Exploring the potential for item banking in assessing quality of life for evaluating adolescent health interventions

Jyoti Khadka (✉ jyoti.khadka@flinders.edu.au)

Flinders University <https://orcid.org/0000-0003-1012-2119>

Gang Chen

Monash University

Julie Ratcliffe

Flinders University

Research

Keywords: preference-based instrument, quality of life, item banking, Rasch analysis

Posted Date: August 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-50156/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

To develop and validate an item bank using existing items from both preference and non-preference-based health-related quality of life (HRQoL) instruments targeted for application with adolescent populations.

Methods

Australian adolescents completed the Child Health Utility 9D (CHU9D), demographic details, self-rated health status, disability and one other instrument (either preference-based: Health Utilities Index (HUI) or EuroQoL-5 Dimension-Youth (EQ-5D-Y) or Assessment of Quality of Life (AQoL)-6D Adolescent or non-preference-based: Pediatric Quality of Life Inventory (PedsQL) Short Form 15 or KIDSCREEN-10) using online survey portals. Data from all the instruments were pooled together to create an item bank, which was then subjected to Rasch analysis to tests its psychometric properties. Validity of the item bank was assessed by testing its ability to discriminate varying levels of self-reported health, disability and socio-economic status (as approximated by the Family Affluence Scale).

Results

A total of 4,352 Australian adolescents aged 11–17 years from the general population participated. The CHU9D (common instrument) items were used as anchor items to create the item bank. Rasch analysis demonstrated that 2 of the 75 items (2.7%) misfit and were removed from the item bank. The 73-item item bank had adequate precision, no misfitting items, no item bias and demonstrated sufficient unidimensionality. In general boys exhibited higher HRQoL scores than girls ($p = 0.001$). More severe health status/disability and lower socio-economic status being associated with lower HRQoL ($p < 0.0001$).

Conclusions

The item bank demonstrated adequate Rasch-based psychometric properties and validity demonstrating the feasibility of the construction of an item bank in this context. The addition of more targeted items for adolescents in the general population and the addition of participants with more health impairments may improve the general applicability of the item bank for both general population and patient cohorts.

Background

The use of patient-reported outcome measures (PROMs) as central components of quality assessment and as endpoints to the evaluation of health-specific interventions and products has been endorsed by

regulatory, pricing and reimbursement authorities in several countries [1–3]. As a result, a plethora of PROMs has been developed and utilised to assess health-related quality of life (HRQoL) across health, social care and public health sectors [4–6]. In the context of evaluation of adolescent health programmes, several PROMs have gained prominence in recent years [4, 7]. These PROMs can be broadly divided into preference and non-preference-based on whether they produce (1) a single preference-based health state utility index or (2) a simple summative (unweighted) score and/or a profile of individual dimension scores respectively [8]. Preference-based PROMs can be utilised to generate quality-adjusted life year (QALY) estimates that are useful for assessing clinical effectiveness and form an integral component of cost-utility analysis. The relative preferences/weights for health states defined by the respective PROMs descriptive system are typically obtained from large samples of the general public [9].

Generally, there is a lack of consensus regarding the selection of appropriate PROMs in specific contexts because a PROM developed for a specific population may not be directly applicable to other populations [10–12]. A relatively new approach called item banking has emerged to address many shortcomings of the existing PROMs [13]. An item bank is a large collection of calibrated items to measure different HRQoL dimensions. Whilst the application of item banks to health economics and preference-based HRQoL instruments in particular is new [14, 15], item banks have previously been created and are gaining prominence in health services research through the adoption of modern psychometric methods such as Rasch analysis and computer-based adaptive testings (CAT) [12, 14, 16, 17]. Studies have shown that item banking may significantly improve accuracy in a PROM measurement, thereby reducing the sample sizes required to achieve meaningful differences in HRQoL between groups in clinical trials and subsequently reduces the cost of large scale clinical studies [18, 19]. Such item banking has been developed in a series of health fields, for example, PROMIS item banks, the Eye-tem Bank [13, 20, 21]. However, in health economics which heavily relies on PROMs data for economic evaluations the application of item banking is relatively novel. The approach has been applied recently in adult populations by the PROMIS group in the US [15, 22].

Our group carried out a series of mapping studies for predicting preference-based CHU9D estimates from preference-based (Assessment of Quality of Life (AQoL)-6D Adolescent version, AQoL-6D, Health Utilities Index (HUI) Mark 2 and HUI Mark 3; EQ-5D-Youth, EQ-5D-Y) [23, 24] and non-preference-based instruments (KIDSCREEN-10 and Pediatric Quality of Life Inventory-Short Form, PedsQoL-SF15) [25]. For the mapping project, data was collected from over four thousand healthy adolescents from the community in Australia. Building on this database, the aim was to investigate the feasibility of developing an adolescent specific item bank by pooling data from the existing PROMs for subsequent application in health economic evaluations. The study described in this manuscript is the first phase of a multi-phased programme of research which has an overarching aim to validate the HRQoL item bank in adolescents living in the community with and without health condition/s and to develop preference-based utility scores for the relevant item bank dimensions.

