The Development of a Preference-Based Health-Related Quality of Life Measure for Children Based on the PedsQL: the PedsUtil Scoring System

by

Ellen Kim DeLuca

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Services Organization and Policy)
in the University of Michigan
2023

Doctoral Committee:

Professor Lisa A. Prosser, Chair
Professor Kim Dalziel, University of Melbourne
Professor David W. Hutton
Professor Scott Roberts

Ellen Kim

ekdeluca@umich.edu

ORCID iD: 0000-0002-0894-1118

Dedication

To my son Ellis, my husband Kevin,

and my parents Hyunjoo and Seong-Min

Acknowledgements

I would like to take this opportunity to express my deepest gratitude and appreciation to the amazing people who have supported me throughout my academic career and have guided me to the completion of this dissertation.

First and foremost, I would like to thank to Dr. Lisa A. Prosser, my dissertation chair, mentor, and collaborator. Dr. Prosser is an extraordinary leader who is dedicated to pushing the boundaries of knowledge. She is a constant source of inspiration for me. I have learned more than I ever thought was possible under her guidance and I am so proud to join the field of decision sciences as one of her trainees. Moreover, she was one of my biggest supporters and advocates when I became pregnant during my PhD studies. Her unwavering support provided me with the necessary strength and motivation to overcome the many challenges during this difficult time, and I am forever indebted to her for the invaluable role she has played in my academic and personal journey. I look forward to her mentorship for many years to come.

Next, I would like to thank Dr. Kim Dalziel for her commitment to my growth as an independent researcher. Even with the thousands of miles of distance (approximately 8,870 miles) and 15-hour time difference, she has always made time for me. She has provided me with rigorous training in health economics and has challenged me to push beyond my limits. Her profound expertise, constructive feedback, and genuine belief in my abilities have been instrumental in shaping the researcher I am today. Dr. David Hutton is the best teacher and most wonderful advisor. He has always addressed my countless questions with remarkable patience and has consistently provided me with thoughtful and insightful answers. Thank you for entertaining all of my (not so) great ideas. Your invaluable input and suggestions have greatly enhanced the quality and rigor of my work. Dr. Scott Roberts is the most kind and generous mentor. He has played a fundamental role in shaping my research interests in the world of ELSI. It has been such a privilege to be taught by the "legend of ELSI." I am so excited to be able to continue my ELSI training in my postdoctoral fellowship, and I look forward to more of our adventures to breweries at ELSI conferences!

I am also grateful to the many other faculty and staff at the University of Michigan and beyond for their support throughout this long journey. I especially want to thank the collaborators of my dissertation, including Dr. Eve Wittenberg at Harvard University, Dr. Abram Wagner, Dr. Nicholas Henderson, and Ms. Angela Rose. I also had the great pleasure of working with the best team at the Susan B. Meister Child Health Evaluation and Research (CHEAR) Center, including Acham Gebremariam and Kerra Mercon. Thank you all for sharing your wisdom, which has contributed immensely to my personal and intellectual growth.

I extend my greatest appreciation to the many scholars who have generously shared their time, expertise, and feedback on this dissertation. Thank you to Dr. Janel Hanmer, Ms. Peiwen Jiang, Dr. Brendan Mulhern, and Dr. Deborah Street for their valuable comments and guidance on study design and methods development. I would also like to thank the expert panelists who participated in my dissertation research: Dr. Jeremy Adler, Dr. Melissa Cousino, Dr. Ellen Lipstein, Ms. Allison Martin Nguyen, Dr. Ann Wu, and Dr. Esther Yoon. I am also deeply appreciative to all the study participants whose invaluable contributions made this research possible.

I would like to acknowledge the financial support that has made my PhD studies possible, including the St. David's Neal Kocurek Scholarship, Agency for Healthcare Research and Quality National Research Services Award T32 Predoctoral Traineeship, NIH/NHGRI ELSI Research Training Program T32 Predoctoral Fellowship, Rackham Predoctoral Fellowship, and the PhRMA Foundation Value Assessment Initiative Research Award.

My heartfelt thanks go out to the incredible friends I have made at Michigan whose steadfast support and encouragement have been indispensable in helping me persevere through the most challenging of times. I particularly want to thank my HSOP cohort, including Anton Avanceña, Maria Carabello, Jason Gibbons, Brad Iott, Kasia Klasa, Emily Lawton, and Amanda Mauri, for being the best study buddies and dancing partners. I would also like to thank the ORDS students, including Chris Cadham, Huey-Fen Chen, Grace Chung, Tran Doan, Janamarie Perroud, and David Suh, for being the best teachers and outstanding colleagues. I have learned so much from all of you and I am so lucky to have such a remarkable group of intellectuals by my side.

And most importantly, I would not be where I am today without the constant love and support from my lifelong friends and family. I particularly want to thank Katie and David for being just the most wonderful people. You have graciously endured all of my complaints about PhD life and have always kept me grounded. True friends brave the artic freeze to come visit a lonely graduate student! I also want to thank Jimbo and Clara for being the best cheerleaders (and for being the most fabulous aunt and uncle to Ellis). Words cannot express how fortunate I am to have both of you in my life. I also want to send my appreciation to my exceptional in-laws who always go above and beyond to make me feel cherished. To my mother-in-law Patricia who has watched me grow up since I was 15 years old, thank you for your unconditional love and support all these years. Your constant reminders of how proud you are of me mean the world to me. To my brother Howard, you may be annoying but I do not know who I would be without you. To my parents Hyunjoo and Seong-Min, there are not enough pages in this dissertation to properly express how grateful I am for your love, sacrifice, and unwavering faith in me.

And to my son Ellis and my husband Kevin – you are the reason I have been able to make it this far. You are what motivates me to be the best version of myself every day. Thank you for being my anchor and for holding me up throughout this entire journey. Your endless love and understanding have provided me with the courage and confidence to pursue my dreams wholeheartedly. I love you both so much and I cannot wait to continue to create the most precious memories together as a family.

Table of Contents

List of Tables

List of Figures

Abstract

Economic evaluations play a vital role in guiding resource allocation decisions and setting priorities in healthcare. However, a major limitation is the paucity of child-specific preference-based measures of health-related quality of life (HRQoL) that is suitable for use in economic evaluations. No valid and reliable preference-based HRQoL measure currently exists for use across multiple pediatric age groups.

The Pediatric Quality of Life Inventory (PedsQL) is a HRQoL instrument validated for children 2-18 years. The PedsQL, however, is not a preference-based measure, thus cannot be directly used in economic evaluations. Instead, a preference-based scoring system can be developed to estimate health utilities from responses to the PedsQL. The purpose of this dissertation was to construct a novel preference-based scoring system for the PedsQL that can be applied across multiple pediatric age groups: the PedsUtil scoring system.

The first step of constructing the PedsUtil scoring system involved selecting a subset of dimensions and items from the PedsQL to form the PedsUtil health state classification system. The first study (Chapter 2) identified the core dimension structure of the health state classification system. Using data from the Longitudinal Study of Australian Children (LSAC) (*n* = 45,207), 4 competing dimension structures were evaluated using

confirmatory factor analysis. The findings supported a 7-dimension structure (i.e., Physical Functioning, Pain, Fatigue, Emotional Functioning, Social Functioning, School Functioning, and School Absence) of the PedsUtil health state classification system.

The second study (Chapter 3) used Rasch (i.e., item level ordering, differential item functioning, goodness-of-fit, item spread) and psychometric analyses (i.e., internal consistency, floor and ceiling effects), alongside expert panels and key informant interviews, to select the most representative item within each dimension of the PedsUtil health state classification system. All secondary analyses used data from the LSAC. After considering all decision criteria, the full set of PedsQL items was reduced to a core set of 7 items to create the PedsUtil health state classification system. The PedsUtil health state classification was constructed to be suitable for children 2 years and older, including both children with special healthcare needs and typically functioning children.

The third study (Chapter 4) developed and applied a valuation protocol for the preliminary construction of the PedsUtil scoring system. Pilot data on public preferences for child health states defined by the PedsUtil health state classification system were collected from US adults using a discrete choice experiment (DCE) (n = 213) and a time trade-off (TTO) survey (n = 81). DCE pilot data were analyzed using a conditional logistic regression model. The DCE latent coefficients were mapped onto the mean observed TTO health utility values using ordinary least squares regression to anchor the latent coefficients onto the 0-1 QALY scale. Monotonic models were obtained and the relative importance of dimensions was found to vary by age group, highlighting the importance of age-related differences in health state valuations. The PedsUtil scoring system was developed to value HRQoL for 4 pediatric age groups (i.e., 2-4 years, 5-7 years, 8-12 years, 13-18 years).

This dissertation produced the first preference-based HRQoL measure for children 2-18 years. The development of the PedsUtil scoring system enables the inclusion of quality of life impacts for children across a full range of ages in economic and value-based research for the first time, ultimately allowing decision-makers to support more equitable and efficient use of child health services.

Chapter 1: Introduction and Motivation

Early childhood investment can be integral to ensuring a healthy life. Therefore, it is critical that resources are allocated efficiently across age groups. Economic evaluations, such as cost-effectiveness analyses, are routinely used by many countries to make evidence-based decisions for resource allocation and healthcare agenda setting. Economic evaluation methods for assessing adult-focused interventions are well established,[1,2] but many have not been validated for use in children. Childhood is a unique period of life which has implications for the design, conduct, interpretation, and implementation of economic evaluations of child health interventions.[3-6] Examples of characteristics that distinguish child health from adult health include: 1) the rapid rate of developmental change in children[3,7-9]; 2) the dependency of children on parents and adults to access care[10-12]; 3) children's unique patterns of health, disease, and disability[13-15]; and 4) children's demographic patterns which impact healthcare resource use.[16-18] Such child-adult differences make existing guidelines on the conduct of health economic evaluations difficult to apply to pediatric populations. Select methodological challenges include how to incorporate differences in outcomes by age and developmental stage, the scarcity of long-term child-specific data, how to value productivity for children, how to measure and integrate family and caregiver spillover effects, and how to best value child health.[3,4,7,19-26] While research is underway to develop recommendations around best practices for conducting child health economic evaluations, many gaps remain to be addressed. This dissertation addresses one of the key limitations in child health research by constructing a novel measure of child health-related quality of life (HRQoL) that is suitable for use in economic evaluations.

Health-Related Quality of Life

HRQoL can be defined as "the value assigned to duration of life as modified by impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy."[27] HRQoL incorporates the idea that health is a major component in determining overall well-being while recognizing that elements other than health may also affect overall quality of life.[1,2]

There are various approaches to measuring HRQoL These approaches can be broadly categorized as either health status measures (i.e., non-preference-based measures) or preference-based measures. Health status measures describe the presence, absence, severity, duration, and/or frequency of specific symptoms, diseases, or disabilities.[22] Examples of health status measures include the 36-Item Short Form Health Survey (SF-36),[28] Pediatric Quality of Life Inventory (PedsQL),[29] and the Child Health Questionnaire (CHQ).[30] Many health status measures have a simple summative scoring system that is used to assess health across various dimensions, such as physical, mental, social, and emotional health. However, there are limitations to directly using health status measures in economic evaluations. For example, Treatment A may perform better than Treatment B on some dimensions of health but perform worse on others. This means that Treatment A could have a more favorable cost per unit of health gain on some dimensions of health but be less favorable on others. Additionally, all dimensions of health are generally weighted equally in health status measures, but some people may place more value on certain dimensions than others.[31]

On the other hand, preference-based measures, such as health utility assessment, provide a summary score for a respondent's valuation of the quality of life associated with a specific health state. Health utilities provide a way of summarizing morbidity and mortality into a single metric.[32-34] Health utility values are typically scaled from 0, which represents being dead, to 1, which represents perfect health. Health utility values also exhibit interval-scaling, such that a change in health from 0.1 to 0.2 is equivalent in value to a change from 0.8 to 0.9.[35] These preference-based measures of HRQoL are used to calculate quality-adjusted life years (QALYs), which are a standard measure of health benefit used in cost-effectiveness analyses.

Quality Adjusted Life Years

QALYs represent survival, adjusted to reflect the HRQoL of health states.[34-36] A major advantage of QALYs is that they provide a common metric to compare diverse health outcomes, which facilitates the use of cost-effectiveness analyses to aid in decision-making across different interventions. QALYs are estimated by multiplying the time spent in a health state by the health utility or preference weight for that health state:

$$U(q,y) = H(q) \times \ y \qquad \textbf{Eq. 1.1}$$

$H(q)$ is the health utility function for a health state and y is the time spent in that health state (usually measured in years). For example, 1 QALY is considered equivalent to 1 year in perfect health. There are three main assumptions that are needed to calculate QALYs.[33,37,38] The first assumption is utility independence between life years and health status. For example, if one is indifferent between the choice of living 3 years for certain in a given health state and a 50:50 lottery between living 1 year or 6 years at the same health state, then one would also be indifferent between these alternatives for any other health state.[33] The second assumption is constant proportional trade-off of life years for health. For example, if one is willing to give up 2 years of their remaining 10 years of life (20%) in order to achieve some defined improvement in their health, then one would also be willing to sacrifice the same proportion of life years (4 years) if faced with a remaining life span of 20 years in order to gain the same improvement in health.[33] The third assumption is risk neutrality with respect to life years, which requires the utility function over life years to be linear.[33,38] Though there have been multiple studies that have discussed violations to these assumptions,[39-41] QALYs remain the standard metric for valuing health in cost-effectiveness analyses.

Techniques to Measure Health Utilities

There are two primary approaches to measure health utilities or preference weights ($H(q)$ in **Eq. 1.1**) that are used to calculate QALYs in cost-effectiveness analyses (**Table 1.1**). The first is direct valuation, in which preferences for health states are elicited either by patients valuing their own health or by community members valuing a hypothetical vignette. Direct elicitation methods include cardinal[32,34,42] (e.g., standard gamble [SG], time trade-off [TTO]) and ordinal methods[43-47] (e.g., discrete choice experiment [DCE], best-worst scaling [BWS]).

