're feeling today. And tick one box for each question, and yeah, you just got to like put a tick in the box for what you feel like, I guess." (P45, child, male, 12–14 years, physical)
Recall period	'Today' was interpreted and understood as the current day that participants were completing the questionnaire (as opposed to yesterday or tomorrow)	Slight majority of adults and children identified problems with 'today', including difficulty in responding whether the questionnaire was administered earlier (vs. later) in the day and alternative preference for a longer recall period, such as a week	"That's today, as in right now, as opposed to yesterday or tomorrow." (P1, parent/guardian, female, 8–11 years, physical)
Response options	Problems distinguishing qualitatively between 'a little bit' and 'a bit' for many adults and children. Minority had problems distinguishing between 'a bit' and 'quite' and between 'some' and 'a few'	Most adults thought the five response options were relevant and appropriate. Most children preferred fewer response options, such as three	"I struggle with 'a little bit' and 'a bit'. I would find it hard to give an example between him being 'a little bit' tired and 'a bit' tired." (P37, parent/guardian, female, 12–14 years, mental)
Q1: Worried	Well understood by almost all adults and children as being about how 'worried' they or their child felt. Typically described in relation to a future situation/event. 'Worry' was synonymous with feeling anxious or having feelings of fear and/or panic. Some participants noted the question was too general and did not specify what the child was worried about	Most adults and children felt the item was relevant but may be age dependent	"Worried, it's kind of asking if anything's bothering you, you feel uneasy at all, anxious." (P42, child, female, 15–17 years, physical)
Q2: Sad	Well understood by almost all adults and children as being about how 'sad' they/their child felt. Typically described in relation to a past or current situation or event. 'Sad' was synonymous with upset, unhappy, or feeling down	Most adults and children felt the item was relevant. Some participants, particularly children, questioned the relevance of 'sad' as a separate question due to the conceptual overlap with 'worried'	"It's asking you if you feel sad, if you're feeling upset, anything bothering you, you're feeling down." (P42, child, female, 15–17 years, physical)
Q3: Pain	Most adults and children interpreted the item to be referring to physical pain. A non-trivial minority thought of emotional pain first or in combination with physical pain. 'Pain' was considered synonymous with 'hurt' or 'discomfort'	Most children and adults regarded this as relevant, with some noting that it would be relevant to assess both physical and emotional pain	"It is all aspects for me. For my child it would be physical pain … She's too young [for] psychological pain. But she's just like, OK, it's hurting, you know." (P3, parent/guardian, female, 8–11 years, no)
Q4: Tired	Well understood by adults and children. Inferred as both physical and mental tiredness, with synonyms of 'lethargic', 'worn out', 'low in energy'. Interpretative overlap with 'sleep' (Q7), particularly for children, with item defined in relation to sleep (or being 'sleepy')	Relevant to most participants. However, a few adults questioned whether increased tiredness was related to worse health-related quality of life, given it is a normal experience in certain situations	"I think just not having enough sleep." (P35, child, female, 12–14 years, mental)
Q5: Annoyed	Conceptually understood by almost all adults and children. Synonymous with feelings of 'frustration', 'anger', 'agitation', or 'irritation'. Understood as a situational feeling state	Relevant to most adults and children but identified as not relevant to approximately one-third	"On that scale how, how annoyed do you feel … For me it's just like being frustrated, people, things, events." (P46, child, male, 8–11 years, mental)

Table 2 (continued)

Component	Comprehensibility	Relevance	Illustrative quote
Q6: Schoolwork/Homework	Understood appropriately by many adults and children. Problems with schoolwork or homework, could refer to (academic) ability, problems concentrating and/or focusing, motivation, or health problems preventing work	Almost universally identified as relevant to children/adolescents	"I think it's asking if you have any problems with homework, schoolwork, so you have problems writing, during the lessons, reading, all that." (P42, child, female, 15–17 years, physical)
Q7: Sleep	Clear to most adults and children. Interpreted as problems with sleep (in general), which could encompass a range of aspects (e.g., getting to sleep, staying asleep, sleep quality, etc.)	Most children and adults viewed this as relevant. Overlap with 'tired' was noted, but most people viewed them as sufficiently distinct	"Like the quality of sleep people are getting, whether they're waking up, whether they aren't like falling asleep easily." (P15, child, female, 15–17 years, physical)
Q8: Daily routine	Understood as being about daily tasks/activities. Unclear whether the question referred to doing daily tasks independently or with assistance (e.g. from parents)	Most participants identified this question as relevant. A few adults suggested separating 'eating' from 'self-care' as distinct activities	"So, impact on daily routine: eating, having a bath, shower, getting dressed. It's sort of independent living, feeling independent, which is naturally what children and young people are trying to become." (P23, parent/guardian, female, 15–17 years, physical/mental)
Q9: Join in activities	Most adults and children identified this item to be about participation and joining in with activities. However, some children failed to identify the social element of the question, instead focusing on whether they could do activities per se (including sports and reading)	Considered relevant to most adults and children who commented on this question	"It kind of fuses two things. There's the doing sports and the being physical, should be separated out from the being with your friends. Being with your friends is an important thing. But being physical is another important thing. And they've put the two together." (P13, parent/guardian, female, 8–11 years, physical)