Methods

A series of web-based surveys were developed for administration to Australian adolescents (age 11–17 years) living in the community through Pureprofile (an online survey panel company) following informed dyad (parent and adolescent) consent to participate. All consenting adolescents were asked to complete the CHU9D and one of the other five PROMs (AQoL-6D, HUI 2&3, EQ-5D-Y, KIDSCREEN-10 and PedsQL SF15). In addition, each respondents also completed self-reported questions on general health, disability status and socio-demographic characteristics including the Family Affluence Scale (FAS, a 4-item adolescent specific measure of family socio-economic status; the total score ranges from 0 to 7 and were categorised into three groups as 0–3 low, 4–5 medium and 6–7 high) [26]. More details on the web-based surveys have been published elsewhere [23, 24].

Generic Preference-based Instruments

The CHU9D is a generic preference-based measure of HRQoL specifically developed with and validated for use in paediatric populations aged 7–17 years [27, 28]. The CHU9D has nine dimensions (worried, sad, pain, tired, annoyed, schoolwork/homework, sleep, daily routine, and activities) each represented by a single item and has five levels of severity response categories.

The AQoL-6D Adolescent version is a generic preference-based utility instrument originally develop for application in adults but later adapted for use in adolescents [29]. The instrument has 20 items distributed across six dimensions (independent, mental health, coping, relationships, pain, and senses) with four to six levels of severity response categories.

The HUI 2/3 are generic preference-based utility instruments to measure health status and HRQoL [30]. The HUI2 was initially developed in the early 1990s for measuring and valuing long-term outcomes in patients with childhood cancer. The HUI3 represents a further development for use in both clinical and general populations. These two instruments are independent but complementary systems. The HUI2 has seven dimensions of HRQoL (sensation, mobility, emotion, cognition, self-care, pain and fertility) assessed on 4 or 5 levels and the HUI3 has eight dimensions (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain) on 6 distinct levels of response categories. Together, the 15-item HUI 2/3 instruments were included excluding the item related to fertility (not relevant to our study population).

The EQ-5D-Y is a pediatric version adaptation of the widely applied generic instrument (EQ-5D-3L) originally developed for adults [31]. The EQ-5D-Y was developed by revising the content and wording of the adult version to ensure their relevance and clarity for children and adolescents. The EQ-5D-Y contains five dimensions (mobility, looking after myself, usual activities, pain or discomfort, and feeling worried, sad or unhappy) assessed on 3 levels of severity response categories.

Generic Non-preference-based Instruments

The KIDSCREEN-10 is developed to assess pediatric health and well-being specifically relevant for aged 8 to 18 years [32]. It was simultaneously developed across 13 European countries. The original instrument was a 52-item covering 10 dimensions of HRQoL [33]. A 27-item shorter version was derived from the 52-item instrument and subsequently a 10-item instrument was derived, which was used in the study.

The PedsQL SF15 is developed to be used in in paediatric health research [34]. It has four dimensions and consists of 15 items which measure physical (5 items), emotional (4 items), social (3 items) and school functioning (3 items) rated on a 5-point Likert scale. In this study, the adolescent self-report version was used.

Item Banking

A total of 77 items from the 6 instruments were available and initially considered for creating an item bank. Two items on global health (KIDSCREEN-10 item-11, “In general, how would you say your health is?” and EQ-VAS, “How good is your health today?”) were not considered for further analysis as they did not measure a specific HRQoL domain leaving 75 items. After pooling raw response data from the 6 instruments, two approaches were taken:

1. Overall item bank: All 75 items were pooled together to assess whether these items together could be calibrated on a single continuum scale to form a valid item bank. As all the respondents completed the CHU9D, it was used as the anchoring instrument to link data from the other 5 instruments. For this, a separate Rasch analysis (describe below in detail) was conducted on the CHU9D data only and item parameter values were obtained for its 9 items, which were anchored to build a 75-item item bank.
2. Dimension-specific item banks: 75 items were classified and binned across 9 dimensions of HRQoL by the authors separately. The separate classifications were reviewed by the authors together and reconciled into one, with any discrepancies resolved through group discussion. The final dimensions were physical activities (13 items), mobility (6 items), mental health (17 items), social relationship (8 items), pain (7 items), fatigue & memory (5 items), coping (5 items), sensory (9 items) and school activities (5 items), See Supplementary Table 1.