The second is through the use of generic preference-based measures of health, in which patients rate their health using a health state classification system that is defined by a set of health dimensions and levels. A preexisting scoring system or value set is then used to indirectly estimate the health utilities of health states defined by the health state classification system.[48,49] The scoring system is usually obtained by asking members of the general public to value a sample of health states defined by the health state classification system using direct valuation methods. Examples of generic preference-based measures of health include the EuroQol-5 Dimension (EQ-5D)[50] and the Child Health Utility 9D Index (CHU-9D).[51]

Table 1.1 Summary of Techniques to Measure Health Utilities

	Direct Methods: Cardinal (e.g., SG, TTO) and Ordinal (e.g., DCE, BWS)	**Indirect Methods**: Generic Preference-Based Measures of Health
Community Perspective	Individual values hypothetical health states	Community sample values health states described by a health state classification system; Patient completes health state classification system
Patient Perspective	Individual values own current health	Patient sample values health states described by health state classification system (including own current health); Patient completes health state classification system

Abbreviations: SG, standard gamble; TTO, time trade-off; DCE, discrete choice experiment; BWS, best worst scaling.

A variant of the second approach is the development of a health utility scoring system for an existing non-preference-based health status measure.[31,52] One major benefit of this method is that it allows for health utilities to be estimated directly from the existing health status measure without the need for additional data collection if the existing measure is already being used as part of a study protocol. The Short-Form Six-Dimension health index (SF-6D) was developed from the SF-36 using this technique and allows for the estimation of health utilities for any study in which responses to the SF-36 have already been collected, such as in clinical trials.[53] This considerably expands the scope for obtaining health utility values for use in economic evaluations in healthcare.

Challenges in Valuing Child Health

Accurately valuing the health of children is crucial for decision-makers to make consistent and strong policy recommendations. However, an important methodological challenge that arises in conducting child health economic evaluations is applying standard health valuation methods to estimate health utilities for pediatric populations. There are special considerations for children that pose a number of additional challenges when valuing child health compared to valuing adult health. For example, health valuation tasks require identifying relevant dimensions of health. However, as mentioned earlier, childhood is a unique period of time characterized by rapid developmental changes. This makes it difficult to identify a common set of dimensions that can be applied across multiple pediatric age groups from infancy to late adolescence.[54-56] It is especially difficult to characterize the experiences of very young children due to their underdeveloped physical, social, emotional and cognitive abilities.[55] Due to these challenges, many researchers have focused on valuing child health for narrow age bands and predominantly for children older than 5 years. However, this introduces bias when making comparisons across age groups due to the wide variability in valuation methods.

A second methodological challenge in valuing child health is the necessity of using proxy respondents. Studies have demonstrated that direct elicitation methods, such as SG and TTO, require respondents to be at a 6th grade reading level or higher, implying that children younger than 12 years old require proxy respondents to complete such tasks.[57] Even the use of indirect measures, which are cognitively less demanding than direct approaches to health state valuation, may pose comprehension challenges for young children and require proxy respondents. However, studies have shown that there are systematic differences in HRQoL values obtained from child self-report versus parent proxy-report.[58-66] For example, a review by Eiser and Morse[59] found that parents and children have greater agreement rating observable dimensions of HRQoL (e.g., physical functioning), but have poorer agreement for non-observable dimensions of HRQoL (e.g., social or emotional functioning). Similarly, Saigal et al.[61] found limited concordance between parent proxy-report and child self-report for health utilities elicited using SG, with adolescents providing lower valuations than their parents. Sung et al.[60] also found that health utilities elicited by TTO were significantly different between parent proxy-report and child self-report, though TTO health utilities were higher for children than their parents. There are a multitude of reasons for why such differences in valuing child health may exist between proxy-report and child self-report, including differences in respondents' attitudes toward risk, perception of time, and/or understanding and experience with health and illness.[22] These factors which influence child health state valuations may also be age dependent and change as children progress.[67] Moreover, parent valuations of child health states may be influenced by competing priorities such as other children in the family or parental guilt.[22] Therefore, health utilities derived from adult or proxy preferences are not interchangeable with those derived from children, which has important implications for decision-making based on cost-effectiveness analyses.

A third challenge to valuing child health is determining the perspective from which respondents are asked to assess health states. For example, an adult proxy respondent could be asked to value their own health as an adult (self-adult), value their own health imagining they were a child (self-child), value the health of a hypothetical child (other-child), or value the health of another adult (other-adult). Previous research has demonstrated that the perspective prescribed influences health state valuations.[68-71] For example, Kreimeier et al.[68] found some evidence for higher valuations from the other-child perspective compared to the self-adult perspective. Shah et al.[69] reached similar conclusions. In contrast, Kind et al.[70] found evidence for lower valuations from the other-child perspective compared to the self-adult and other-adult perspectives. And Lipman et al.[71] found consistent differences between child and adult perspectives, though the direction of the effect depended on the specific health state valued. Therefore, these studies suggest that using different perspectives yields systematic differences in health state valuations, though it is difficult to predict the scale and direction of the discrepancy. Relatedly, framing effects are also important to consider.[72] For example, the framing recommended for valuing child health states defined by the EuroQol-5 Dimension Youth (EQ-5D-Y) asks adults to consider what they think is best for a 10-year-old child.[73] However, rather than having proxies paternalistically decide what is best for the child, valuation tasks could ask proxies to value health states from the point of view of the child. Lipman et al.[74] found that the use of these different proxy framing approaches yielded differences in EQ-5D-Y valuations for severe health states. These diverse methods to valuing child health highlight the need for greater empirical evidence to better inform the ongoing normative discussion around whose preferences should be elicited and from which perspective and framing respondents should value child health states.

Problem Statement

Given the numerous methodological challenges of valuing child health, there has been a lack of detailed guidelines for conducting economic evaluations for child health interventions. A critical limitation is the paucity of child-specific preference-based measures of HRQoL. In particular, there is not a valid and reliable preference-based measure of HRQoL that can be applied across multiple pediatric age groups. This presents a major problem as many child health economic evaluations include outcomes for a wide range of ages, thus requiring the use of a variety of measures and approaches, including adult-specific instruments, to value child health.

A number of reviews have been published that provide an overview of the available generic preference-based measures of HRQoL which have been developed or adapted for use in children.[75-80] **Table 1.2** summarizes these generic preference-based measures of child HRQoL with existing scoring systems. However, as shown in the table, these measures apply only to limited age groups, and very few are available to value HRQoL in children younger than 5 years. Furthermore, there is no consensus on the methodology used to generate the value sets for these child-specific measures. The preference elicitation techniques used to derive these value sets vary considerably, and research has demonstrated that different elicitation techniques often do not yield the same health utility values.[81-83] In addition, the choice of whose values and which perspective to use in the elicitation tasks varies widely across measures. There is even variation within the same measure, such as the CHU-9D, for valuations implemented in different countries. Therefore, combining multiple measures of HRQoL to derive health utilities for use in child health economic evaluations may introduce bias and lead to more uncertainty around the value of child health interventions. This means that studies may be undervaluing (or overvaluing) certain child health interventions, compromising the ability of healthcare systems to be more efficient and equitable. Consequently, there is a need for a single preference-based measure of HRQoL that can be applied across multiple pediatric age groups in order to achieve greater consistency in valuing child health.

Table 1.2 Summary of Child-Specific Generic Preference-Based Measures of Health

Measure	Country[a]	Age[b]	Elicitation method	Whose preferences?	Which perspective?	Age of child in valuation task[c]	References
16D	Finland	12-15 years	VAS	Child	Self	12-15 years old	Apajasalo,[84] 1996
17D	Finland	8-11 years	VAS	Adult	Hypothetical child	8-11 years old	Apajasalo,[85] 1996
AHUM	UK	12-18 years	TTO	Adult	Self	N/A	Beusterien,[86] 2012
AQoL-6D	Australia, New Zealand, Fiji, Tonga	15-17 years	TTO	Child	Self	15-17 years old	Moodie,[87] 2010
CHU-9D	UK	7-17 years[d]	SG	Adult	Self	N/A	Stevens,[51] 2012
	Australia		BWS	Child	Self	11-17 years old	Ratcliffe,[88] 2016
	Netherlands		DCE	Adult	Self	N/A	Rowen,[89] 2018
	China		BWS	Child	Self	9-17 years old	Chen,[90] 2019
EQ-5D-Y	Slovenia	8-15 years[e]	DCE	Adult	Hypothetical child	10 years old	Prevolnik Rupel,[91] 2021
	Japan						Shiroiwa,[92] 2021
	Spain						Ramos-Goñi,[93] 2022
	Germany						Kreimeier,[94] 2022
	Hungary						Rencz,[95] 2022
	Netherlands						Roudijk,[96] 2022
	Belgium						DeWilde,[97] 2022
	Indonesia						Fitriana,[98] 2022
	China						Yang,[99] 2022
HUI-2	Canada	≥5 years	SG	Adult	Self as child	10 years old	Torrance,[100] 1996
	UK						McCabe,[101] 2005
HUI-3	Canada	≥5 years	SG	Adolescents (16 years and older) and adults	Self	N/A	Feeny,[102] 2002
IQI	Hong Kong, UK, USA	0-12 months	DCE	Adult	Hypothetical child	0-12 months old	Krabbe,[103] 2020
QWB	USA	Unclear	VAS	Adult	Self	N/A	Seiber,[104] 2008

Abbreviations: 16D, 16-Dimension; 17D, 17-Dimension; AHUM, Adolescent Health Utility Measure; AQoL-6D, Assessment of Quality of Life-6 Dimensions; Child Health Utility 9D Index; EQ-5D-Y, EuroQol-5 Dimension Youth; HUI-2, Health Utilities Index 2; HUI-3, Health Utilities Index 3; IQI, Infant health-related Quality of life Instrument; QWB, Quality of Well-Being; VAS, visual analog scale; TTO, time trade-off; SG, standard gamble; BWS, best-worst scaling; DCE, discrete choice experiment; N/A, not applicable.

[a] Available country-specific preferences.

[b] Ages considered appropriate for use with the measure.

[c] Age of child that study participants were asked to consider in valuation tasks.

[d] Originally developed for use with children 7-11 years old but mostly applied to adolescents 11-17 years old.

[e] Proxy version available for children ages 4-7 years.

Contribution to Literature

The overall objective of this dissertation was to construct a novel preference-based HRQoL measure for children. Specifically, a valuation protocol was developed and applied in order to estimate a health utility scoring system for the PedsQL: the PedsUtil scoring system. The PedsQL is a widely used *non*-preference-based measure of HRQoL validated for children 2-18 years.[29] While the PedsQL cannot be directly used in economic evaluations because it is not preference-based, deriving a preference-based scoring system for the PedsQL allows for health utilities to be estimated directly from the PedsQL. The PedsUtil scoring system was developed to apply across multiple pediatric age groups, but separate preference weights were derived for 4 different age groups (i.e., 2-4 years, 5-7 years, 8-12 years, 13-18 years) in order to account for the significant changes that occur at different developmental stages of children.

The construction of the PedsUtil scoring system addresses a crucial limitation in measures applied to value assessment in healthcare by providing a method to consistently and accurately value child health. This dissertation is the first study to derive a preference-based measure of HRQoL that can estimate health utilities for children across a full range of ages – from young children to adolescents. The development of the PedsUtil scoring system will ultimately allow decision-makers to support more equitable and efficient use of child health services.

Overview of Chapters

Chapter 2: Establishing the Core Dimension Structure of the PedsUtil Health State Classification System Based on the PedsQL

The PedsQL includes more items than is manageable to value in the preference valuation exercise required to derive the PedsUtil scoring system. Therefore, it is necessary to first reduce the length of the original PedsQL to a core set of dimensions and items to form the PedsUtil health state classification system. Chapter 2 describes the first step of this reduction process, which was to establish instrument dimensionality of the PedsUtil health state classification system using confirmatory factor analysis.

Chapter 3: Item Reduction of the PedsQL to Derive the PedsUtil Health State Classification System

Once the dimension structure of the PedsUtil health state classification system was established, the most representative item within each dimension was selected for. Chapter 3 describes how the design of the PedsUtil health state classification was finalized and validated by conducting Rasch and psychometric analyses, expert panels, and key informant interviews.

Chapter 4: Developing a Valuation Protocol to Construct the PedsUtil Scoring System Using a Discrete Choice Experiment with Time Trade-Off

Once the PedsUtil health state classification system was created, a valuation protocol was developed and implemented to value child health states defined by the PedsUtil health state classification system in order to derive preference weights for the preliminary construction of the PedsUtil scoring system. Chapter 4 describes the development of the valuation protocol and how the provisional PedsUtil scoring system was estimated using pilot data.

Chapter 5: Conclusion

Chapter 5 provides a summary of contributions and suggestions for future research on valuing child health.

Chapter 2: Establishing the Core Dimension Structure of the PedsUtil Health State Classification System Based on the PedsQL

Introduction

Preference-based measures of HRQoL that can be utilized to calculate QALYs in cost-effectiveness analyses are comprised of two parts. The first is a health state classification system that is made up of various dimensions of HRQoL, and each dimension is defined by a number of levels. For example, the EQ-5D-Y-3L describes 5 dimensions of HRQoL (mobility, looking after myself, doing usual activities, having pain or discomfort, and feeling worried, sad or unhappy) and each dimension spans 3 severity levels (no problems, some problems, and a lot of problems).[105] A health state is formed by selecting 1 level from each dimension in the health state classification system, thus the EQ-5D-Y-3L is comprised of $3^5 = 243$ unique health states. The second part is the scoring system or value set. The scoring system assigns a health utility value to each unique health state defined by the health state classification system, and this scoring system is derived through a direct valuation exercise, typically from a community-based sample.