Mental=mental impairment, illness, or disability; NA=not applicable; no=no physical or mental impairment, illness, or disability; Physical/mental=physical and mental impairment, illness, or disability; physical=physical impairment, illness, or disability

expressed that they would like to see additional social content (e.g., relationships, friendships, communication, social issues); two adults identified a need for a separate diet/eating item; two adults and two children mentioned the influence of social media/screen time; and one adult and two children suggested an overall HRQoL question. Other isolated examples were raised by one participant only (e.g., maintaining focus, feeling overwhelmed, anger, etc.).

Finally, a few parents expressed that they would have difficulty in answering some of the questions on the CHU9D (as a proxy) without talking to their child. This was particularly, but not exclusively, mentioned in relation to the subjective feeling state questions (Q1–5).

3.3 EQ-5D-5L/EQ-5D-Y-3L

Table 3 presents a summary of the comprehensibility and relevance results for the EQ-5D instruments with supporting quotes. Further quotations supporting the findings are provided in File F in the ESM. Almost all adults understood the instructions on the EQ-5D-5L/EQ-5D-Y-3L, but a few children/adolescents found it more difficult to articulate clearly what they had to do on the questionnaire overall. The recall period of 'today' appeared to be interpreted as intended by adults and children. However, the relevance of the recall period was mixed, with most participants preferring a longer recall period as this was more 'representative' of their/their child's health. Those who were in favour of retaining 'today' referred to a greater perceived reliability of recall. On the EQ-5D-5L, most adults and adolescents thought that the five-response options were relevant and appropriate for the scale. However, there were problems in distinguishing the response options qualitatively, with most adolescents who commented expressing difficulty in distinguishing between 'severe' and 'extreme'. Several adolescents and adults also questioned the distinction between 'slight' and 'moderate'. Many adults who commented expressed that they could distinguish between the response options on the scale. On the EQ-5D-Y-3L, a slight majority of adults and children found the three-level response options relevant and appropriate, but some children and more adults expressed a need for more response options for finer granularity. A few adults were critical of the three-level response scale, stating that most people would choose the middle. Most people who commented on the issue were able to distinguish between the response levels.

For question one ('mobility'), almost all participants correctly understood this question as being about the ability to walk. Some adults questioned whether the question meant walking independently or with aids, as this was not specified. It was viewed as relevant by almost all adults and children. However, a few adults and children expressed that 'mobility' is more than walking, particularly for children who also like to run or for specific health conditions (e.g., for those in a wheelchair). This meant that some participants questioned the concordance between the label of 'mobility' and the question itself.

Question two ('self-care/looking after myself') was understood predominantly in terms of personal care, with a focus on washing and dressing. However, a range of reasons underlying potential problems in washing and dressing were expressed (e.g., both physical and mental health problems, including – in one instance – a lack of fashion sense). Further, the use of a label that differed from the possible answers (i.e., washing and dressing) created some confusion, and several adults and children viewed the question as being about more than just washing and dressing, such as feeding yourself or emotional self-care. Those who saw both questionnaires had mixed views on whether 'self-care' or 'looking after myself' was superior as a label, with most favouring the latter, as it was simpler to understand and better matched the response options. Almost all adults and children viewed the question as relevant, but it was noted that relevance would differ by age.

Most adults and adolescents understood question three ('usual activities/doing usual activities') on both questionnaires as involving activities that someone would do day-to-day. Participants differed in the extent they interpreted this question to be including physical or social activities. Considering the examples provided, some adults described the question as too 'blunt' or a 'catch-all' as a child may be able to do some activities but not others. Of those who saw both questionnaires, adults and adolescents preferred the examples given on the EQ-5D-Y-3L as they were more relevant to childhood, and they noted that some examples on the EQ-5D-5L (e.g., work, housework) were too 'adult'. One adolescent did not understand the term 'leisure activities'. For EQ-5D-Y-3L, both adults and children considered the item relevant. For the EQ-5D-5L, all adolescents who commented acknowledged its relevance, but adults were concerned about the relevance to adolescents of the examples provided.