Psychometric Properties Assessment Using Rasch Analysis

Rasch analysis is a probabilistic mathematical model which assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item’s location (i.e. item parameter) and the respondent’s health status [35]. Assessing these probabilities across the items, Rasch analysis estimates the health status of individual respondents (person parameters) and the values of the health states represented by items (weighted item parameters) on the same latent scale expressed in log of the odds units (or Logits, a measure in interval-level scale). Rasch analysis is a widely accepted methodology to develop and validate PROMs including item banks [13, 36]. The advantage of Rasch

analysis is that it also provides surgical insights into whether an instrument could form a valid scale based on its assessment on a series of psychometric properties [37]. The following Rasch-based psychometric properties were assessed for the item bank [20, 38].

Measurement precision

It is the ability of an instrument to discriminate between respondents with differing levels of the underlying construct. Precision is indicated by person separation index (PSI) and person reliability (PR) coefficients, where values of ≥ 2.0 and ≥ 0.8 , respectively, are considered minimally acceptable levels where the instrument could differentiate people into three strata of the underlying construct. PR is also equivalent to Cronbach's Alpha (a traditional test of reliability), therefore a $PR \geq 0.8$ indicates acceptable internal consistency of the instrument [39].

Targeting

Ideally, there should be a good spread of items across the full range of respondents' scores. When respondents have higher (ceiling) or lower (floor) construct than most of the items in the instrument, the range of construct coverage by the instrument may not be adequate. This leads to poor targeting. Targeting is estimated by determining the difference between the item mean (defined as 0 logits by default) and the mean of respondents' measure; a difference of < 1.0 logits is desired.

Unidimensionality

Item fit statistics and principal component analysis (PCA) of residuals were used to assess whether the item bank attained the requirements of unidimensionality.

Item fit statistics

Chi-square fit statistics (mean square, MnSq) were used to assess how well the data fit the Rasch model. Misfitting items may indicate that they are measuring a construct different than other items in the instrument, indicating multi-dimensionality. There are two item fit statistics, infit and outfit. Infit is more sensitive to the pattern of responses to items closely targeted to the respondents' ability whereas outfit statistics is more influenced by outliers (respondents with very high or low construct). A fit statistics value between 0.50 and 1.50 is considered acceptable and is still conducive for productive measurement [40].

Principal Component Analysis (PCA) of residuals

A PCA of residuals was conducted to assess for patterns in the data that did not accord with the Rasch requirements, suggesting that groups of items may be forming a secondary dimension. An instrument is considered unidimensional if the raw variance explained for the first factor (i.e. primary dimension) is expected to be $\geq 50\%$, and the unexplained variance by the first contrast (i.e. first component in the correlation matrix of the residuals) is < 2 eigenvalues [38].

Differential Item Functioning (DIF)

DIF determines whether item bias exists for sample subgroups (e.g. age group, gender, disability). A DIF contrast of < 1.0 logits and corresponding p-value of < 0.05 is acceptable [20]. DIF was assessed for age group (11–14 yrs Vs 15–17 yrs), gender (boys vs girls) and disability (yes vs no).

Validity And Reliability Assessments Of The Hrql Item Bank

Known group validity

Known group validity (the extent to which the item bank could discriminate between groups known to be different) was assessed by demonstrating that respondents with different ratings of self-reported health, disability and affluence levels (measured by FAS) have significantly different HRQoL scores.

Construct validity

Item separation index (ISD) and item reliability (IR) coefficient were used to verify validity of the item hierarchy in the item bank. An ISD value of > 3.00 and $IR > 0.9$ imply that the sample is large enough to establish reproducible item calibration hierarchy. These statistics inform construct validity (the extent to which an instrument measure what it purports to be measuring) of the item bank [41].

Statistical analysis

Rasch analysis was performed on the 75-item pool (with anchored CHU9D items) and each of 9 dimensions separately using Winsteps ver 4.4.3 software (Chicago, USA) using Andrich rating scale model per question format (i.e. common item stem and response categories) [42].

Firstly, the 75 items in the pool were classified into 38 groups based on whether they shared common rating scale (i.e. the same preceding statement and categories). Secondly, the response polarity was reversed for all 10 KIDSCREEN items to make them consistent with other items, such that a higher response scores meant worse HRQoL. Finally, the pool data for 75-items was subjected to a group Rasch analysis with 1 Andrich Rating scale per question format (i.e. 38 rating scales) [43]. Rasch analysis also assessed the psychometric properties of the item pools without items from the 2 non-preference based instruments (PedsQL and KIDSCREEN).