The PedsQL is one of the most widely used HRQoL instruments in pediatric clinical trials.[29] However, as mentioned previously, the PedsQL is not a preference-based measure, thus cannot be directly used in cost-effectiveness analyses. One solution to expand its use into cost-effectiveness analyses is to develop a health utility scoring system for the PedsQL: the PedsUtil scoring system. Constructing the scoring system is a multi-step process that first requires deriving a health state classification system that is amenable to preference elicitation methods. Specifically, the PedsQL includes more items than is manageable to value in the preference elicitation exercise required to construct the PedsUtil scoring system. With 23 items with 5 levels each ranging from "Never" to "Almost always", the PedsQL in its entirety would generate 5^{23} health states which are too many to value. Therefore, it is necessary to reduce the length of the original instrument to a core set of dimensions and items to derive the PedsUtil health state classification system.

The first step of deriving the PedsUtil health state classification system is to establish instrument dimensionality. A requirement of a multidimensional health state classification system is that the dimensions are sufficiently independent in order to avoid nonsensical health states.[31] This means that there should be minimal correlation between dimensions. One method to help identify structurally independent dimensions is factor analysis. Therefore, the purpose of this study was to establish the core dimension structure of the PedsUtil health state classification system by using confirmatory factor analysis.

Methods

The PedsQL

The PedsQL 4.0 Generic Core Scales is a modular instrument that measures HRQoL in children and adolescents from ages 2 to 18 years.[29,106] It was developed through focus groups and cognitive interviews, and consists of child self-report and parent proxy-report versions. The child self-report version includes age groups 5-7 years, 8-12 years and 13-18 years. The parent proxy-report version includes age groups 2-4 years, 5-7 years, 8-12 years, and 13-18 years. The items between the different versions differ only in developmentally appropriate vocabulary and first or third person tense. A 5-point response scale (0 = Never; 1 = Almost never; 2 = Sometimes; 3 = Often; 4 = Almost always) is used for child self-report ages 8 years and older, and for all age groups with the parent proxy-report. For child self-report ages 5 to 7 years, a 3-point response scale (0 = Not at all; 2 = Sometimes; 4 = A lot) is used to increase comprehension. A child self-report version is understandably not available for ages 2 to 4 years.

Overall, there are 21 to 23 multi-level items that fall under 4 different dimensions or subscales: 1) Physical Functioning; 2) Emotional Functioning; 3) Social Functioning; and 4) School Functioning (**Table 2.1**). The items are reverse-scored and linearly transformed to a 0 to 100 scale (0 = 100; 1 = 75; 2 = 50; 3 = 25; 4 = 0), where higher scores indicate better HRQoL. The total HRQoL score is calculated as the sum of the item scores divided by the number of items answered.[29]

Table 2.1 Summary of PedsQL

Dimensions	Number of Items			
	2-4 years	5-7 years	8-12 years	13-18 years
Physical Functioning	8	8	8	8
Emotional Functioning	5	5	5	5
Social Functioning	5	5	5	5
School Functioning	3	5	5	5
Total	21	23	23	23

Data Source

This study used data from the Longitudinal Study of Australian Children (LSAC).[107] The LSAC is a study that follows a representative sample of 10,000 children and their families from all parts of Australia. It delivers a comprehensive national dataset on children as they age, and it is one of the very few large-scale nationally representative surveys of children in the world. The LSAC adopted a dual cohort cross-sectional sequential design. Data are collected from 2 cohorts every 2 years. The first cohort (infant cohort) consists of 5,000 children that were 0-1 years old in 2003. The second cohort (child cohort) consists of 5,000 children that were 4-5 years old in 2003.[108] The LSAC dataset contains responses to a very large number of questions, including the parent proxy-report version of the PedsQL. It is important to note that for children aged 2-3 years, responses to only 19 out of the 21 PedsQL items were collected as part of the LSAC. The 2 items omitted[*] were the items about school absence. Therefore, all analyses conducted in this study for children aged 2-3 years only utilized the 19 items that were administered. For all other age groups, responses to the full PedsQL were available. A cross-sectional master dataset (*n* = 45,207) was created using data from

[*] The omitted items were "missing school/daycare because of not feeling well" and "missing school/daycare to go to the doctor or hospital."

the first 7 waves (2003-04 to 2015-16) of the LSAC in order to obtain fully completed PedsQL responses for children aged 2 to 17 years for use in this analysis.

Statistical Analysis – Confirmatory Factor Analysis

Factor analysis tries to identify unobserved latent variables that explain patterns of correlation among observed variables.[109] There are different types of factor analysis, namely exploratory and confirmatory factor analyses.[110,111] Because the PedsQL is a well validated instrument with an established dimensional structure, confirmatory factor analysis was determined to be the more appropriate method to identify the core dimension structure of the PedsUtil health state classification system. The main reason was because the confirmatory approach allows for a specific dimensional structure (the conceptual model) to be tested. The exploratory approach, on the other hand, does not incorporate any prior assumptions about the dimension structure into the analysis. This can be problematic since it does not guarantee that the identified dimension structure will be clinically coherent. Therefore, by applying confirmatory factor analysis, theoretically or clinically driven decisions can be integrated into the general method of item assessment at the outset, rather than as an afterthought.[110]

Four main competing conceptual models of the PedsUtil health state classification system were developed a priori, drawing on published literature and expert opinion (**Table 2.2**). The first conceptual model (Model A) was the original 4-dimension structure of the PedsQL. The second conceptual model (Model B) was comprised of 5 dimensions. Model B was specified based on previously published literature that found model fit to be better when School Functioning items were split into 2 separate dimensions.[29,106,112] Specifically, the 5-item School Functioning dimension was split into 3 items measuring school cognitive functioning and 2 items measuring school absence as related to illness. The third conceptual model (Model C) included 6 dimensions, 2 of which were single-item dimensions. Model C was constructed based on expert opinion and by referencing other available preference-based HRQoL instruments for children,[51,100,102,105,113] which suggested that the items measuring Pain and Fatigue (originally in the Physical Functioning dimension) may be clinically important to separate out as independent dimensions. The fourth conceptual model (Model D) was composed of 7 dimensions, and was formed by combining Models B and C.

Table 2.2 Summary of Dimension Structures for Models A-D

Item[a] *Problems with...*	**Model A** (4 Dimensions)	**Model B** (5 Dimensions)	**Model C** (6 Dimensions)	**Model D** (7 Dimensions)
Phys 1. Walking	Physical	Physical	Physical	Physical
Phys 2. Running	Physical	Physical	Physical	Physical
Phys 3. Participating in exercise	Physical	Physical	Physical	Physical
Phys 4. Lifting something heavy	Physical	Physical	Physical	Physical
Phys 5. Taking a bath or shower	Physical	Physical	Physical	Physical
Phys 6. Doing chores	Physical	Physical	Physical	Physical
Phys 7. Having hurts or aches	Physical	Physical	Pain	Pain
Phys 8. Low energy level	Physical	Physical	Fatigue	Fatigue
Emot 1. Feeling afraid or scared	Emotional	Emotional	Emotional	Emotional
Emot 2. Feeling sad or blue	Emotional	Emotional	Emotional	Emotional
Emot 3. Feeling angry	Emotional	Emotional	Emotional	Emotional
Emot 4. Trouble sleeping	Emotional	Emotional	Emotional	Emotional
Emot 5. Worrying	Emotional	Emotional	Emotional	Emotional
Soc 1. Getting along with others	Social	Social	Social	Social
Soc 2. Others not wanting to be friends	Social	Social	Social	Social
Soc 3. Getting teased	Social	Social	Social	Social
Soc 4. Unable to do things others can do	Social	Social	Social	Social
Soc 5. Keeping up with other children	Social	Social	Social	Social
School 1. Paying attention in class	School	School	School	School
School 2. Forgetting things	School	School	School	School
School 3. Keeping up with schoolwork	School	School	School	School
SchAbs 1. Missing school because sick	School	School Absence	School	School Absence
SchAbs 2. Missing school to go to doctor	School	School Absence	School	School Absence

Abbreviations: Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.

[a] Item descriptions are summarized in the table (exact wording of items not displayed).

The 4 conceptual models were evaluated using the robust weighted least squares estimation method, which is recommended for modeling ordinal data.[114] Single-item dimensions (i.e., Pain and Fatigue) that were included in the conceptual models were deemed clinically important to be considered for inclusion in the PedsUtil health state classification system, but could not be empirically tested using confirmatory factor analysis. In other words, only multi-item dimensions were analyzed because confirmatory factor analysis requires at least 2 items in each dimension in order to estimate the measurement model. In addition, the PedsUtil health state classification should ideally be applicable across diverse pediatric populations. Therefore, this study stratified the analyses by age group and child health status in order to identify a common dimension structure across all subgroups. Age groups were stratified by 2-year age intervals (from 2 to 17 years) to reflect the study design of the LSAC. Child health status was defined as children with special healthcare needs[†] or typically functioning children[‡]. This study examined standardized factor loadings and evaluated the fit of the models using the comparative fit index (CFI),[115] Tucker-Lewis Index (TLI),[116] and the root mean square error of approximation (RMSEA).[117] Previously established guidelines suggest adequate fitting models have CFI and TLI values ≥0.90 and RMSEA values ≤0.08.[118] Modification indices and residual correlations were examined to re-specify the models to improve model fit when necessary. Models were modified and re-fit until a

[†] Children with special healthcare needs were identified as children whose parents responded "Yes" in the last wave of the LSAC that data were available for to the following question: "Child has a condition which has lasted or is expected to last for at least 12 months, which causes the child to use medicine prescribed by a doctor, other than vitamins, or use more medical care, mental health or educational services." Data from the last available wave was used to determine child health status because younger children are less likely to be identified with special healthcare needs.

[‡] Typically functioning children excluded children with special healthcare needs.

conceptually meaningful model was achieved across all subgroups that also adequately fit the data. All analyses were conducted in Mplus v8.[114] This study was determined to be exempt by the University of Michigan Institutional Review Board (IRBMED # HUM00182088).

Results

Sample Characteristics

Table 2.3 presents characteristics of the LSAC participants across child health status groups. As shown in the table, there were statistically significant differences between child health status groups for child age, child sex, parent age, and parent education. And as expected, the PedsQL scale scores were significantly lower for children with special healthcare needs than for typically functioning children. Sex of the parent was not significantly different across child health status groups, and most of the parents (96%) who answered the PedsQL were female.

Table 2.3 Summary of LSAC Participants Across Child Health Status

Characteristic	Children with Special Healthcare Needs (*n* = 8,793)	Typically Functioning Children[a] (*n* = 36,414)	*p*-value[b]
CHILD			
Age, y			
Mean (SD)	9.1 (4.1)	8.8 (4.0)	<0.01
Age Distribution [*n* (%)]			
2-3 years	519 (5.9)	2,615 (7.2)	<0.01
4-5 years	1,330 (15.1)	5,862 (16.1)	
6-7 years	1,343 (15.3)	5,765 (15.8)	
8-9 years	1,380 (15.7)	5,938 (16.3)	
10-11 years	1,450 (16.5)	5,933 (16.3)	
12-13 years	1,315 (15.0)	5,434 (14.9)	
14-15 years	775 (8.8)	2,588 (7.1)	
16-17 years	681 (7.7)	2,279 (6.3)	
Sex [*n* (%)]			
Male	4,642 (52.8)	18,482 (50.8)	<0.01
Female	4,151 (47.2)	17,932 (49.2)	
PedsQL Scale Scores [mean (SD)]			
Physical Functioning	78.9 (17.2)	83.9 (14.7)	<0.01
Emotional Functioning	68.8 (18.0)	75.8 (15.3)	<0.01
Social Functioning	76.7 (19.9)	83.6 (15.7)	<0.01
School Functioning	73.0 (20.0)	79.8 (16.8)	<0.01
Total Score	74.8 (14.8)	81.0 (11.9)	<0.01
PARENT[c]			
Age, y			
Mean (SD)	40.3 (6.8)	39.9 (6.5)	<0.01
Sex [*n* (%)]			
Male (%)	344 (3.9)	1,366 (3.8)	0.48
Female (%)	8,449 (96.1)	35,048 (96.3)	
Education [*n* (%)]			
Less than high school	117 (1.3)	387 (1.1)	<0.01
Some high school	2,649 (30.2)	10,510 (28.9)	
High school graduate	2,807 (31.9)	12,431 (34.2)	
College degree	1,704 (19.4)	6,981 (19.2)	
Graduate degree	1,510 (17.2)	6,065 (16.7)	

Abbreviations: LSAC, Longitudinal Study of Australian Children; y, years; SD, standard deviation.
[a] Typically functioning children excluded children with special healthcare needs.
[b] *p*-values derived from chi-squared tests for categorical variables and t-tests for continuous variables.
[c] Parent that answered the PedsQL about their child.

Confirmatory Factor Analysis

Table 2.4 reports the confirmatory factor analysis fit indices for Models A-D across all age and child health status subgroups. Model D was the best fitting model and Model A was the worst fitting model for all subgroups. Models A-C generally did not meet the cutoffs for the various model fit indices for most subgroups, especially for children with special healthcare needs. On the other hand, Model D exhibited adequate model fit for most subgroups with CFI values that ranged from 0.929 to 0.954, TLI values from 0.916 to 0.946, and RMSEA values from 0.058 to 0.102. Additionally, despite the acceptable fit for Model D across most subgroups, model fit was slightly better for typically functioning children than for children with special healthcare needs.

Table 2.5 displays the standardized factor loadings for Model D for all subgroups. All items had salient factor loadings with factor loadings >0.4 and *p*-values <0.001, though most items had factor loadings >0.7. A general trend in factor loadings was observed across subgroups for certain items. For example, the factor loading for item Phys 4 ("lifting something heavy") increased as children got older. This may be because children have a greater opportunity to experience or express this kind of behavior as they get older, suggesting that the construct validity of some items may improve as children age.