Question four ('pain/discomfort/having pain or discomfort') was understood in different ways by participants. Although many adults and children interpreted 'pain' as physical pain (with synonyms such as 'hurt'), 'discomfort' was viewed as distinct from 'pain' and interpreted by many as having both an emotional component (i.e., feeling 'unease') and a physical aspect. Most participants who commented on the item viewed it as relevant.

Question five ('anxiety/depression/feeling worried, sad or unhappy') on the EQ-5D-5L was understood as assessing anxiety (worry) and depression (low mood), but there was ambiguity about whether this referred to a medical diagnosis of anxiety and/or depression or a descriptive feeling state. The same question in the EQ-5D-Y was

Table 3 Summary of comprehensibility and relevance results: EQ-5D-5L and EQ-5D-Y-3L

Component	Comprehensibility	Relevance	Illustrative quotes
Instructions	Instructions on 5L/Y-3L understood well by almost all adults. Several children/adolescents struggled to articulate clearly what they were meant to do on the questionnaire	NA	"I guess it's just asking me to think about my child's health and tick what I think about their health today, in terms of these areas." (P37, parent/guardian, female, 12–14 years, mental, 5L)
Recall period	Recall period of 'today' appeared to be interpreted as intended	Mixed preferences. Most participants preferred a longer recall period (e.g., a week) as more representative	"Maybe put it like, not 'today', maybe put it in a month. Because, I think these all these questions like, 'today', because maybe today you don't feel pain. But tomorrow you can feel pain. And same with anxiety and depression. Because, for example, maybe your anxiety's OK, but then tomorrow you can have, a panic attack." (P19, child, female, 15–17 years, physical, 5L)
Response options	Problems distinguishing qualitatively between 'severe' and 'extreme' and – to a lesser extent – 'slight' and 'moderate' on the 5L, particularly for adolescents. Most adults could distinguish between options on the 5L. Most participants were able to distinguish response levels on the Y-3L	Most adults and adolescents found response scales relevant and appropriate on both instruments. Some children and more adults would prefer more than three response choices for the Y-3L	"The three I would say is a lot more simpler and straightforward, I mean I guess it's less repetitive as well because 'severe' and 'extreme' they do sound the same and I feel they do have the same impact, but 'some' or and 'a lot' they are different, you can tell the difference." (P47, child, male, 12–14 years, mental, Y-3L)
Q1: Mobility/mobility (walking about)	Correctly understood as being about the ability to walk by almost all adults and children. Mobility was considered as more than just walking. Unclear whether the question referred to walking independently or with aids	Considered relevant by almost all adults and children. However, 'mobility' can be more than the ability to walk (e.g., running, bodily movement)	"I'd say mobility and walking about, mobility is the title but walking about is maybe something different. Mobility obviously includes walking about, but it would probably also include arms. Walking about is very much focused on walking." (P8, child, female, 15–17 years, no, 5L)
Q2: Self-care/looking after myself	Understood by most as being about personal care, with a focus on washing/dressing. Difference between labels and response options created confusion, with a few adults and children viewing the question as more than just washing/dressing, including feeding oneself or emotional self-care. 'Looking after myself' was considered superior to 'self-care' as a simpler label and more consistent with response options	Viewed as relevant by almost all participants, but relevance may differ by age	"So, looking after yourself, like if you're looking after your body, like having showers, baths. If you're looking after your hair or skincare. Or if you're giving enough self to have some me time, to have some alone time." (P19, child, female, 15–17 years, physical, Y-3L)

Table 3 (continued)