Descriptive data were analysed using STATA Version 15.1, Stata Corp LLC, Texas, USA (Texas, USA) [44]. To compare median HRQoL item bank scores, Wilcoxon rank-sum was used to compare between two groups (gender, age-group and disability) and Kruskal-Wallis test (self-reported health and affluence levels) was used to test between multiple groups. Dunn's test was carried out following Kruskal-Wallis test for multiple pairwise comparison between the groups [45]. All statistical tests used a level of significance at 2-sided alpha of 0.05.

Results

The study cohort included 4,352 Australian adolescents from the community. After exclusion of 246 (5.7%) respondents who demonstrated very high ceiling effects (i.e. they responded to only the highest category “no problem/ no issue” to all the items), there were 4,106 eligible individuals. The mean age of the respondents was 14.7 ($\pm$ 1.88) years and there were approximately equal males and females. The majority of the respondents self-reported to have no disability (87.6%), had excellent or very good health ratings (72.1%) and were from moderate to high family affluence backgrounds (Table 1). Typical of a general population response, 58 out of 75 items in the item pool demonstrated a ceiling effect (> 15% respondents reported no problems/issue).

Table 1
Demographic characteristics of the study population

Variables	(n = 4,106)
Age	14.7 (1.88)
<i>Mean, SD</i>	
<i>Median, IQR</i>	15 (11–16) years
Range	11–17 years
Age groups, n(%)	
11–15 years	2,412 (58.7)
16–17 years	1,694 (41.3)
Females, n (%)	2,078 (50.6)
Self-reported Health ratings, n(%)	
Excellent	1,159 (28.2)
Very good	1,803 (43.9)
Good	876 (21.3)
Fair	226 (5.5)
Poor	42 (1.02)
Self-reported disability, n(%)	
Yes	516 (12.6)
No	3,590 (87.4)
Family Affluence Scale (FAS), n (%)	
Low affluence	70 (1.7)
Medium affluence	1,454 (35.4)
High affluence	2,582 (62.9)
Instrument	
CHU9D	4,106
KIDSCREEN-10	583
PedsQL SF15	739
AQoL-6D Adolescent	476
EQ-5D-Y	1839

Variables	(n = 4,106)
HUI	468
Note: CHU9D = Child Health Utility 9D; PedsQL SF15 = Pediatric Quality of Life Inventory Short Form 15; AQoL = Assessment of Quality of Life; EQ-5D-Y = EQ-5D-Youth; HUI = Health Utilities Index	

Psychometric properties of the HRQoL Item bank

Rasch analysis revealed that two items (KIDSCREEN-Lonely and KIDSCREEN-Sad) were misfitting items ($MNSQ > 1.5$), hence they were deleted from the item bank. The resulting 73-item item bank demonstrated good psychometric properties including adequate measurement precision ($PSI = 2.11$), acceptable unidimensionality as shown by no misfitting items, and no item bias by population subgroups as shown by no item bias (i.e. no DIF) (Table 2). The PCA of the residuals showed the raw variance explained by the measure was 58%. However, the eigen values for the first contrast was 2.92 and 4 items (PedsQL-run, PedsQL-sports, PedsQL-walk, PedsQL-lift) clustered (loading > 0.4) separate from the main Rasch dimension. Deleting these items from the item bank reduced the measurement precision significantly ($PSI < 2.00$), hence the items were retained in the item bank. The 73-item item bank demonstrated a poor targeting to population (person-item mean difference = 1.87 logit).

Table 2
Rasch based psychometric properties of the childhood quality of life item bank

Rasch metrics (expected values)	Item bank (All 6 instruments)	Final iteration (All 6 instruments)
No of items	75	73#
Sample size	4,106	4,106
Person separating index	2.05	2.11
Measurement precision (person reliability ≥ 0.80)	0.81	0.82
Item separation index	13.40	12.26
Item reliability	0.99	0.99
Item fit statistics (MNSQ < 1.5) Infit	2 (KIDSCREEN_Lonely, KIDSCREEN_Sad)	None
Outfit MNSQ > 1.5	4 (KIDSCREEN_Lonely, KIDSCREEN_Sad, AQoL_Vision, HUI-Vision)	None
Targeting (difference between person & item means < 1.00)	1.86	1.87
PCA analysis Raw variance explained by the measure > 50%	50%	50%
PCA analysis: 1st contrast eigen value (< 3.00)	2.93	2.92
Measurement range (min and maximum)	5.22 (-2.84 to 2.38)	4.08 (-1.70 to 2.38)
Differential item functioning (DIF) by age group, gender and disability (DIF < 1.00)	No DIF	No DIF
#Two items “KIDSCREEN-Lonely” and “KIDSCREEN-Sad” misfit, therefore removed from the item bank.		