Table 2.4 Confirmatory Factor Analysis Fit Indices for Models A-D Across Subroups

Fit Indices	Children with Special Healthcare Needs								Typically Functioning Children							
	2/3 years[a]	4/5 years	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years	2/3 years[a]	4/5 years	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years
Model A																
RMSEA	0.087	0.107	0.113	0.112	0.117	0.112	0.118	0.109	0.071	0.092	0.101	0.098	0.107	0.094	0.109	0.103
RMSEA	0.080-	0.104-	0.110-	0.109-	0.114-	0.109-	0.114-	0.105-	0.068-	0.091-	0.099-	0.097-	0.106-	0.092-	0.107-	0.101-
90% CI	0.093	0.111	0.116	0.115	0.120	0.115	0.122	0.114	0.074	0.094	0.102	0.100	0.108	0.095	0.111	0.105
CFI	0.907	0.881	0.885	0.879	0.892	0.898	0.909	0.908	0.918	0.887	0.902	0.889	0.896	0.906	0.922	0.902
TLI	0.892	0.863	0.870	0.864	0.878	0.885	0.897	0.896	0.905	0.870	0.890	0.875	0.882	0.894	0.912	0.890
Model B																
RMSEA	–	0.082	0.101	0.098	0.110	0.104	0.109	0.102	–	0.067	0.089	0.088	0.100	0.085	0.104	0.098
RMSEA	–	0.079-	0.098-	0.095-	0.107-	0.101-	0.105-	0.097-	–	0.066-	0.088-	0.087-	0.098-	0.083-	0.102-	0.095-
90% CI		0.086	0.104	0.101	0.113	0.107	0.113	0.106		0.069	0.091	0.090	0.101	0.087	0.106	0.100
CFI	–	0.912	0.910	0.908	0.906	0.914	0.923	0.921	–	0.929	0.925	0.913	0.911	0.924	0.931	0.914
TLI	–	0.899	0.897	0.895	0.892	0.901	0.912	0.909	–	0.917	0.914	0.900	0.898	0.913	0.920	0.901
Model C																
RMSEA	–	0.109	0.110	0.111	0.111	0.104	0.109	0.099	–	0.093	0.096	0.094	0.100	0.087	0.100	0.088
RMSEA	–	0.105-	0.107-	0.108-	0.108-	0.100-	0.104-	0.091-	–	0.092-	0.094-	0.093-	0.098-	0.085-	0.098-	0.085-
90% CI		0.113	0.113	0.114	0.115	0.107	0.113	0.100		0.095	0.096	0.096	0.102	0.089	0.102	0.090
CFI	–	0.898	0.909	0.898	0.917	0.925	0.934	0.930	–	0.901	0.925	0.912	0.923	0.929	0.945	0.937
TLI	–	0.881	0.896	0.882	0.904	0.913	0.925	0.920	–	0.885	0.913	0.899	0.912	0.919	0.937	0.928
Model D																
RMSEA	0.077	0.072	0.093	0.094	0.102	0.093	0.096	0.089	0.061	0.058	0.079	0.080	0.089	0.075	0.092	0.078
RMSEA	0.070-	0.068-	0.089-	0.090-	0.098-	0.090-	0.092-	0.084-	0.057-	0.057-	0.077-	0.078-	0.087-	0.073-	0.090-	0.076-
90% CI	0.085	0.076	0.096	0.097	0.105	0.097	0.101	0.094	0.064	0.060	0.080	0.081	0.090	0.076	0.095	0.081
CFI	0.937	0.942	0.937	0.929	0.932	0.941	0.950	0.948	0.947	0.953	0.950	0.939	0.941	0.949	0.954	0.951
TLI	0.925	0.932	0.926	0.916	0.920	0.930	0.941	0.939	0.937	0.945	0.942	0.928	0.930	0.940	0.946	0.943

Abbreviations: RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker-Lewis index; CI, confidence interval.

[a] For children aged 2-3 years, the measurement models for Models A and B and for Models C and D were the same because only 1 School Functioning item was included in the LSAC dataset.

Table 2.5 Factor Loadings for Model D Across Subgroups

Item	Children with Special Healthcare Needs								Typically Functioning Children							
	2/3 years	4/5 years	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years	2/3 years	4/5 years	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years
Phys 1	0.867	0.872	0.888	0.904	0.894	0.882	0.900	0.880	0.907	0.910	0.911	0.877	0.933	0.902	0.932	0.902
Phys 2	0.921	0.947	0.950	0.917	0.913	0.914	0.902	0.910	0.864	0.944	0.972	0.922	0.932	0.903	0.917	0.872
Phys 3	0.824	0.843	0.925	0.896	0.933	0.929	0.943	0.918	0.753	0.807	0.934	0.920	0.943	0.923	0.946	0.902
Phys 4	0.618	0.625	0.761	0.749	0.801	0.809	0.848	0.807	0.566	0.590	0.703	0.686	0.753	0.783	0.847	0.858
Phys 5	0.631	0.637	0.716	0.727	0.825	0.794	0.849	0.773	0.573	0.686	0.766	0.745	0.890	0.837	0.957	0.915
Phys 6	0.478	0.416	0.604	0.621	0.613	0.654	0.759	0.602	0.567	0.492	0.633	0.661	0.680	0.646	0.774	0.631
Pain[a]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Fatigue[a]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Emot 1	0.693	0.701	0.751	0.786	0.795	0.804	0.857	0.858	0.655	0.718	0.730	0.743	0.785	0.771	0.811	0.814
Emot 2	0.712	0.733	0.755	0.789	0.817	0.847	0.872	0.833	0.713	0.732	0.756	0.788	0.792	0.826	0.847	0.848
Emot 3	0.621	0.593	0.688	0.675	0.698	0.749	0.789	0.705	0.642	0.621	0.638	0.667	0.688	0.692	0.768	0.728
Emot 4	0.481	0.535	0.622	0.617	0.641	0.640	0.727	0.691	0.482	0.492	0.566	0.605	0.591	0.617	0.676	0.676
Emot 5	0.800	0.791	0.708	0.747	0.788	0.789	0.818	0.801	0.785	0.775	0.704	0.740	0.758	0.790	0.801	0.811
Soc 1	0.764	0.748	0.749	0.751	0.780	0.803	0.785	0.750	0.708	0.725	0.773	0.765	0.787	0.777	0.810	0.818
Soc 2	0.730	0.782	0.779	0.832	0.878	0.860	0.865	0.850	0.787	0.771	0.736	0.814	0.808	0.833	0.835	0.844
Soc 3	0.744	0.706	0.712	0.780	0.804	0.814	0.840	0.794	0.723	0.733	0.719	0.789	0.783	0.789	0.839	0.808
Soc 4	0.793	0.802	0.802	0.821	0.816	0.804	0.847	0.859	0.764	0.735	0.695	0.749	0.714	0.781	0.801	0.784
Soc 5	0.838	0.821	0.872	0.859	0.855	0.835	0.870	0.902	0.756	0.767	0.892	0.823	0.877	0.844	0.854	0.863
School 1[b]	N/A	N/A	0.857	0.857	0.867	0.892	0.883	0.903	N/A	N/A	0.838	0.862	0.848	0.877	0.918	0.874
School 2[b]	N/A	N/A	0.716	0.752	0.729	0.764	0.766	0.755	N/A	N/A	0.646	0.702	0.660	0.765	0.744	0.751
School 3[c]	–	–	0.961	0.894	0.942	0.867	0.889	0.864	–	–	0.957	0.904	0.930	0.874	0.891	0.854
SchAbs 1[d]	–	0.865	0.795	0.792	0.784	0.752	0.876	0.877	–	0.873	0.785	0.811	0.823	0.812	0.827	0.874
SchAbs 2[d]	–	0.898	0.915	0.921	0.883	0.917	0.867	0.823	–	0.855	0.917	0.863	0.881	0.838	0.830	0.794

Abbreviations: Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence; N/A, not applicable.

[a] Pain and Fatigue are single item dimensions so could not be empirically tested using confirmatory factor analysis.

[b] School 1 and School 2 are not included in the parent proxy-report version of the PedsQL for young children.

[c] School 3 was a single item dimension for children aged 2-5 years since School 1 and School 2 are not included in the PedsQL for those age groups, thus the School Functioning dimension could not be empirically tested using confirmatory factor analysis for age groups 2-5 years.

[d] SchAbs 1 and SchAbs 2 were not administered for children aged 2-3 years in the LSAC.

Though model fit was generally adequate for Model D, the RMSEAs were >0.08 for select age groups 6 years and older for both children with special healthcare needs and typically functioning children. When modification indices and residual correlations were examined, items Soc 1 ("getting along with others"), Soc 4 ("unable to do things others can do"), and Soc 5 ("keeping up with other children") appeared to cross-load onto the Physical Functioning dimension for most age groups 6 years and older. Therefore, Model D was re-specified with these cross-loadings and the refined model fit indices for these age groups are shown in **Table 2.6**. Fit indices for the refined model indicated slightly better model fit for typically functioning children (CFI ≥0.96, TLI ≥0.952, and RMSEA ≤0.071) than for children with special healthcare needs (CFI ≥0.954, TLI ≥0.945, and RMSEA ≤0.076) across all age groups. Nevertheless, model fit improved across all subgroups, resulting in model fit indices that were considered to demonstrate acceptable model fit (i.e., cutoff criteria of CFI and TLI ≥0.90 and RMSEA ≤0.08).

Although overall model fit improved when items Soc 1, Soc 4, and Soc 5 were cross-loaded onto the Physical Functioning dimension, these items need to be allocated to a single dimension in order to create the PedsUtil health state classification system. Therefore, these items were retained in the Social Functioning dimension as originally hypothesized, and Model D was determined to be the core dimension structure for the PedsUtil health state classification system.

Table 2.6 Fit Indices for Re-specified Model D[a]

Fit Indices	Children with Special Healthcare Needs						Typically Functioning Children					
	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years	6/7 years	8/9 years	10/11 years	12/13 years	14/15 years	16/17 years
RMSEA	0.072	0.076	0.074	0.076	0.074	0.074	0.061	0.065	0.062	0.060	0.071	0.067
RMSEA 90% CI	0.068-0.075	0.073-0.080	0.071-0.077	0.072-0.079	0.069-0.079	0.069-0.079	0.060-0.063	0.063-0.067	0.061-0.064	0.058-0.061	0.068-0.073	0.064-0.070
CFI	0.963	0.954	0.965	0.961	0.971	0.964	0.970	0.960	0.971	0.968	0.973	0.965
TLI	0.955	0.945	0.958	0.954	0.965	0.957	0.965	0.952	0.966	0.962	0.968	0.958

Abbreviations: RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker-Lewis index; CI, confidence interval.
[a] Model D re-specified by cross-loading items Soc 1, Soc 4, and Soc 5 onto the Physical Functioning dimension.

Discussion

The objective of this study was to identify the core dimension structure of the PedsUtil health state classification system. The findings from this study support a 7-dimension structure (i.e., Physical Functioning, Pain, Fatigue, Emotional Functioning, Social Functioning, School Functioning, and School Absence) of the health state classification system. Moreover, this analysis suggests that the 7-dimension structure may be applicable across diverse pediatric populations, including children with special healthcare needs and typically functioning children, as well as children 2-17 years.

Previous studies have examined the dimension structure of the PedsQL using confirmatory factor analysis for various pediatric populations. Similar to this study, these studies found that splitting the 5 items in the original School Functioning dimension into 2 separate dimensions of School Functioning and School Absence was more appropriate.[112,119-123] However, none of the other studies evaluated the dimension structure of the PedsQL with Pain and Fatigue as single-item dimensions. Previous studies have also examined factorial invariance for the PedsQL across pediatric subpopulations. These studies similarly found that PedsQL items are comparable across age[119] and child health status subgroups,[120] as well as across various race,[121] gender,[122] and socioeconomic status subgroups.[123]

There are some limitations to this study. First, all data for this analysis came from an Australian population. The final PedsUtil scoring system will be developed using a US-representative population sample, and so analyzing data from a US-based sample would have been preferred when establishing the core dimension structure of the PedsUtil health state classification system. However, the LSAC dataset was used for this study because it is one of the most extensive pediatric datasets with responses to the PedsQL and no equivalent US dataset exists. Second, this analysis utilized responses to the parent proxy-report version of the PedsQL as that was the only available response type administered in the LSAC. Future research should validate dimensionality of the health state classification system using the child self-report version of the PedsQL. Third, data were not available for the 2 School Absence items for the youngest age group. Nevertheless, the 7-dimension model was found to be superior compared to the other 3 models for the closest comparable age group of 4-5 years. Fourth, this study did not model any single-item dimensions (i.e., Pain and Fatigue). It may be possible to model these single-item dimensions through estimating higher order models. However, this study did not take this approach because there are numerous higher order models that could be tested using confirmatory factor analysis, which may introduce conceptual issues regarding the causal structure amongst dimensions. Furthermore, these higher order models would not have much bearing on identifying the core dimension structure of the PedsUtil health state classification system, thus evaluating higher order models was deemed outside the scope of this analysis. Fifth, this study did not attempt to test for metric and scalar invariance across subgroups, which are stricter conditions of measurement invariance. Instead, only configural invariance was established by estimating measurement models for each subgroup separately. Assessing stricter conditions of measurement invariance was considered beyond the framework of this study since the goal of this study was to identify a conceptually sound dimension structure where each common dimension is associated with identical sets of items across various pediatric subgroups. Nevertheless, individual item functioning and item invariance across subgroups will be further examined in Chapter 3 using Rasch analysis. Lastly, there were statistically significant differences across subgroups for certain sample characteristics, such as child sex, parent age, and parent education. Despite these differences in characteristics, this study still identified a clinically coherent common dimension structure across subgroups.

In conclusion, this study established the core dimension structure of the PedsUtil health state classification system using confirmatory factor analysis. This is the first key step to developing the PedsUtil scoring system. The next step of deriving the PedsUtil health state classification system is to select the most representative item within each dimension.