Component	Comprehensibility	Relevance	Illustrative quotes
Q3: Usual activities/doing usual activities	Understood as day-to-day activities by most adults and adolescents. Participants differed in the extent to which the question was appraised as having a social dimension	Y-3L question considered relevant by adults and children. Concern from adults on relevance of examples used in 5L for adolescents	"Well, it's asking is the child able to do their usual activities and whether they've got, and if so, are there no problems doing those activities, or some or a lot and then it gives a bit of context I guess by including the word in brackets there, going to school, hobbies, sports, etcetera." (P28, parent/guardian, male, 15–17 years, physical, Y-3L)
Q4: Pain/discomfort/having pain or discomfort	'Pain' and 'discomfort' interpreted as distinct. Former perceived as physical by most adults and children. Latter perceived as having an emotional component (i.e., feeling uncomfortable) by many	Noted to be relevant by most participants who commented	"I guess pain I'd describe as something that hurts, discomfort would be more around kind of, I find that harder actually, discomfort would be more, like not necessarily saying something hurts or, I'd link pain probably with a specific issue as well, like my knee hurts, my head hurts, whereas discomfort I'd see as a more general thing that they may be feeling just generally uncomfortable" (P37, parent/guardian, female, 12–14 years, mental, 5L)
Q5: Anxiety/depression/worried, sad, or unhappy	Well understood by most adults and children. Ambiguity about whether 'anxiety/depression' in 5L referred to a medical diagnosis. 'Anxiety/worry' seen as conceptually distinct and separable from 'depression/sad or unhappy'	Most participants viewed the item as relevant	"Yeah, anxious means– I normally think of nervousness. But anxiousness, I normally associate it with nervousness. I would view anxiousness and depression as very different things. Anxiousness is more related to nervousness and you're more hesitant in what you do. Whereas depression is more based on sadness and just unhappiness." (P27, child, male, 15–17 years, physical, 5L)

Mental=mental impairment, illness, or disability; NA=not applicable; no=no physical or mental impairment, illness, or disability; physical=physical impairment, illness, or disability; 5L=EQ-5D-5L; Y-3L=EQ-5D-Y-3L

generally well-understood by adults and children. On both versions, anxiety/worry was understood as conceptually distinct from depression/sad or unhappy, and a slight majority of participants would prefer to see separate items for these questions, whereas others identified them as sufficiently similarly assessing emotional distress. Of those who saw both questionnaires, most preferred 'worried, sad, or unhappy' (for child health) as it was less medicalised and did not imply a mental health diagnosis. Most participants viewed the item as relevant.

Views on the EQ VAS were mixed. Almost all participants understood the task and what they had to do clearly. However, some participants questioned what was meant by 'health' and the terms 'worst health' and 'best health'. People differed in the extent they focused on physical and/or mental health when interpreting the EQ VAS, and some advocated for having separate questions focusing on these aspects of health. Understanding was concerning for three younger children (8–12 years) completing the EQ-5D-Y-3L, who referred to healthy/unhealthy behaviours informing their answer, such as whether they had eaten sugar or unhealthy food that day or whether they were doing things they liked (e.g., playing with friends online).

Regarding the comprehensiveness of the EQ-5D instruments, one of the most common aspects noted as missing was a social dimension (including relationships, friends, and family), which was mentioned by four adults and three children. Similarly, a question on eating/diet was mentioned by four adults and three children. Including a discrete question on school was viewed as important by six people (three adults and three children). Finally, for the EQ-5D-5L, three people (one adult and two adolescents) expressed a desire for a question on happiness (or sadness), with a particular focus on positive emotions. Similarly to the CHU9D, a few adults noted that they may find more subjective questions difficult to answer as a proxy (Q4–5).

3.4 Preference of instruments

Results on preference were heterogenous and differed according to respondent and the order of exposure to the instruments. Children/adolescents who commented upon the EQ-5D-Y-3L and CHU9D instruments did not indicate a clear preference. Adult participants who commented only upon the EQ-5D-Y-3L and CHU9D instruments favoured the CHU9D. Children/adolescents who commented upon all three instruments favoured the CHU9D, whereas there was no clear preference across adult participants. The results are shown in File G in the ESM.

4 Discussion

This study is the first of its kind to explore in a structured manner the content validity of three commonly used generic PWMs in this age group. Although the instruments performed well in some regards, there were areas that could be actioned to improve content validity for both adults and children.