Validity Assessment Of The 73-item Hrql Item Bank

The 73-item item bank demonstrated very high ISI (13.40) and IR coefficient (0.99) supporting its construct validity. The median QoL scores for the cohort was – 1.87 (IQR, -1.05 to -2.64) logits. Girls ($z=-5.007$, $p < 0.001$, Fig. 1a), adolescents aged 16–17 years ($Z=-4.74$, $p < 0.001$) and people with self-reported disability ($z = 14.23$, $p < 0.001$, Fig. 1b) reported worse QoL. Self-reported health ratings were associated with poor QoL scores (Chi-squared = 1390.2, $df = 4$, $p < 0.001$, Fig. 1c) with significant

differences within and between all the groups ($P < 0.001$ between all the groups except for between health rating 3 & 5, $p = 0.01$).

Similarly, HRQoL scores were significantly better for high (Chi-squared = 3.05, 2 d.f., $P = 0.001$) and medium socio-economic status (as approximated by FAS) (Chi-squared = 8.58, $P < 0.001$) groups than for those in the low affluence group (Fig. 1d). When compared between groups, there was a significant difference in HRQoL between low and medium (Chi-squared = 8.62, $P < 0.001$), and low and high affluence groups (Chi-squared = 3.07, $P = 0.001$). However, high and medium affluence groups (Chi-squared = 0.72, $P = 0.24$) did not have a significantly different HRQoL scores. These results also demonstrate known-group validity of the 73-item HRQoL item bank.

Rasch Analysis Of The Dimensions

As expected a priori none of the 9 dimensions demonstrated adequate measurement precision (all had a $PSI < 2.00$ or PR coefficient < 0.80) indicating that they lacked enough sensitivity to form a standalone dimension, hence no further analysis was carried out.

Discussion

The main objective of this study was to demonstrate “proof of concept” of item banking in the development of adolescent specific HRQoL instruments for subsequent application in health economic evaluations. This was achieved by pooling data from six PROMs suitable for application in adolescent populations, four (CHU9D, EQ-5D-Y, HUI2/3, AQoL-6D) are preference based and two (PedsQL SF15 and KIDSCREEN-10) are non-preference based but have established mapping algorithms which facilitate their application in economic evaluation [46, 47]. By utilising Rasch analysis and linking pooled PROMs data, an item bank was constructed that contained a large volume of items calibrated on a single continuum scale of HRQoL.

Although the item bank contained items representing different QoL dimensions, the 73-item bank met all the Rasch-based psychometric requirements to qualify as a unidimensional scale. The PCA analysis also demonstrated that four items referring to sports activities clustered together suggesting that these items might form a secondary dimension. However, the removal of the four items reduced the precision of the item bank, which suggests that these items were adding more signal than noise, hence they were retained. Similar approaches were utilised to tackle item clustering and multi-dimensionality while developing item banks in other health fields [20, 48]. It is found that when a range of items representing different HRQoL dimensions pooled together a valid unidimensional latent scale was identified. This unidimensional scale represents a latent concept of HRQoL. This finding may look puzzling at first but it is analogous to a mathematic test which constitutes different components (e.g. word problems, algebra, geometry, calculus etc), which are conceptually different, but they all contribute to the measurement of the overall performance in mathematics of students.

One issue with the item bank was that inadequate coverage towards the higher end of HRQoL. This is likely to be a by-product of the study population itself which was drawn from Australian general population. The majority of the respondents had good health, no disability and were from relatively affluent family backgrounds (Table 1), hence they were expected a priori to have relatively good HRQoL. However, whilst the item bank may not differentiate well between adolescents who have a high HRQoL status, it may match well for those with low to moderate HRQoL. This is likely for this item bank for which the items were driven from generic instruments developed to assess HRQoL impact in populations demonstrating more health-related quality of life impairments e.g. patient samples rather than for healthy populations. Several disease-specific PROMs have demonstrated poor targeting when used on less affected or healthy individuals [49–51]. Similarly, none of the HRQoL dimensions formed a valid scale either due to inadequate number of items and/or content coverage for this largely healthy population. The next phase of this programme of research will test the validity of these dimensions in adolescent patients with more regular and on-going engagement with health services due to the presence of chronic health conditions. Once valid dimensions are identified, we will take a similar methodological approach to that pursued by the PROMIS group to develop the PROMIS-Preference PROPr for application in adults (≥ 18 years), to produce adolescent-specific preference scores (or utility weights) for these HrQoL dimensions [22].