Chapter 3: Item Reduction of the PedsQL to Derive the PedsUtil Health State Classification System

Introduction

Previous studies have adopted modern psychometric approaches to develop health state classification systems from existing non-preference-based measures.[31,109-111,124-126] This study applies and adapts these previously used methods in order to establish a framework for deriving the PedsUtil health state classification system. Chapter 2 described the first step of this process, which was to establish the core dimension structure of the PedsUtil health state classification system. This chapter describes the next step, which is to select the most representative item within each dimension in order to develop a health state classification system that is amenable to preference elicitation methods.

One useful technique in the process for item selection is Rasch analysis. Rasch analysis is a mathematical modeling technique that transforms categorical responses to points on a unidimensional continuous latent scale using a logistic model.[127] The response patterns are tested against what should be expected if measurement is to be achieved, which is a probabilistic form of Guttman scaling.[128] When applied to HRQoL measures, the Rasch model assumes that respondents with more severe HRQoL problems should indicate that they have difficulties with more items than respondents with less severe problems and vice versa. In addition, the model assumes that the easier an item is, the more likely it will be affirmed.[129-132] In terms of development of a health state classification system, Rasch analysis provides empirical evidence on how well items on a scale or dimension measure the construct of interest (e.g., physical functioning), and thus helps inform which items to include or exclude from the health state classification system. The objective of this study was to utilize Rasch analysis alongside other psychometric testing and qualitative work in order to select a subset of PedsQL items to include in the PedsUtil health state classification system.

Methods

Overview of Approach

A 2-step procedure was used to identify the PedsQL items to include in the PedsUtil health state classification system. Step 1 was to exclude any poorly functioning items in each dimension by examining various Rasch criteria (described in detail below). Since Rasch models assume unidimensionality, Rasch analysis was performed for each multi-item dimension (i.e., Physical Functioning, Emotional Functioning, Social Functioning, School Functioning, and School Absence) using the Rasch partial credit model.[133] Single-item dimensions (i.e., Pain and Fatigue) were not analyzed using Rasch analysis since these dimensions do not need to go through the process of item selection. Step 2 was to then select a single item to represent each dimension among the remaining items by using Rasch analysis and other psychometric analysis, as well as conducting expert panels and key informant interviews.

All secondary analyses were conducted using the LSAC dataset previously described in Chapter 2.[107] Please refer to the methods section of Chapter 2 for more detailed information about the LSAC. Empirical analyses were stratified by age group (i.e., 2-5 years, 6-13 years, and 14-17 years) in order to select items for the PedsUtil health state classification system that were suitable to use for a wide range of ages spanning from young children to adolescents. These specific age groupings were selected to represent the different developmental stages of children, as well as to reflect the study design of the LSAC. Rasch analyses were conducted using RUMM2030.[134] This study was granted an exempt determination by the University of Michigan Institutional Review Board (IRBMED # HUM00182088).

Step 1: Excluding Items Using Rasch Analysis

Data were fitted to the Rasch model to test how well the observed data meet the expectations of the measurement model. If there was any misfit, adjustments were made until a well-fitting model was achieved, but items that exhibited misfit were considered for exclusion from the PedsUtil health state classification system. Three main Rasch criteria were used to assess item performance in Step 1 – item level ordering, differential item functioning (DIF), and Rasch model goodness-of-fit.

Item Level Ordering

The pattern of item response thresholds was first examined to determine if disordering was present. For a well-fitting item to the Rasch model, each item response level (i.e., Never, Almost Never, Sometimes, Often, and Almost Always) should systematically take turns showing the highest probability of being chosen. However, item misfit occurs when respondents inconsistently use the response levels. In other words, respondents are unable to distinguish between the item response levels. This can occur if there are too many item response levels or when the levels are poorly labeled or open to misinterpretation.[131] Items that exhibit disordered thresholds fail to respond to the full range of severity across the dimension being measured, thus are not ideal to include in a health state classification system.

The ordering of thresholds was evaluated graphically using individual item threshold probability curves. For items that exhibited disordered thresholds, ordering of items was achieved by collapsing adjacent item response levels. If there was more than one possible combination to merge item response levels, the distribution of responses across the item response levels was analyzed to help guide which levels to merge.[*] The combination with the best overall fit to the Rasch model was selected. Disordered items were assessed for exclusion from the PedsUtil health state classification system.

Differential Item Functioning

Once all items were ordered, the Rasch model was used to test for DIF. DIF is a form of bias where responses to items systematically differ across respondent characteristics (e.g., males vs. females), despite having the same level of the underlying dimension being measured.[130] For example, males and females with equal levels of physical ability may systematically respond differently to item(s) on the Physical Functioning dimension.

[*] The item response levels were merged to achieve a more balanced distribution of responses across levels when possible.

However, any difference in scores observed for items exhibiting DIF is biased because the difference may be an artifact of measurement nonequivalence rather than reflecting a true difference in scores between subgroups. Therefore, assessment of DIF yields crucial information about measurement invariance and can be used to determine cross-population validity of items.[110,130]

Previous studies have documented gender differences in HRQoL outcomes,[135,136] thus this study examined DIF by sex in order to ensure that meaningful comparisons between males and females could be made using the PedsUtil health state classification system. Similarly, DIF by child health status (children with special healthcare needs or typically functioning children) was examined since the PedsUtil health state classification system needs to apply across diverse pediatric populations. Two types of DIF were tested for in this study using analysis of variance – uniform and nonuniform DIF.[†,137] Uniform DIF is when groups show a consistent difference in their responses to an item across the entire severity range of the dimension being measured (e.g., male responses are always higher than female responses). Nonuniform DIF is when the differences in responses between groups is not constant across the severity range.[130,131] For any items exhibiting DIF, the items were separated into different person factors and the Rasch model was refit. For example, if an item exhibited DIF by sex, the responses were separated for males and females. If splitting the item did not improve model fit, the item was considered for removal from the Rasch model. Any items that exhibit DIF threaten construct validity and are of limited value for making cross-

† A *p*-value < 0.05 (with Bonferroni adjustment) was used to determine the presence of DIF.

population comparisons, thus were considered for exclusion from the PedsUtil health state classification system.

Rasch Model Goodness-of-Fit

After issues of disordered thresholds and DIF were resolved, Rasch model fit statistics were assessed to determine overall model goodness-of-fit. The overall model fit was specified by the item-trait interaction statistic, which was reported as a χ^2 statistic. The item-trait interaction statistic reflects the property of invariance and a well-fitting Rasch model should exhibit the property of invariance. Accordingly, a statistically significant χ^2 statistic (with a Bonferroni correction) indicates poor model fit. If overall model fit was statistically significant (i.e., p-value < 0.01), the fit of the individual items was examined via fit residuals and individual item χ^2 statistics. Fit residuals quantify the divergence between expected and observed responses and are standardized to approximate a *Z*-score, representing a standardized normal distribution. Items with fit residuals greater than the standard cutoff of $\pm$2.5 were considered to not fit the Rasch model.[111,126,131] And similar to the overall model goodness-of-fit statistic, items with individual χ^2 statistics that were statistically significant were also considered to not fit the Rasch model. In this analysis, poorly fitting items were dropped from the model sequentially, beginning with the worst fitting item. The Rasch model was refit after each item was removed. This process was repeated until only well-fitting items remained and the overall item-trait interaction statistic was nonsignificant. Items that were dropped from the Rasch model poorly represent the underlying dimension being measured, thus were considered for exclusion from the PedsUtil health state classification system.

Robustness Check

In order to enhance robustness, Rasch analysis was conducted on 5 subsamples of the LSAC dataset for each age group (i.e., 2-5 years, 6-13 years, 14-17 years) for a total of 15 subsamples. Stratified random sampling was used to obtain subsamples of approximately 500 responses in each, which is the recommended sample size for Rasch analysis.[138] Sampling was stratified on child sex, age, and health status. Each item was given a total score indicating the number of subsamples that the item performed well on all Rasch criteria. The total score for each item was out of 5 since there were 5 subsamples per age group that were fitted to the Rasch model. In general, any item that performed poorly across all 5 subsamples in any age group (i.e., score of 0/5) or was the worst fitting item in any age group (i.e., lowest total score) was excluded from the PedsUtil health state classification system.

Step 2: Selecting Items

Following Step 1, a single best item was selected for each dimension from the remaining items for inclusion in the PedsUtil health state classification system. A range of criteria was considered for item selection, which are described below.

Rasch Analysis

Individual item goodness-of-fit statistics were assessed, and the item with the better fit to the underlying Rasch model was generally considered to be the better item to represent the dimension. In addition, the spread of item thresholds was examined. An item that covers a wider severity range of the latent variable was considered to be a better item to represent the dimension than an item that covers a narrow range.

Psychometric Analysis

Internal consistency and floor and ceiling effects were also examined when selecting the best item for each dimension. Internal consistency was determined by estimating the correlation (i.e., Spearman's correlation coefficient) of an item score and its dimension score. If a low correlation was observed, then the item was deemed to not be a good representative of the dimension. Items exhibiting either floor or ceiling effects were also regarded to be poor candidates for inclusion in the PedsUtil health state classification system. Floor and ceiling effects were estimated by counting the frequencies of item response levels for each item, and a threshold of 30% was adopted in this study.

Expert Panels and Key Informant Interviews

Expert panels and key informant interviews were conducted to supplement Rasch and psychometric analyses for item selection and to establish content and face validity of the PedsUtil health state classification system. Previous studies on the development of health state classification systems have similarly engaged with various stakeholders to aid in item selection.[126,139]

A convenience sample of 6 pediatricians and clinical trialists were recruited to participate in expert panels and 12 parents were recruited to participate in key informant interviews. The clinicians included general pediatricians, as well as pediatric specialists, including a gastroenterologist and psychologist. The interviewees included parents of children with special healthcare needs and parents of typically functioning children from ages 2 to 17 years (**aTable 3.1**). The parents were recruited through referrals from expert panelists, online Facebook groups, and pediatrician offices affiliated with the University of Michigan Health System. All eligible participants were at least 18 years old and English speakers. Expert panelists were compensated $75 and parents were compensated $25 for their participation.

A semi-structured interview format was used to conduct the expert panels and key informant interviews. All interviews were conducted online. Two main types of questions were asked. In cases where Rasch and psychometric analyses were able to identify a single item to represent a dimension, participants were asked if they agreed with the item selected. In cases where Rasch and psychometric analyses were only able to exclude some items but not identify a single best item, participants were asked to choose which item they believed best represents the dimension among the remaining items and to provide justifications for their choices. Participants could also disagree with the item(s) excluded based on Rasch and psychometric criteria and could instead select a different item from the full list of items in each dimension as the best item to represent that dimension. **aFigure 3.1** in the Appendix provides sample questions regarding item selection that were presented to expert panelists and key informant interviewees.

Final Item Selection

The core research team considered results from all criteria listed above to make the final decisions for item selection. The PedsUtil health state classification system was also reviewed with a health status measurement expert to ensure that the items selected were cohesive and amenable to preference valuation required to construct the PedsUtil scoring system.

Results

<u>Step 1: Item Exclusion</u>

Table 3.1 displays the total scores indicating how many subsamples each item performed well on all Rasch criteria (i.e., item ordering, DIF, and item fit). **aTables 3.2-3.7** in the Appendix provide more detailed Rasch analysis results for each item.

Table 3.1 Summary of Item Performance on Rasch Criteria for Item Exclusion

Item Description	Total Score[a]			Item
Problems with…	2-5 years	6-13 years	14-17 years	Excluded
School Absence				
SchAbs 1. Missing school because sick	2/5[b]	4/5	3/5	
SchAbs 2. Missing school to go to doctor	0/5[b]	0/5	0/5	✓
School Functioning				
School 1. Paying attention in class	–[c]	0/5	2/5	✓
School 2. Forgetting things	–[c]	0/5	0/5	✓
School 3. Keeping up with schoolwork	Only item[d]	0/5	4/5	
Social Functioning				
Soc 1. Getting along with others	0/5	0/5	0/5	✓
Soc 2. Others not wanting to be friends	4/5	4/5	3/5	
Soc 3. Getting teased	4/5	1/5	2/5	
Soc 4. Unable to do things others can do	1/5	0/5	0/5	✓
Soc 5. Keeping up with other children	0/5	0/5	0/5	✓
Emotional Functioning				
Emot 1. Feeling afraid or scared	4/5	2/5	2/5	
Emot 2. Feeling sad or blue	1/5	5/5	0/5	✓
Emot 3. Feeling angry	4/5	1/5	1/5	
Emot 4. Trouble sleeping	2/5	0/5	2/5	✓
Emot 5. Worrying	4/5	5/5	5/5	
Physical Functioning[e]				
Phys 1. Walking	0/3[f]	1/5	1/3[f]	✓
Phys 2. Running	0/3	4/5	3/3	
Phys 3. Participating in exercise	0/3	4/5	3/3	
Phys 4. Lifting something heavy[g]	–	–	–	✓
Phys 5. Taking a bath or shower[g]	–	–	–	✓
Phys 6. Doing chores[g]	–	–	–	✓

Abbreviations: SchAbs, School Absence; School, School Functioning; Soc, Social Functioning; Emot, Emotional Functioning; Phys, Physical Functioning.

[a] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).

[b] SchAbs 1 and SchAbs 2 were not administered for children aged 2-3 years in the LSAC so results reflect responses for children aged 4-5 years.

[c] School 1 and School 2 are not included in the PedsQL for children under 5 years old.

[d] Only School 3 is included in the PedsQL for this age group.

[e] Most subsamples did not fit the Rasch model for the Physical Functioning dimension, thus supplemental Rasch analyses were conducted for items Phys 1-Phys 3. Results from the supplemental analyses are shown in the table.

[f] Insufficient sample size to obtain 5 subsamples so only 3 subsamples were created for supplemental analyses.

[g] Items Phys 4-Phys 6 were omitted from the supplemental Rasch analyses, thus total scores were not calculated for these items. Refer to Results section for more details.