4.1 Comprehensibility

Full assessment of comprehensibility was limited in this study as no published documentation of intended item constructs/meanings is available for either the CHU9D or the EQ instruments. However, it could be argued that some concepts are self-explanatory (e.g., 'sad') and that terms were often understood in a comparable way by participants. Beyond this, the study suggests that there are some notable differences in interpretation across items within all instruments, which could benefit from clarification. For example, 'pain' in the CHU9D and 'pain/discomfort' in the EQ instruments were interpreted heterogeneously as encompassing physical and/or emotional pain/discomfort. 'School work/homework' in the CHU9D could be interpreted as the ability to do them in an educational sense (i.e., level of attainment, whereby a problem may relate to whether an individual has difficulty in completing them), the ability to undertake them (i.e., concentration), or the willingness to do them (i.e., motivation, enjoyment, etc.). The EQ VAS was interpreted as referring to either physical health or mental health or both. Additional discordance was identified between some of the labels in the EQ instruments and the item content, such as 'mobility' referring to wider bodily movement than just walking. Further, item meaning across instruments was not always considered comparable (e.g., 'anxiety/depression' referring to mental health diagnoses vs. 'worried, sad or unhappy'). Problems were also observed when items combined two or more discrete constructs within an item (e.g., anxiety/depression). Combining two constructs in one item is not recommended in best practice criteria for developing PWMs [27]. Finally, problems were observed in qualitatively distinguishing between response options (e.g., 'a little bit' and 'a bit' in the CHU9D; 'severe' and 'extreme' in the EQ-5D-5L). Similar issues with comprehension in the EQ-5D-Y-3L measure have been highlighted elsewhere [28].

Guidance from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force [29] notes the importance of establishing and reporting evidence of content validity in PROs for medical product evaluation. The EQ-5D and/or CHU9D instruments have been used in studies assessing new drugs or interventions without this evidence being formally well-established. It must be

noted that the guidance was published after the development of the instruments included in this study. This study has demonstrated that both children/adolescents and parent/guardians interpret item meanings differently. Heterogeneity in the interpretation of items means that they may not be measuring the same underlying construct across participants. However, the practical implication of this finding is not currently known, and further research is recommended.

4.2 Relevance

Consideration of item relevance can be closely associated with an individual's understanding and interpretation of the item itself (i.e., comprehensibility). This was the case for some item content explored within this study. Perhaps predictably, this study identified that individuals have differing perspectives on item relevance. It is not inconceivable that their views were influenced by the presence of their (or their child's) health condition(s). For example, a person with a mental health condition, illness, or disability, such as autism, may view an item on 'pain' to be irrelevant. However, some participants expressed views beyond their (or their child's) health condition(s) and considered child/adolescent HRQoL more holistically (i.e., beyond individual experiences). It could be argued that endorsement of relevance of all items by all participants within any instrument is unrealistic. Despite this, there was endorsement of items included in each of the instruments from both adult and children/adolescents alike. Where item relevance was questioned, it related to potentially overlapping constructs, an issue more prevalent within the CHU9D measure. Related constructs (such as 'tired' and 'sleep', and 'worried' and 'sad') led individuals to question the need to include both.

The relevance/appropriateness of the recall period of 'today' was identified as problematic in both questionnaires. Specifically, participants' concerns related to the acute nature of the recall period and the use of the word 'today' relative to time of administration (i.e., near the start vs. the end of the day). Other PWMs have different recall periods, such as the 1-week or 4-week versions of the SF-6D [30], and a recent review suggested that different recall periods may provide different HRQoL information [31]. Whereas a 1-day recall is a viable option for further testing, the results suggest that rephrasing 'today' to 'past 24 hours' may help mitigate concerns about the time of administration of the measures.

A final area for consideration is the relevance of certain more 'subjective' questions for proxy administration. Some adults in our sample expressed difficulty in answering certain questions in the CHU9D and EQ-5D instruments, which were interpreted as more 'subjective' and only possible to answer by asking the child/adolescent directly. Examples included subjective states (e.g., emotions and pain). This finding is consistent with guidance from the US Food and Drug Administration, who discourage the use of proxy-reported outcomes, particularly for questions on subjective experience (vs. observable behaviour) [32]. However, it raises wider conceptual questions requiring consideration about the value of subjective experience items to HRQoL measurement, where children cannot self-report.

4.3 Comprehensiveness

This study identified some aspects of health (and wellbeing) that could be considered important to child/adolescent HRQoL but that were not assessed by the CHU9D or EQ-5D instruments. The most common aspects mentioned by participants were the social dimension of HRQoL (i.e., relationships, social connections, communication); positive emotions (i.e., happiness); and eating (i.e., diet, nutrition, and relationship with food). The omission of social content is consistent with a recent study looking at the content validity of the EQ-5D-Y-3L in Chinese children and adolescents [33]. The results are not altogether surprising, as it is unlikely that a generic instrument will include all aspects of HRQoL deemed relevant for every health condition. However, the goal of a generic instrument should be to include items related to the target construct that are important and relevant for most people living with different levels of health. Published evidence has shown that the EQ-5D instruments lack comprehensiveness for adults. Shah et al. [34] noted that, in the UK adult general population, the EQ-5D-5L did not cover several important aspects of health. Similar findings were observed in an online survey of adults recruited from New Zealand [35]. Further, literature exists that demonstrates that, in given (adult) health conditions, the measure is lacking in key areas [36]. When considering the suitability of the EQ-5D-5L and EQ-5D-Y-3L instruments for use in a child/adolescent population, published research examining their content validity is sparse. A recent publication reported the EQ-5D-Y-5L to be insufficiently comprehensive in a particular health condition (children with intellectual disability) [10]. Krig et al. [37] examined the acceptability of the EQ-5D-Y-5L instrument in a survey study of children and adolescents in psychiatric inpatient care, reporting insufficient content relating to psychological or psychosocial aspects.