A key advantage of Rasch analysis is that it also assesses whether respondents who have approximately similar HRQoL status across different demographic groups perform in a similar way on each individual item, a test of item invariance. None of the 73 items had significant DIF indicating that the item bank and its items were invariant by key socio-demographic differences. This is a key psychometric property for the item bank which implies that items in the item bank work in similar ways for the different groups to be compared [52]. No DIF also confirms that it is valid to compare QoL scores across different demographic groups. Further, the item bank was able to discriminate between adolescents with different health ratings, disability and affluence levels, demonstrating its validity. That is, adolescents with better self-reported health ratings, no disability and a higher affluence were associated with better QoL scores and vice versa.

A clear advantage of an item bank is that it contains a large collection of calibrated items to provide a comprehensive measure of HRQoL suitable for wide range of people. However, following its initial construction and due to its length, an item bank typically requires a CAT system to administer it. A computer enabled “adaptive test” presents items that are more accurately targeted to the individuals’ status based upon their previous responses. This process provides highly accurate measurement and the test can be continued until a desired level of measurement precision is achieved [49]. By tailoring the test to the individual, the problem of poor targeting which we observed in our item bank can be eliminated. The CAT system not only streamlines the administration of the item banks but also help in expansion of item banks by adding uncalibrated items to the item bank and determining their calibration with Rasch analysis against the calibrated items already in the bank [49]. The CAT also creates an opportunity for electronic implementation via digital portals including smart phones for real time scoring and recording of data. Such systems have been widely developed across different health fields [20, 48, 51, 53]. This study adds to the previous study reported upon by the PROMIS group conducted in adult populations by

providing early evidence that item banking is feasible in the development of HRQoL instruments that may subsequently be applied in health economics context [14, 22].

Leading from this feasibility study, the next steps are to test the validity of the item bank in adolescent patient groups who have regular engagement with health services due to the presence of one or more chronic health condition/s, generate adolescent specific scoring algorithms for utility assessment and test feasibility of a CAT to elicit tailored utility scores in real-time for health economics evaluations. Studies have shown that item banks implemented via CAT need fewer items to obtain superior precision and sensitivity compared to the traditional full-length paper-pencil PROMs, minimising respondent burden [53–55]. This is an important consideration for enhancing participation and completion rates in adolescent populations.

There are some limitations to this study, unlike other item banks that utilised common question format and response categories across all the items to improve measurement accuracy, [11, 56] the items in our item bank retained their original question formats and response scales which might have introduced noise in the measurement. Further the data were obtained from a web-based survey which raises questions around data quality and whether the respondents may or may not provide accurate information. However, appropriate data checks (including data completeness, time taken to complete, identifying respondents who selected perfect responses (flatliners)) were used to deal with this limitation [46].

In conclusion, an item bank has been developed by pooling data from six PROMs. The item bank has demonstrated adequate Rasch-based psychometric properties demonstrating the feasibility of the construction of an item bank in the field of health economics and for the development of instruments suitable for quality of life assessment with adolescents for economic evaluation. The addition of more targeted items for adolescents in the general population or the addition of respondents with more health impairments may improve the applicability and generalisability of the item bank for general population and patient cohorts. Generating adolescent-specific scoring algorithms and utilities and development of a CAT system to administer the item bank are the natural next steps.

Abbreviations

HRQOL	Health-related Quality of Life
CHU-9D	Child Health Utility 9D
HUI	Health Utility Index
EQoL-5D-Y	Euro Quality of Life- 5 Dimension-Youth
EQoL-VAS	Euro Quality of Life- Visual Analogue Scale
AQoL-6D	Assessment of Quality of Life-6 Dimension
PedsQL-SF15	Pediatric Quality of Life Inventory-Short Form
KIDSCREEN-10	KIDSCREEN-10 Index
QALY	Quality-adjusted life year
FAS	Family Affluent Scale
CAT	Computer Adaptive Testing
PROMIS	Patient-reported Outcome Measurement Information System
PSI	Person separation index
PR	Person reliability
PROM	Patient-reported Outcome Measure
ISD	Item separation index
IR	Item reliability
PCA	Principal Component Analysis
DIF	Differential Item Functioning
MnSQ	Mean Square

Declarations

Ethics approval: This study was approved by the Social and Behavioural Research Ethics Committee, Flinders University (project number 4701).

Consent for publication: The manuscript is original and is not under consideration for publication elsewhere. The data, models, or methodology used described in the manuscript are not proprietary. On the behalf of all the co-authors,

Availability of data and material: Data will be made available on request

Conflict of interest/Competing interests: None declared for all the authors

Funding: None

Authors' contribution: JK led data the analysis, interpretation and drafting of the manuscript. All the co-authors have participated in the study design, interpretation of the data and revising the manuscript.