For the School Absence dimension, none of the 5 subsamples across all age groups could be fitted to the Rasch model (i.e., item-trait interaction χ^2 statistic was statistically significant) (**aTable 3.2**). Nevertheless, SchAbs 2 performed worse across all age groups (total score 0/5) than SchAbs 1. The individual χ^2 statistics for SchAbs 2 were statistically significant for all subsamples across all age groups, indicating poor item fit to the Rasch model. On the other hand, the individual χ^2 statistics for SchAbs 1 were not statistically significant for most subsamples across age groups. Therefore, SchAbs 2 was excluded from the PedsUtil health state classification system.

Since the School Functioning dimension consisted of only 1 item for age group 2-5 years, Rasch analysis was not conducted for this dimension for this age group. For age group 6-13 years, none of the items performed well in any of the subsamples. School 1 and School 3 exhibited disordered thresholds and DIF and School 2 did not fit the Rasch model (**aTable 3.3**). Rasch analysis provided little insight for item selection for this age group. Consequently, Rasch analysis results for age group 14-17 years, as well as results from additional criteria in Step 2, were used to help guide item selection for this dimension for age group 6-13 years. For age group 14-17 years, School 2 did not fit the Rasch model in any of the subsamples, and so was excluded from the health state classification system. School 1 was also excluded at this stage of the analysis because School 1 is not a validated item for the School Functioning dimension of the PedsQL for children under 5 years old, and the PedsUtil health state classification system needs to apply across all age groups. Moreover, School 1 (total score 2/5) performed worse than School 3 (total score 4/5) for age group 14-17 years.

For the Social Functioning dimension, Soc 1 and Soc 5 did not perform well in any of the subsamples across all age groups (total score 0/5), and Soc 4 scored 0/5 for age groups 6-13 years and 14-17 years (**aTable 3.4**). These 3 items were excluded from the PedsUtil health state classification system.

Emot 2 was the worst performing item for age groups 2-5 years and 14-17 years in the Emotional Functioning dimension, with total scores of 1/5 and 0/5, respectively (**aTable 3.5**). Consequently, Emot 2 was excluded from the PedsUtil health state classification system. For age group 6-13 years, Emot 4 exhibited disordered thresholds and/or item misfit in all subsamples (total score 0/5), thus Emot 4 was also excluded.

For the Physical Functioning dimension, 4 out of the 5 subsamples did not fit the Rasch model for age groups 2-5 years and 6-13 years. For the 1 subsample that did fit the Rasch model for these age groups, only Phys 3 performed well on all Rasch criteria (**aTable 3.6**). For age group 14-17 years, all 5 subsamples fit the Rasch model, but Phys 1, Phys 5 and Phys 6 performed poorly in all subsamples (total score 0/5). Because most of the subsamples did not fit the Rasch model for the Physical Functioning dimension, supplemental Rasch analyses were performed on just items Phys 1-Phys 3, which were considered the most relevant items in this dimension according to expert judgement (**aTable 3.7**). As a result, items Phys 4-Phys 6 were excluded from the health state classification system. For age group 2-5 years, none of the supplemental subsamples fit the Rasch model, thus supplemental results from the other age groups were used to help with item exclusion. Phys 1 was the worst performing item for age groups 6-13 years (total score 1/5) and 14-17 years (total score 1/3[‡]) in the supplemental analyses, thus this item was excluded from the PedsUtil health state classification system.

Step 2: Item Selection

Table 3.2 provides a summary of results for Rasch and psychometric criteria for the remaining 9 items following Step 1.

‡ Total score was out of 3 instead of 5 for the supplemental analyses for age groups 2-5 years and 14-17 years because only 3 subsamples could be created due to limited sample size.

Table 3.2 Summary of Rasch and Psychometric Criteria for Remaining Items

Item Description *Problems with…*	Age Group	Rasch Criteria							Psychometric Criteria[a,b]		
		Total Score[c]	Mean Item Level Performance[d] (Range Across Subsamples): *p*-value[f]	Fit Residual[g]	Spread	Disordered Thresholds	DIF	Item Misfit	% Response Ceiling (Never)	% Response Floor (Almost always)	Corr.[e]
School Absence											
SchAbs 1. Missing school because sick	2-5y[h]	2/5	0.15 (0.10, 0.19)	0.37 (0.17, 0.57)	2.08 (1.86, 2.30)	1 sample	2 samples (Health status[i])	No misfit	58.2%	0.1%	0.91
	6-13y	4/5	0.06 (0.01, 0.15)	0.53 (0.33, 0.73)	1.67 (1.10, 2.03)	None	No DIF	1 sample	47.0%	0.3%	0.90
	14-17y	3/5	0.04 (0.02, 0.06)	0.09 (0.02, 0.16)	1.22 (1.14, 1.32)	None	No DIF	2 samples	38.3%	1.2%	0.91
School Functioning											
School 3. Keeping up with schoolwork	2-5y	Only item included – Rasch analysis not performed[j]							62.5%	2.6%	N/A
	6-13y	0/5	–	–	–	3 samples	5 samples (Sex)	No misfit	39.2%	5.0%	0.89
	14-17y	4/5	0.83 (0.76, 0.87)	0.44 (0.34, 0.60)	1.13 (1.08, 1.22)	1 sample	1 sample (Sex)	No misfit	26.5%	5.4%	0.89
Social Functioning											
Soc 2. Others not wanting to be friends	2-5y	4/5	0.11 (0.01, 0.36)	1.28 (0.18, 2.23)	1.14 (0.88, 1.34)	1 sample	No DIF	No misfit	46.8%	0.2%	0.75
	6-13y	4/5	0.24 (0.02, 0.90)	1.17 (0.42, 1.64)	1.25 (0.75, 1.84)	1 sample	1 sample (Sex)	No misfit	43.4%	0.9%	0.78
	14-17y	3/5	0.17 (0.09, 0.27)	0.65 (0.46, 1.0)	1.04 (0.76, 1.35)	1 sample	2 samples (Sex)	No misfit	51.4%	0.6%	0.80
Soc 3. Getting teased	2-5y	4/5	0.10 (0.01, 0.27)	1.05 (0.63, 2.15)	1.02 (0.81, 1.18)	1 sample	No DIF	No misfit	60.8%	0.1%	0.68
	6-13y	1/5	0.22[k]	0.31	0.92	3 samples	1 sample (Health status)	No misfit	42.4%	0.7%	0.76
	14-17y	2/5	0.57 (0.44, 0.70)	0.43 (0.01, 0.86)	0.87 (0.77, 0.97)	2 samples	1 sample (Sex)	No misfit	54.0%	0.6%	0.77

Item Description *Problems with…*	Age Group	Rasch Criteria							Psychometric Criteria[a,b]		
		Total Score[c]	Mean Item Level Performance[d] (Range Across Subsamples)			Disordered Thresholds	DIF	Item Misfit	% Response Ceiling (Never)	% Response Floor (Almost always)	Corr.[e]
			***p*-value**[f]	**Fit Residual**[g]	**Spread**						
Emotional Functioning											
Emot 1. Feeling afraid or scared	2-5y	4/5	0.16 (0.05, 0.40)	0.90 (0.30, 1.38)	1.10 (0.99, 1.28)	1 sample	No DIF	No misfit	23.0%	0.3%	0.72
	6-13y	2/5	0.27 (0.08, 0.45)	1.13 (0.86, 1.41)	1.09 (0.96, 1.22)	1 sample	2 samples (Sex)	No misfit	35.8%	0.5%	0.77
	14-17y	2/5	0.06 (0.005, 0.12)	1.23 (0.85, 1.62)	0.94 (0.76, 1.12)	1 sample	2 samples (Sex) 1 sample (Health status and Sex)	No misfit	53.0%	0.5%	0.78
Emot 3. Feeling angry	2-5y	4/5	0.51 (0.09, 0.84)	0.72 (0.26, 1.72)	1.57 (1.24, 1.97)	1 sample	1 sample (Sex)	No misfit	12.2%	0.2%	0.67
	6-13y	1/5	0.25[k]	1.13	0.97	None	3 samples (Sex) 1 sample (Health status and Sex)	4 samples	15.4%	0.7%	0.69
	14-17y	1/5	0.57[k]	1.50	1.08	None	3 samples (Sex)	1 sample	19.0%	1.0%	0.75
Emot 5. Worrying	2-5y	4/5	0.05 (0.004, 0.11)	1.26 (0.03, 1.81)	0.99 (0.76, 1.28)	None	No DIF	1 sample	50.9%	0.2%	0.72
	6-13y	5/5	0.25 (0.04, 0.51)	0.52 (0.38, 0.74)	0.98 (0.75, 1.32)	None	No DIF	No misfit	39.7%	0.9%	0.77
	14-17y	5/5	0.33 (0.01, 0.68)	1.20 (0.47, 1.69)	0.99 (0.81, 1.12)	None	No DIF	No misfit	36.6%	1.1%	0.81

Item Description *Problems with…*	Age Group	Rasch Criteria: Total Score[c]	Mean Item Level Performance[d] (Range Across Subsamples): *p*-value[f]	Fit Residual[g]	Spread	Disordered Thresholds	DIF	Item Misfit	Psychometric Criteria[a,b]: % Response Ceiling (Never)	% Response Floor (Almost always)	Corr.[e]
Physical Functioning[l]											
Phys 2. Running	2-5y	0/3[m]	–	–	–	2 samples	No DIF	3 samples	91.8%	0.3%	0.86[n]
	6-13y	4/5	0.08 (0.004, 0.13)	1.13 (1.09, 1.17)	0.68 (0.64, 0.72)	None	None	1 sample	76.0%	1.8%	0.93
	14-17y	3/3[m]	0.06 (0.02, 0.09)	0.70 (0.11, 1.10)	0.63 (0.56, 0.78)	None	No DIF	None	70.6%	2.3%	0.92
Phys 3. Participating in exercise	2-5y	0/3[m]	–	–	–	3 samples	No DIF	3 samples	80.5%	0.6%	0.83[o]
	6-13y	4/5	0.26 (0.06, 0.63)	0.42 (0.08, 1.16)	0.59 (0.54, 0.64)	1 sample	None	1 sample	73.1%	3.1%	0.91
	14-17y	3/3[m]	0.05 (0.02, 0.13)	0.37 (0.002, 0.57)	0.61 (0.50, 0.72)	None	No DIF	None	67.7%	3.5%	0.91

Abbreviations: SchAbs, School Absence; School, School Functioning; Soc, Social Functioning; Emot, Emotional Functioning; Phys, Physical Functioning; DIF, differential item functioning; Corr, correlation; y, years; N/A, not applicable.

[a] Psychometric criteria were assessed using the full LSAC sample (n=45,207) rather than using just the Rasch analysis subsamples.

[b] Results from psychometric analysis for all items shown in Appendix aTables 3.7-3.11.

[c] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).

[d] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.

[e] Correlation of item score with dimension score.

[f] *p*-value for the individual item χ^2 statistics.

[g] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.

[h] School Absence items were not asked for children aged 2-3 years in the LSAC, thus results reflect responses for children aged 4-5 years.

[I] Health status defined as children with special healthcare needs or typically functioning children.

[j] School 3 is only item included in the PedsQL for this dimension for this age group, thus Rasch analysis was not performed.

[k] Only 1 subsample performed well on all Rasch criteria, thus no range reported in table.

[l] Results from supplemental Rasch analyses shown in table.

[m] Insufficient sample size to obtain 5 subsamples so only 3 subsamples were created for supplemental analyses.

[n] Correlation of Phys 2 with Physical Functioning dimension if the dimension included all 6 items (Phys 1-Phys 6) was 0.62 for age group 2-5 years, 0.82 for age group 6-13 years, and 0.82 for age group 14-17 years.

[o] Correlation of Phys 3 with Physical Functioning dimension if the dimension included all 6 items (Phys 1-Phys 6) was 0.67 for age group 2-5 years, 0.86 for age group 6-13 years, and 0.87 for age group 14-17 years.

School Absence

Only 1 item (SchAbs 1) remained in the School Absence dimension based on the exclusion criteria defined in Step 1. Compared to SchAbs 2, SchAbs 1 better fit the Rasch model and had higher correlation with the dimension score (≥0.90) (**aTables 3.2 and 3.8**). Both SchAbs 1 and SchAbs 2 exhibited ceiling effects across age groups, although it was less severe for SchAbs 1. Lastly, all expert panelists and key informant interviewees agreed that SchAbs 1 was the best item to represent the School Absence dimension. Therefore, SchAbs 1 was selected for inclusion in the PedsUtil health state classification system.

School Functioning

The School Functioning dimension also had only 1 item remaining after Step 1 (School 3). On average, School 3 better fit the Rasch model than the other 2 items and had the greatest item spread (**aTable 3.3**). School 3 also had high correlation (0.89) with the dimension score (**aTable 3.9**). However, School 3 did exhibit ceiling effects for age groups 2-5 years (63%) and 6-13 years (39%). Nevertheless, all expert panelists and key informant interviewees agreed that School 3 was the best item to represent the School Functioning dimension, thus School 3 was included in the PedsUtil health state classification system.

Social Functioning

Items Soc 2 and Soc 3 remained in the Social Functioning dimension after Step 1. Overall, Soc 2 better fit the Rasch model, had larger item spread, less severe ceiling effects, and higher correlation with the dimension score across all age groups compared to Soc 3 (**aTables 3.4 and 3.10**). Moreover, none of the expert panelists and key informant interviewees thought that Soc 3 was the best item to represent the Social Functioning dimension. In contrast, 2 out of 6 expert panelists and 5 out of 12 key informant interviewees chose Soc 2 as the best item to represent the dimension. It should be noted that the remaining expert panelists and key informant interviewees selected Soc 1 as the best item to represent the dimension. But because Soc 1 exhibited disordered thresholds in all subsamples across all age groups, it was excluded from the health state classification system in Step 1. Soc 2 was ultimately selected to be included in the health state classification, and this decision was reviewed with the health status measurement expert who agreed that Soc 2 was the most suitable item to include and best fit with the overall tone of the PedsUtil health state classification system.