While recognising that generic measures will never be fully comprehensive, it is important to consider that instrument development is an iterative process. Measuring what matters to individuals is the crux of content validity – and if what matters to people changes over time, then it may be pertinent to identify a current conceptual framework of HRQoL for children/adolescents. In doing so, it would be possible to identify whether the CHU9D and EQ-5D instruments are sufficiently comprehensive by mapping item

content to the conceptual framework. Further research is recommended in this area. In addition, it should be acknowledged that the EQ-5D-5L, EQ-5D-Y-3L, and CHU9D instruments were developed at a time when evidencing content validity for the target population was not a requirement. The results of this study may not be surprising given the application of assessment of modern-day international standards for instrument development to legacy measures. Notwithstanding, the findings presented here represent areas for potential improvement of the instruments. Any modifications need to be carefully balanced against the need to ensure comparability and standardisation in data collected using these established instruments.

4.4 Strengths and limitations

This study used existing best practice guidance to explore the content validity of three generic PWMs used in children/adolescents. The large study sample covered a range of child/adolescent ages and included those with and without physical and/or mental health conditions, illnesses, or disabilities. The online data collection method allowed for recruitment of participants across the UK. Interviews were conducted by two experienced qualitative researchers, who discussed preliminary findings regularly to allow for consideration of additional probing questions to be introduced to subsequent interviews.

The study is not without limitations. Recruitment methods not linked to healthcare assessments and/or interventions may have resulted in participants with (or parent/guardians of children/adolescents with) relatively mild health conditions. More severe health conditions were unlikely to have been included within the study sample. Further, consideration of the instruments with respect to health condition(s) may not have been reflected upon as fully by participants had they been recruited through a hospital or clinic. In addition, the study was conducted within the UK, and findings may differ in other countries and/or cultures. It is possible that the online method of data collection may have biased the sample towards people who are digitally literate and have access to the internet and/or devices.

The study was not designed to assess the content validity across specific child health conditions, therefore the specifics of any health condition of the child/participant were not collected. Instead, this was determined by self-report responses under broad categories ('physical', 'mental', or 'no' health condition, disease, or disability). The recruitment methodology employed means it was not possible to validate responses against medical records. Finally, although the order of the questionnaires was randomised across participants, the overlap of content between the EQ-5D instrument(s) and the CHU9D meant individuals' perspectives on content may have been influenced by the questionnaire they were first exposed to.

5 Conclusion

This is the first study of its kind to assess the content validity of three commonly used generic PWMs for children in a mixed group of participants, representing the generic target population for the measures. The research provides novel insights to the evidence base to support instrument selection in future research and complements ongoing discussions regarding the appropriateness of existing measures and their refinement.

Key findings from this research suggest that, although the content of the respective instruments is supported, the measures are not without issues. Potential problems were identified in aspects of comprehensibility, relevance, and comprehensiveness of all the PWMs, which present opportunities for future research and refinement to improve the content validity of the measures for use in children/adolescents (and their parent/guardian proxies). Balancing refinements against the need to preserve the integrity and consistency of established legacy measures is important, but outcome measure development is a constructive, iterative, and ongoing process that should aim to adapt to new evidence that emerges over time. The current study provides important new evidence that contributes to that process.

Glossary

HRQoL	Health related quality of life
NICE	National Institute for Health and Care Excellence
PWM	Preference-weighted measure
QALY	Quality adjusted life year

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40271-025-00743-9.

Acknowledgements The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, NICE, the Department of Health and Social Care or its arm's-length bodies, or other UK government departments. Any errors are the responsibility of the authors. The authors thank Koonal Shah and Elizabeth Coates from NICE for their comments and guidance throughout the project. We also acknowledge and thank all the study participants. We also thank Donna Davis for her professional support on the project.