Acknowledgement: Not applicable

References

1. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Version). [Accessed on October 30, 2018]. Available from <https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines>. 4th ed. Canada2017.
2. National Institute of Health and Care Excellence. NICE Guide to the methods of technology appraisal 2013. [Accessed October 30, 2018]. Available from: <http://www.nice.org.uk/article/pmg9/chapter/foreword>. In: NICE, editor. London2013.
3. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0). [Accessed October 30, 2018]. Available from <https://pbac.pbs.gov.au/>. In: Australian Government Department of Health, editor. Australia2016.
4. Chen G, Ratcliffe J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics*. 2015;33(10):1013–28.
5. Cleland J, Hutchinson C, Khadka J, Milte R, Ratcliffe J. A Review of the Development and Application of Generic Preference-Based Instruments with the Older Population. *Appl Health Econ Health Policy*. 2019;17(6):781–801.
6. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90(8):720–44.
7. Khadka J, Kwon J, Petrou S, Lancsar E, Ratcliffe J. Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: A systematic review. *Soc Sci Med*. 2019;240:112543.
8. Mpundu-Kaambwa C, Chen G, Huynh E, Russo R, Ratcliffe J. A review of preference-based measures for the assessment of quality of life in children and adolescents with cerebral palsy. *Qual Life Res*. 2018;27(7):1781–99.
9. Brazier J, Ratcliffe J. Measurement and Valuation of Health for Economic Evaluation. In: Inc E, editor *International Encyclopedia of Public Health* 2nd ed2017. p. 586–93.
10. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–25.
11. Khadka J, McAlinden C, Gothwal VK, Lamoureux EL, Pesudovs K. The importance of rating scale design in the measurement of patient-reported outcomes using questionnaires or item banks. *Invest Ophthalmol Vis Sci*. 2012;53(7):4042–54.

12. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. *Qual Life Res.* 2007;16(Suppl 1):109–19.
13. Khadka J, Fenwick E, Lamoureux E, Pesudovs K. Methods to Develop the Eye-tem Bank to Measure Ophthalmic Quality of Life. *Optom Vis Sci.* 2016;93(12):1485–94.
14. Hanmer J, Feeny D, Fischhoff B, Hays RD, Hess R, Pilkonis PA, et al. The PROMIS of QALYs. *Health Qual Life Outcomes.* 2015;13:122.
15. Hartman JD, Craig BM. Comparing and transforming PROMIS utility values to the EQ-5D. *Qual Life Res.* 2018;27(3):725–33.
16. Askew RL, Cook KF, Keefe FJ, Nowinski CJ, Cella D, Revicki DA, et al. A PROMIS Measure of Neuropathic Pain Quality. *Value Health.* 2016;19(5):623–30.
17. Uy EJB, Xiao LYS, Xin X, Yeo JPT, Pua YH, Lee GL, et al. Developing item banks to measure three important domains of health-related quality of life (HRQOL) in Singapore. *Health Qual Life Outcomes.* 2020;18(1):2.
18. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007;45(5 Suppl 1):3–11.
19. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther.* 2011;13(5):R147.
20. Khadka J, Fenwick EK, Lamoureux EL, Pesudovs K. Item Banking Enables Stand-Alone Measurement of Driving Ability. *Optom Vis Sci.* 2016;93(12):1502–12.
21. Smith AB, Hanbury A, Retzler J. Item banking and computer-adaptive testing in clinical trials: Standing in sight of the PROMISed land. *Contemp Clin Trials Commun.* 2019;13:005–5.
22. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, et al. Estimation of a Preference-Based Summary Score for the Patient-Reported Outcomes Measurement Information System: The PROMIS((R))-Preference (PROPr) Scoring System. *Med Decis Making.* 2018;38(6):683–98.
23. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer MG. Whose values in health? An empirical comparison of the application of adolescent and adult values for the CHU-9D and AQOL-6D in the Australian adolescent general population. *Value Health.* 2012;15(5):730–6.
24. Chen G, Flynn T, Stevens K, Brazier J, Huynh E, Sawyer M, et al. Assessing the Health-Related Quality of Life of Australian Adolescents: An Empirical Comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. *Value Health.* 2015;18(4):432–8.
25. Mpundu-Kaambwa C, Chen G, Russo R, Stevens K, Petersen KD, Ratcliffe J. Mapping CHU9D Utility Scores from the PedsQL(TM) 4.0 SF-15. *Pharmacoeconomics.* 2017;35(4):453–67.
26. Boyce W, Torsheim T, Currie C, Zambon A. The family affluence scale as a measure of national wealth: Validation of an adolescent self-report measure. *Soc Indic Res.* 2006;78(3):473–87.