Emotional Functioning

Among the 3 remaining items (Emot 1, Emot 3, and Emot 5) for the Emotional Functioning dimension, Emot 3 was the worst performing item based on Rasch criteria with a total score of 1/5 for both age groups 6-13 years and 14-17 years (**aTable 3.5**). Since Emot 3 exhibited DIF and/or item misfit for most subsamples, it was not selected to be included in the health state classification system. Between Emot 1 and Emot 5, Emot 5 had higher total scores across all age groups and had slightly higher correlation with dimension score (**aTables 3.5 and 3.11**). However, Emot 1 had larger item spread for age groups 2-5 years and 6-13 years, while Emot 5 had larger item spread for age group 14-17 years. Ceiling effects for Emot 1 and Emot 5 depended on the age group, with Emot 5 exhibiting large ceiling effects (51%) for age group 2-4 years and Emot 1 exhibiting large ceiling effects (53%) for age group 14-17 years. Compared to Emot 1, Emot 5 was selected more often as the best item to represent the Emotional Functioning dimension by expert panelists and key informant interviewees. None of the expert panelists chose Emot 1 as the best item, while 1 panelist chose Emot 5. Four out of 12 key informant interviewees chose Emot 1, while 6 chose Emot 5. It is important to note that the majority of expert panelists (5 out of 6) chose Emot 2 as the best item to represent the dimension, though many were torn between items Emot 2 and Emot 5. This finding was inconsistent with the Rasch analysis results which indicated that Emot 2 was the worst performing item for age groups 2-5 years (total score 1/5) and 14-17 years (total score 0/5). The health status measurement expert reviewed all results and concluded that Emot 5 may better express the pathology of emotional functioning compared to Emot 1 and Emot 2. For example, it is normal and/or expected to experience some items in the Emotional Functioning dimension, such as Emot 2 (feeling sad or blue). In fact, experiencing some level of such emotions may actually demonstrate better emotional functioning than if a child never experiences them. Therefore, the health status measurement expert considered Emot 5 as the best item to illustrate symptoms of emotional functioning when using the 5 item response levels (i.e., Never, Almost Never, Sometimes, Often, and Almost Always) of the PedsQL. After careful consideration of all findings, the core research team selected Emot 5 for inclusion in the health state classification system.

Physical Functioning

Items Phys 2 and Phys 3 remained in the Physical Functioning dimension following Step 1. Both items similarly fit the Rasch model in the supplemental analyses, but Phys 3 had marginally better item fit to the Rasch model than Phys 2 (**aTable 3.7**). Phys 2 and Phys 3 had similar item spread, and both demonstrated large ceiling effects which were less severe for Phys 3 (**aTable 3.12**). Both items also had high correlations with dimension score across age groups. Though both items performed similarly on Rasch and psychometric criteria, most expert panelists (5 out of 6) and key informant interviewees (11 out of 12) thought Phys 3 was the best item to represent the Physical Functioning dimension. Therefore, Phys 3 was included in the health state classification system.

Final PedsUtil Health State Classification System

Figure 3.1 displays the final PedsUtil health state classification system. **Table 3.3** presents the correlations between the items selected to represent the dimensions of the PedsUtil health state classification system. As shown in the table, there was minimal correlation between the dimensions, with most correlations ≤0.37. The only exception was for dimensions Pain and Fatigue, which had a correlation of 0.46, but was still regarded as a weak correlation. The limited correlations between the dimensions of the PedsUtil health state classification system suggest that the dimensions are structurally independent.

Figure 3.1 PedsUtil Health State Classification System[a]

Physical Functioning – Participating in sports activity or exercise
Never has problems with participating in sports activity or exercise
Almost never has problems with participating in sports activity or exercise
Sometimes has problems with participating in sports activity or exercise
Often has problems with participating in sports activity or exercise
Almost always has problems with participating in sports activity or exercise

Pain – Having hurts or aches
Never has problems with having hurts or aches
Almost never has problems with having hurts or aches
Sometimes has problems with having hurts or aches
Often has problems with having hurts or aches
Almost always has problems with having hurts or aches

Fatigue – Low energy level
Never has problems with low energy level
Almost never has problems with low energy level
Sometimes has problems with low energy level
Often has problems with low energy level
Almost always has problems with low energy level

Emotional Functioning – Worrying about what will happen to them
Never has problems with worrying about what will happen to them
Almost never has problems with worrying about what will happen to them
Sometimes has problems with worrying about what will happen to them
Often has problems with worrying about what will happen to them
Almost always has problems with worrying about what will happen to them

Social Functioning – Other kids not wanting to be their friend
Never has problems with other kids not wanting to be their friend
Almost never has problems with other kids not wanting to be their friend
Sometimes has problems with other kids not wanting to be their friend
Often has problems with other kids not wanting to be their friend
Almost always has problems with other kids not wanting to be their friend

School Functioning – Keeping up with schoolwork
Never has problems with keeping up with schoolwork
Almost never has problems with keeping up with schoolwork
Sometimes has problems with keeping up with schoolwork
Often has problems with keeping up with schoolwork
Almost always has problems with keeping up with schoolwork

School Absence – Missing school because of not feeling well
Never has problems with missing school because of not feeling well
Almost never has problems with missing school because of not feeling well
Sometimes has problems with missing school because of not feeling well
Often has problems with missing school because of not feeling well
Almost always has problems with missing school because of not feeling well

[a] Wording for PedsUtil health state classification system differs slightly between age groups but items selected are the same across all age groups. PedsUtil health state classification system for age group 8-12 years shown in this table.

Table 3.3 Correlations Between Dimensions

Dimension	Physical Functioning	Pain	Fatigue	Emotional Functioning	Social Functioning	School Functioning	School Absence
Physical Functioning	1.00	–	–	–	–	–	–
Pain	0.24	1.00	–	–	–	–	–
Fatigue	0.31	0.46	1.00	–	–	–	–
Emotional Functioning	0.20	0.32	0.34	1.00	–	–	–
Social Functioning	0.21	0.25	0.26	0.37	1.00	–	–
School Functioning	0.35	0.19	0.25	0.26	0.28	1.00	–
School Absence	0.23	0.33	0.32	0.25	0.19	0.28	1.00

Discussion

This study conducted extensive Rasch and psychometric analyses to reduce the full set of PedsQL items to a core set in order to derive the PedsUtil health state classification system. Child health experts and parents were also involved in the item selection process to ensure content and face validity of the health state classification system. The PedsUtil health state classification system was constructed to be suitable for use with children 2 years and older, including both children with special healthcare needs and typically functioning children. This is the first study to derive a health state classification system based on the PedsQL.

There are some limitations to this study. Similar to the previous chapter, this study used data from an Australian population. However, the PedsUtil scoring system will be developed using a US-representative population sample, thus analyzing data from a US sample would have been preferred when selecting items for the health state classification system. Nevertheless, incorporating input from US-based experts and parents in the item selection process helped establish that the PedsUtil health state classification system is acceptable to use with a US population. A second potential limitation is that the small, purposive sample of expert panelists and key informant interviewees used in this study may not reflect all viewpoints of the broader general population. Although participants were selected so that children of different ages and different health statuses were widely represented, the resulting sample included mostly females and highly educated individuals. Researchers conducting similar studies should consider further diversifying their sample to investigate potential variations in opinions among different groups. A third potential limitation is that this study used parent-proxy responses to the PedsQL. Future research should validate item selection with child self-report responses. A fourth limitation is that the age groups used in this analysis (i.e., 2-5 years, 6-13 years, and 14-17 years) are slightly different from the age groups of the PedsQL (i.e., 2-4 years, 5-7 years, 8-12 years, and 13-18 years). The PedsQL age groups could not be adopted in this analysis because the LSAC collects data every 2 years, requiring age groups to be grouped into 2-year intervals. For example, 2-year-olds and 3-year-olds were grouped together, 4-year-olds and 5-year-olds, and so forth. Therefore, age groups for this analysis were constructed to closely match the age groups of the PedsQL, though were not identical given the study design of the LSAC. It should also be noted that the 6-year-olds and 7-year-olds were not constructed to be a separate age group, but instead were combined with children aged 8-13 years. This was because children aged 6-7 years only represent a single wave of data collection for each LSAC cohort. Prior work also suggests that combining children aged 6-7 years with children aged 8-13 years may be appropriate for this analysis as those years represent the middle childhood years.[140] Given that the objective of the study was to identify a single best item to represent each dimension *across* all ages, the slight variation in age groups most likely does not have a significant effect on study findings. And lastly, this study did not explore collapsing item response levels when deriving the final PedsUtil health state classification system. Other studies have reduced the number of item response levels because some respondents may find it difficult to distinguish between levels in preference valuation exercises.[109,124] However, this study did not choose to reduce the number of item response levels for two reasons. First, reducing the number of levels after selecting items may contradict Rasch criteria (i.e., item level ordering and spread of item thresholds) used earlier in the item selection process to evaluate item performance. Second, collapsing item response levels changes the original structure of the PedsQL, which may result in respondents valuing items with collapsed levels differently than if the original item response levels were preserved in the health state classification system. Since the aim of the study was to derive a health state classification system that most closely resembles the original PedsQL instrument, item response levels were not reduced.

In summary, this study identified the most representative item for each dimension to include in the PedsUtil health state classification system. The next step is to develop and apply a valuation protocol for the preliminary construction of the PedsUtil scoring system.

Appendix

aTable 3.1 Summary of Parent Key Informant Interviewee Characteristics (*n* = 12)

Characteristic	***n***	**%**
Sex[a]		
Male	1	8%
Female	11	92%
Highest level of education		
High school/GED	0	0%
2-year college/Associate's degree	0	0%
4-year college degree	1	8%
Advanced degree (Master's, Doctorate or Professional)	7	58%
Unknown	4	33%
Number of children		
1	3	25%
2	8	67%
3 or more	1	8%
Age of child(ren)[b]		
2-5 years	6	–[c]
6-13 years	6	–
14-17 years	3	–
18 years or older	1	–
Child health status[d]		
Typically functioning child	8	–[c]
Child with special healthcare needs	8	–

[a] Sex of the parent.
[b] Some parents had children in different age groups so total adds up to more than *n* = 12.
[c] % not calculated because total equals to more than *n* = 12 since parents can be in multiple categories.
[d] Some parents had both typically functioning children and children with special healthcare needs so total adds up to more than *n* = 12.

aFigure 3.1 Questions Provided to Expert Panelists and Key Informant Interviewees

Question 1: School Absence

Do you think this is the best question to represent the **school absence** category for children?

School Absence Questions

Problems with…

2-5 years	1. Missing school/daycare because of not feeling well
	2. Missing school/daycare to go to the doctor or hospital

6-13 years AND 14-17 years	1. Missing school because of not feeling well
	2. Missing school to go to the doctor or hospital

Best Question

Problems with…

1. Missing school/daycare because of not feeling well

Question 2: School Functioning

Do you think this is the best question to represent the **school functioning** category for children?

School Functioning Questions

Problems with…

2-5 years	1. Doing the same school activities as peers[1]

[1]Only 1 question in this category for this age group

6-13 years	1. Keeping up with school activities/school work[2]
	2. Paying attention in class
	3. Forgetting things

[2]Children up to 7 years old asked about 'school activities' and children 8 years and older asked about 'school work'

14-17 years	1. Keeping up with school work
	2. Paying attention in class
	3. Forgetting things

Best Question

Problems with…

1. Doing the same school activities as peers/Keeping up with school activities/school work

Question 3: Social Functioning

Do you think this is the best question to represent the **social functioning** category for children?

Social Functioning Questions
Problems with…

2-5 years	1. Playing with other children
	2. Other kids not wanting to play with him/her
	3. Getting teased by other children
	4. Not able to do things that other children his/her age can do
	5. Keeping up when playing with other children
6-13 years	1. Getting along with other children
	2. Other kids not wanting to be his/her friend
	3. Getting teased by other children
	4. Not able to do things that other children his/her age can do
	5. Keeping up when playing with other children
14-17 years	1. Getting along with other teens
	2. Other teens not wanting to be his/her friend
	3. Getting teased by other teens
	4. Not able to do things that other teens his/her age can do
	5. Keeping up with other teens

Best Question
Problems with…

2. Other kids not wanting to play with him or her/Other kids/teens not wanting to be his or her friend

Question 4: Emotional Functioning

Which of these 2 questions do you think is the best question to represent the **emotional functioning** category for children?

Emotional Functioning Questions
Problems with…

2-5 years	1. Feeling afraid or scared
	2. Feeling sad or blue
	3. Feeling angry
	4. Trouble sleeping
	5. Worrying
6-13 years AND 14-17 years	1. Feeling afraid or scared
	2. Feeling sad or blue
	3. Feeling angry
	4. Trouble sleeping
	5. Worrying about what will happen to him/her

Candidates for Best Question
Problems with…

1.Feeling afraid or scared
5. Worrying/Worrying about what will happen to him/her

Best Question
Problems with…

???

Question 5: Physical Functioning

Which of these 2 questions do you think is the best question to represent the **physical functioning** category for children?