Declarations

Funding This research is funded through the National Institute for Health and Care Research (NIHR) Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions (reference NIHR206111). The views expressed are those of the authors and

not necessarily those of the NIHR or the Department of Health and Social Care. The funders had no role in data collection, interpretation, and reporting.

Conflicts of Interest Jill Carlton, Philip Powell, and Donna Rowen have received funding from the EuroQol Research Foundation. Donna Rowen is a member of the EuroQol Group. The CHU9D instrument is owned and licensed by the University of Sheffield. Andrew Kirkcaldy has no conflicts to declare.

Ethics Approval Ethical approval was obtained from the University of Sheffield (Ref: 045650). All participants gave informed consent before interviews were conducted.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication All participants consented to their anonymised data being used to support publication.

Availability of Data and Materials Data are provided in the manuscript and accompanying supplementary files.

Code availability All data supporting the findings of this study are available within the paper and its Supplementary Information.

CRediT Authorship Contribution Statement JC: study conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, visualisation, writing—original draft, and writing—revisions. PAP: study conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, visualisation, writing—original draft, and writing—revisions. AK: formal analysis and writing—revisions. DR: study conceptualisation, funding acquisition, methodology, and writing—revisions.

References

1. UK National Institute for Health and Care Excellence. NICE Health Technology Evaluations: the manual. NICE process and methods [PMG36]. 31 January 2022. https://www.nice.org.uk/process/pmg36
2. Hill H, Rowen D, Pennington B, Wong R, Wailoo A. A review of the methods used to generate utility values in NICE technology assessments for children and adolescents. Value Health. 2020;23(7):907–17. https://doi.org/10.1016/j.jval.2020.02.011.
3. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res. 2010;19(6):875–86. https://doi.org/10.1007/s11136-010-9648-y.
4. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res. 2009;18(8):1105–13. https://doi.org/10.1007/s11136-009-9524-9.
5. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49. https://doi.org/10.1007/s11136-010-9606-8.
6. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45. https://doi.org/10.1016/j.jclinepi.2010.02.006.
7. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A review of the psychometric performance of selected child and adolescent preference-based measures used to produce utilities for child and adolescent health. Value Health. 2021;24(3):443–60. https://doi.org/10.1016/j.jval.2020.09.012.
8. O'Loughlin R, Jones R, Chen G, et al. Quality Of Life in Kids: Key evidence to strengthen decisions in Australia (QUOKKA) project team. Comparing the psychometric performance of generic paediatric health-related quality of life instruments in children and adolescents with ADHD, anxiety and/or depression. [Published erratum appears in: Pharmacoeconomics. 2024;42(Suppl 1):79. https://doi.org/10.1007/s40273-024-01373-z.] Pharmacoeconomics. 2024;42(Suppl 1):57-77. https://doi.org/10.1007/s40273-024-01354-2.
9. Jones R, O'Loughlin R, Xiong X, et al. Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA) Project Team. Comparative psychometric performance of common generic paediatric health-related quality of life instrument descriptive systems: results from the Australian Paediatric Multi-Instrument Comparison Study. Pharmacoeconomics. 2024;42(1):39–55. https://doi.org/10.1007/s40273-023-01330-2.
10. Blackmore AM, Mulhern B, Norman R, et al. How well does the EQ-5D-Y-5L describe children with intellectual disability? There's a lot more to my child than that she can't wash or dress herself. Value Health. 2024;27(2):190–8. https://doi.org/10.1016/j.jval.2023.11.005.
11. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159–70. https://doi.org/10.1007/s11136-018-1829-0.
12. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1–eliciting concepts for a new PRO instrument. Value Health. 2011;14(8):967–77. https://doi.org/10.1016/j.jval.2011.06.014.
13. Terwee CB PC, Chiarotto A, de Vet HCW, et al. COSMIN methodology for assessing the content validity of PROMs. User manual. Version 1. 2018. https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf
14. Tan RL, Soh SZY, Chen LA, Herdman M, Luo N. Psychometric properties of generic preference-weighted measures for children and adolescents: a systematic review. Pharmacoeconomics. 2023;41(2):155–74. https://doi.org/10.1007/s40273-022-01205-y.
15. Stevens K. Working with children to develop dimensions for a preference based generic paediatric health related quality of life measure. Sheffield: University of Sheffield; 2010. https://eprints.whiterose.ac.uk/42973/

16. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. Qual Health Res. 2010;20(3):340–51. https://doi.org/10.1177/1049732309358328.
17. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. Appl Health Econ Health Policy. 2011;9(3):157–69. https://doi.org/10.2165/11587350-000000000-00000.
18. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. Value Health. 2012;15(8):1092–9. https://doi.org/10.1016/j.jval.2012.07.011.
19. Canaway AG, Frew EJ. Measuring preference-based quality of life in children aged 6–7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y–the WAVES pilot study. Qual Life Res. 2013;22(1):173–83. https://doi.org/10.1007/s11136-012-0119-5.
20. Stevens K. Valuation of the child health utility 9D index. Pharmacoeconomics. 2012;30(8):729–47. https://doi.org/10.2165/11599120-000000000-00000.
21. Rowen D, Mulhern B, Stevens K, Vermaire JH. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. Value Health. 2018;21(10):1234–42. https://doi.org/10.1016/j.jval.2018.03.016.
22. Kreimeier S, Astrom M, Burstrom K, 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.
23. Prevolnik Rupel V, Ogorevc M, IMPACT HTA HRQoL Group. EQ-5D-Y value set for Slovenia. Pharmacoeconomics. 2021;39(4):463–71. https://doi.org/10.1007/s40273-020-00994-4.
24. Rencz F, Ruzsa G, Bato A, Yang Z, Finch AP, Brodszky V. Value set for the EQ-5D-Y-3L in Hungary. Pharmacoeconomics. 2022;40(Suppl 2):205–15. https://doi.org/10.1007/s40273-022-01190-2.
25. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman B, Burgess R, editors. Analyzing qualitative data. London: Routledge; 1994.
26. Kiernan MD, Hill M. Framework analysis: a whole paradigm approach. Qual Res J. 2018;18(3):248–61. https://doi.org/10.1108/QRJ-D-17-00008.
27. Peasgood T, Mukuria C, Carlton J, Connell J, Brazier J. Criteria for item selection for a preference-based measure for use in economic evaluation. Qual Life Res. 2021;30(5):1425–32. https://doi.org/10.1007/s11136-020-02718-9.
28. Khanna D, Lay K, Khadka J, Mpundu-Kaambwa C, Ratcliffe J. Quality of Life in Kids: key evidence to strengthen decisions in Australia (QUOKKA) project team. How do children understand and respond to the EQ-5D-Y-3L? A mixed methods study in a community-based sample of 6–12-year-olds. Health Qual Life Outcomes. 2024;22(1):105. https://doi.org/10.1186/s12955-024-02320-4.
29. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2–assessing respondent understanding. Value Health. 2011;14(8):978–88. https://doi.org/10.1016/j.jval.2011.06.013.
30. Mukuria C, Rowen D, Mulhern B, et al. The short form 6 dimensions (SF-6D): development and evolution. Appl Health Econ Health Policy. 2025;23(1):19–33. https://doi.org/10.1007/s40258-024-00919-8.
31. Peasgood T, Caruana JM, Mukuria C. Systematic review of the effect of a one-day versus seven-day recall duration on patient reported outcome measures (PROMs). Patient. 2023;16(3):201–21. https://doi.org/10.1007/s40271-022-00611-w.
32. U.S. Department of Health and Human Services Food and Drug Administration. Patient-focused drug development: selecting, developing, or modifying fit for purpose clinical outcome assessments. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. https://www.fda.gov/media/159500/download. Accessed 12 Feb 2025.
33. Ding Y, Zhang G, Yang Z, et al. Evaluating the content validity of the EQ-5D-Y for Chinese children and adolescents. Qual Life Res. 2025. https://doi.org/10.1007/s11136-025-03917-y. (**Epub ahead of print 2025**).
34. Shah KK, Mulhern B, Longworth L, Janssen MF. Views of the UK general public on important aspects of health not captured by EQ-5D. Patient. 2017;10(6):701–9. https://doi.org/10.1007/s40271-017-0240-1.
35. Devlin NJ, Hansen P, Selai C. Understanding health state valuations: a qualitative analysis of respondents' comments. Qual Life Res. 2004;13(7):1265–77. https://doi.org/10.1023/B:QURE.0000037495.00959.9b.
36. Powell PA, Carlton J. A comprehensive qualitative framework for health-related quality of life in Duchenne muscular dystrophy. Qual Life Res. 2023;32(1):225–36. https://doi.org/10.1007/s11136-022-03240-w.
37. Krig S, Astrom M, Kulane A, Burstrom K. Acceptability of the health-related quality of life instrument EQ-5D-Y-5L among patients in child and adolescent psychiatric inpatient care. Acta Paediatr. 2021;110(3):899–906. https://doi.org/10.1111/apa.1554.