27. 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.
28. 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.
29. Richardson JR, Peacock SJ, Hawthorne G, Iezzi A, Elsworth G, Day NA. Construction of the descriptive system for the Assessment of Quality of Life AQoL-6D utility instrument. *Health Qual Life Outcomes*. 2012;10:38.
30. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1:54.
31. Wille N, Badia X, Bonsel G, Burstrom 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.
32. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Qual Life Res*. 2010;19(10):1487–500.
33. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Value Health*. 2008;11(4):645–58.
34. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care*. 2001;39(8):800–12.
35. Boone WJ, Staver JR, Mellisa SY. *Rasch Analysis in the Human Sciences*. New York: Springer; 2014.
36. Ho AK, Horton MC, Landwehrmeyer GB, Burgunder JM, Tennant A. European Huntington's Disease N. Meaningful and Measurable Health Domains in Huntington's Disease: Large-Scale Validation of the Huntington's Disease Health-Related Quality of Life Questionnaire Across Severity Stages. *Value Health*. 2019;22(6):712–20.
37. Khadka J, Schoneveld PG, Pesudovs K. Development of a Keratoconus-Specific Questionnaire Using Rasch Analysis. *Optom Vis Sci*. 2017;94(3):395–403.
38. Khadka J, Pesudovs K, McAlinden C, Vogel M, Kernt M, Hirneiss C. Reengineering the glaucoma quality of life-15 questionnaire with rasch analysis. *Invest Ophthalmol Vis Sci*. 2011;52(9):6971–7.
39. Boone WJ, Noltemeyer A. Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*. 2017;4(1):1416898.
40. Khadka J, Ryan B, Margrain TH, Court H, Woodhouse JM. Development of the 25-item Cardiff Visual Ability Questionnaire for Children (CVAQC). *Br J Ophthalmol*. 2010;94(6):730–5.
41. Linacre M. Reliability and separation of measures 2002 [Available from: <http://www.winsteps.com/winman/reliability.htm>.

42. Lincare JM. Winsteps® Rasch measurement computer program. Beaverton: Winsteps.com; 2020.
43. Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*. 2010;36(5):718–32.
44. StataCorp. Stata Statistical Software: Release 15. College Station. TX: StataCorp LLC; 2017.
45. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's tes. *Stata J*. 2015;15(1):292–300.
46. Mpundu-Kaambwa C, Chen G, Huynh E, Russo R, Ratcliffe J. Mapping the PedsQL onto the CHU9D: An Assessment of External Validity in a Large Community-Based Sample. *Pharmacoeconomics*. 2019;37(9):1139–53.
47. Chen G, Stevens K, Rowen D, Ratcliffe J. From KIDSCREEN-10 to CHU9D: creating a unique mapping algorithm for application in economic evaluation. *Health Qual Life Outcomes*. 2014;12:134.
48. Fenwick EK, Khadka J, Pesudovs K, Rees G, Wong TY, Lamoureux EL. Diabetic Retinopathy and Macular Edema Quality-of-Life Item Banks: Development and Initial Evaluation Using Computerized Adaptive Testing. *Invest Ophthalmol Vis Sci*. 2017;58(14):6379–87.
49. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci*. 2010;87(4):285–93.
50. Omara M, Stamm T, Boecker M, Ritschl V, Mosor E, Salzberger T, et al. Rasch model of the Child Perceptions Questionnaire for oral health-related quality of life A step forward toward accurate outcome measures. *J Am Dent Assoc*. 2019;150(5):352–+.
51. Cleanthous S, Barbic SP, Smith S, Regnault A. Psychometric performance of the PROMIS(R) depression item bank: a comparison of the 28- and 51-item versions using Rasch measurement theory. *J Patient Rep Outcomes*. 2019;3(1):47.
52. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health Qual Life Out*. 2017;15.
53. Fries J, Rose M, Krishnan E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *J Rheumatol*. 2011;38(8):1759–64.
54. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol*. 2009;36(9):2061–6.
55. Fries JF, Krishnan E, Bruce B. Items. Instruments, Crosswalks, and PROMIS. *J Rheumatol*. 2009;36(6):1093–5.
56. Khadka J, Gothwal VK, McAlinden C, Lamoureux EL, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.

Figures

Figure 1

a: Box and whisker plot showing median and interquartile range for health-related quality of life scores between boys and girls. b: Box and whisker plot showing median and interquartile range for health-related quality of life scores between adolescents who self-reported having disability (Yes) and no disability (No). c: Box and whisker plot showing median and interquartile range for health-related quality of life scores among adolescents grouped into 5 levels of self-rated health status. d: Box and whisker plot showing median and interquartile range for health-related quality of life scores among adolescents grouped into 3 groups of affluence (Low, Medium and High) as approximated by the Family Affluence Scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)