Physical Functioning Questions
Problems with…

2-5 years	1. Walking
	2. Running
	3. Participating in active play or exercise
	4. Lifting something heavy
	5. Bathing
	6. Helping pick up his/her toys

6-13 years AND 14-17 years	1. Walking more than one block
	2. Running
	3. Participating in sports activity or exercise
	4. Lifting something heavy
	5. Taking a bath or shower by himself/herself
	6. Doing chores around the house (like picking up his/her toys)

Candidates for Best Question
Problems with…

2. Running
3. Participating in active play or exercise/Participating in sports activity or exercise

Best Question
Problems with…

???

aTable 3.2 Summary of Rasch Analysis Results - School Absence[a]

Item Description *Problems with…*	**Total Score**[b]	**Mean Item Level Performance**[c] (Range Across Subsamples)			**Disordered Thresholds**	**DIF**	**Item Misfit**
		***p*-value**[d]	**Fit Residual**[e]	**Spread**			
2-5 years[f]							
SchAbs 1. Missing school because sick	2/5	0.15 (0.10, 0.19)	0.37 (0.17, 0.57)	2.08 (1.86, 2.30)	1 sample	2 samples (Health status[g])	No misfit
SchAbs 2. Missing school to go to doctor	0/5	–	–	–	1 sample	2 samples (Health status)	5 samples
6-13 years							
SchAbs 1. Missing school because sick	4/5	0.06 (0.01, 0.15)	0.53 (0.33, 0.73)	1.67 (1.10, 2.03)	None	No DIF	1 sample
SchAbs 2. Missing school to go to doctor	0/5	–	–	–	1 sample	2 samples (Health status)	5 samples
14-17 years							
SchAbs 1. Missing school because sick	3/5	0.04 (0.02, 0.06)	0.09 (0.02, 0.16)	1.22 (1.14, 1.32)	None	No DIF	2 samples
SchAbs 2. Missing school to go to doctor	0/5	–	–	–	None	1 sample (Health status)	5 samples

Abbreviations: SchAbs, School Absence; DIF, differential item functioning.
[a] All subsamples across all age groups did not fit the Rasch model, thus the item-specific results reported in the table are for Rasch models with statistically significant item-trait interaction χ^2 statistics.
[b] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[c] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[d] *p*-value for the individual item χ^2 statistics.
[e] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[f] School Absence dimension was not asked for children aged 2-3 years in the LSAC, thus results reflect responses for children aged 4-5 years.
[g] Health status defined as children with special healthcare needs or typically functioning children.

aTable 3.3 Summary of Rasch Analysis Results - School Functioning

Item Description *Problems with…*	**Total Score**[a]	**Mean Item Level Performance**[b] (Range Across Subsamples)			**Disordered Thresholds**	**DIF**	**Item Misfit**
		***p*-value**[c]	**Fit Residual**[d]	**Spread**			
2-5 years							
School 3. Keeping up with schoolwork	Only item included – Rasch analysis not performed[e]						
6-13 years							
School 1. Paying attention in class	0/5	–	–	–	2 samples	5 samples (Sex)	No misfit
School 2. Forgetting things	0/5	–	–	–	None	No DIF	5 samples
School 3. Keeping up with schoolwork	0/5	–	–	–	3 samples	5 samples (Sex)	No misfit
14-17 years							
School 1. Paying attention in class	2/5	0.08 (0.06, 0.10)	0.45 (0.43, 0.47)	1.10 (1.06, 1.15)	None	3 samples (Sex)	No misfit
School 2. Forgetting things	0/5	–	–	–	None	1 sample (Health status[f] and Sex)	5 samples
School 3. Keeping up with schoolwork	4/5	0.83 (0.76, 0.87)	0.44 (0.34, 0.60)	1.13 (1.08, 1.22)	1 sample	1 sample (Sex)	No misfit

Abbreviations: DIF, differential item functioning.
[a] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[b] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[c] *p*-value for the individual item χ^2 statistics.
[d] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[e] Only School 3 is included in the PedsQL for this dimension for this age group, thus Rasch analysis was not performed.
[f] Health status defined as children with special healthcare needs or typically functioning children.

aTable 3.4 Summary of Rasch Analysis Results - Social Functioning

Item Description *Problems with…*	Total Score[a]	Mean Item Level Performance[b] (Range Across Subsamples)			Disordered Thresholds	DIF	Item Misfit
		***p*-value**[c]	**Fit Residual**[d]	**Spread**			
2-5 years							
Soc 1. Getting along with others	0/5	–	–	–	5 samples	No DIF	No misfit
Soc 2. Others not wanting to be friends	4/5	0.11 (0.01, 0.36)	1.28 (0.18, 2.23)	1.14 (0.88, 1.34)	1 sample	No DIF	No misfit
Soc 3. Getting teased	4/5	0.10 (0.01, 0.27)	1.05 (0.63, 2.15)	1.02 (0.81, 1.18)	1 sample	No DIF	No misfit
Soc 4. Unable to do things others can do	1/5	0.14[e]	0.84	0.52	4 samples	1 sample (Health status[f]) 1 sample (Sex)	No misfit
Soc 5. Keeping up with other children	0/5	–	–	–	5 samples	No DIF	No misfit
6-13 years							
Soc 1. Getting along with others	0/5	–	–	–	5 samples	1 sample (Health status) 1 sample (Sex)	1 sample
Soc 2. Others not wanting to be friends	4/5	0.24 (0.02, 0.90)	1.17 (0.42, 1.64)	1.25 (0.75, 1.84)	1 sample	1 sample (Sex)	No misfit
Soc 3. Getting teased	1/5	0.22[e]	0.31	0.92	3 samples	1 sample (Health status)	No misfit
Soc 4. Unable to do things others can do	0/5	–	–	–	5 samples	No DIF	2 samples
Soc 5. Keeping up with other children	0/5	–	–	–	5 samples	1 sample (Health status)	2 samples
14-17 years							
Soc 1. Getting along with others	0/5	–	–	–	5 samples	1 sample (Health status)	No misfit
Soc 2. Others not wanting to be friends	3/5	0.17 (0.09, 0.27)	0.65 (0.46, 1.0)	1.04 (0.76, 1.35)	1 sample	2 samples (Sex)	No misfit
Soc 3. Getting teased	2/5	0.57 (0.44, 0.70)	0.43 (0.01, 0.86)	0.87 (0.77, 0.97)	2 samples	1 sample (Sex)	No misfit
Soc 4. Unable to do things others can do	0/5	–	–	–	4 samples	1 sample (Health status) 1 sample (Sex)	No misfit
Soc 5. Keeping up with other children	0/5	–	–	–	5 samples	No DIF	No misfit

Abbreviations: Soc, Social Functioning; DIF, differential item functioning.

[a] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[b] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[c] *p*-value for the individual item χ^2 statistics.
[d] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[e] Only 1 subsample performed well on all Rasch criteria, thus no range reported in table.
[f] Health status defined as children with special healthcare needs or typically functioning children.

aTable 3.5 Summary of Rasch Analysis Results - Emotional Functioning

Item Description *Problems with…*	**Total Score**[a]	**Mean Item Level Performance**[b] (Range Across Subsamples)			**Disordered Thresholds**	**DIF**	**Item Misfit**
		***p*-value**[c]	**Fit Residual**[d]	**Spread**			
2-5 years							
Emot 1. Feeling afraid or scared	4/5	0.16 (0.05, 0.40)	0.90 (0.30, 1.38)	1.10 (0.99, 1.28)	1 sample	No DIF	No misfit
Emot 2. Feeling sad or blue	1/5	0.36[e]	0.50	1.11	4 samples	No DIF	No misfit
Emot 3. Feeling angry	4/5	0.51 (0.09, 0.84)	0.72 (0.26, 1.72)	1.57 (1.24, 1.97)	1 sample	1 sample (Sex)	No misfit
Emot 4. Trouble sleeping	2/5	0.36 (0.34, 0.37)	0.37 (0.25, 0.49)	0.47 (0.43, 0.50)	1 sample	No DIF	2 samples
Emot 5. Worrying	4/5	0.05 (0.004, 0.11)	1.26 (0.03, 1.81)	0.99 (0.76, 1.28)	None	No DIF	1 sample
6-13 years							
Emot 1. Feeling afraid or scared	2/5	0.27 (0.08, 0.45)	1.13 (0.86, 1.41)	1.09 (0.96, 1.22)	1 sample	2 samples (Sex)	No misfit
Emot 2. Feeling sad or blue	5/5	0.13 (0.007, 0.43)	0.71 (0.25, 1.44)	1.20 (0.89, 1.58)	None	No DIF	No misfit
Emot 3. Feeling angry	1/5	0.25[e]	1.13	0.97	None	3 samples (Sex) 1 sample (Health status[f] and Sex)	4 samples
Emot 4. Trouble sleeping	0/5	–	–	–	2 samples	No DIF	4 samples
Emot 5. Worrying	5/5	0.25 (0.04, 0.51)	0.52 (0.38, 0.74)	0.98 (0.75, 1.32)	None	No DIF	No misfit
14-17 years							
Emot 1. Feeling afraid or scared	2/5	0.06 (0.005, 0.12)	1.23 (0.85, 1.62)	0.94 (0.76, 1.12)	1 sample	2 samples (Sex) 1 sample (Health status and Sex)	No misfit
Emot 2. Feeling sad or blue	0/5	–	–	–	None	4 samples (Sex) 1 sample (Health status and Sex)	2 samples
Emot 3. Feeling angry	1/5	0.57[e]	1.50	1.08	None	3 samples (Sex)	1 sample
Emot 4. Trouble sleeping	2/5	0.30 (0.17, 0.43)	2.09 (1.92, 2.26)	0.62 (0.57, 0.67)	None	No DIF	3 samples
Emot 5. Worrying	5/5	0.33 (0.01, 0.68)	1.20 (0.47, 1.69)	0.99 (0.81, 1.12)	None	No DIF	No misfit

Abbreviations: Emot, Emotional Functioning; DIF, differential item functioning.

[a] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[b] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[c] *p*-value for the individual item χ^2 statistics.
[d] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[e] Only 1 subsample performed well on all Rasch criteria, thus no range reported in table.
[f] Health status defined as children with special healthcare needs or typically functioning children.

aTable 3.6 Summary of Rasch Analysis Results - Physical Functioning

Item Description *Problems with…*	**Total Score**[a]	**Mean Item Level Performance**[b] (Range Across Subsamples)			**Disordered Thresholds**	**DIF**	**Item Misfit**
		***p*-value**[c]	**Fit Residual**[d]	**Spread**			
2-5 years[e]							
Phys 1. Walking	0/5	–	–	–	5 samples	1 sample (Health status[f])	4 samples
Phys 2. Running	0/5	–	–	–	5 samples	No DIF	3 samples
Phys 3. Participating in exercise	1/5	0.01[g]	2.12	0.46	4 samples	No DIF	3 samples
Phys 4. Lifting something heavy	0/5	–	–	–	4 samples	4 samples (Sex)	No misfit
Phys 5. Taking a bath or shower	0/5	–	–	–	5 samples	No DIF	4 samples
Phys 6. Doing chores	0/5	–	–	–	2 samples	No DIF	4 samples
6-13 years[e]							
Phys 1. Walking	0/5	–	–	–	5 samples	No DIF	4 samples
Phys 2. Running	0/5	–	–	–	2 samples	1 sample (Sex)	4 samples
Phys 3. Participating in exercise	1/5	0.22[g]	0.17	0.67	1 sample	No DIF	4 samples
Phys 4. Lifting something heavy	0/5	–	–	–	5 samples	No DIF	5 samples
Phys 5. Taking a bath or shower	0/5	–	–	–	5 samples	2 samples (Health status)	2 samples
Phys 6. Doing chores	0/5	–	–	–	1 sample	1 sample (Sex) 1 sample (Health status)	5 samples
14-17 years							
Phys 1. Walking	0/5	–	–	–	5 samples	1 sample (Sex)	3 samples
Phys 2. Running	3/5	0.36 (0.05, 0.58)	0.61 (0.26, 1.11)	0.56 (0.49, 0.67)	None	1 sample (Sex) 1 sample (Health status)	No misfit
Phys 3. Participating in exercise	2/5	0.13 (0.003, 0.25)	2.04 (1.74, 2.35)	0.65 (0.60, 0.70)	None	1 sample (Sex) 2 samples (Health status)	1 sample
Phys 4. Lifting something heavy	3/5	0.51 (0.34, 0.63)	1.34 (0.50, 1.93)	0.64 (0.62, 0.65)	1 sample	No DIF	2 samples
Phys 5. Taking a bath or shower	0/5	–	–	–	5 samples	1 sample (Sex) 1 sample (Health status) 1 sample (Health status	No misfit

						and Sex)	
Phys 6. Doing chores	0/5	–	–	–	None	4 samples (Sex) 1 sample (Health status)	5 samples

Abbreviations: Phys, Physical Functioning; DIF, differential item functioning.
[a] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[b] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[c] *p*-value for the individual item χ^2 statistics.
[d] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[e] 4 subsamples did not fit the Rasch model.
[f] Health status defined as children with special healthcare needs or typically functioning children.
[g] Only 1 subsample performed well on all Rasch criteria, thus no range reported in table.

aTable 3.7 Summary of Rasch Analysis Results - Physical Functioning Supplement[a]

Item Description *Problems with…*	**Total Score**[b]	**Mean Item Level Performance**[c] (Range Across Subsamples)			**Disordered Thresholds**	**DIF**	**Item Misfit**
		***p*-value**[d]	**Fit Residual**[e]	**Spread**			
2-5 years[f]							
Phys 1. Walking	0/3[g]	–	–	–	1 sample	No DIF	3 samples
Phys 2. Running	0/3	–	–	–	2 samples	No DIF	3 samples
Phys 3. Participating in exercise	0/3	–	–	–	3 samples	No DIF	3 samples
6-13 years[h]							
Phys 1. Walking	1/5	0.02[i]	0.94	0.47	None	No DIF	3 samples
Phys 2. Running	4/5	0.08 (0.004, 0.13)	1.13 (1.09, 1.17)	0.68 (0.64, 0.72)	None	None	1 sample
Phys 3. Participating in exercise	4/5	0.26 (0.06, 0.63)	0.42 (0.08, 1.16)	0.59 (0.54, 0.64)	1 sample	None	1 sample
14-17 years							
Phys 1. Walking	1/3[g]	0.01[i]	0.20	0.22	1 sample	No DIF	2 samples
Phys 2. Running	3/3	0.06 (0.02, 0.09)	0.70 (0.11, 1.10)	0.63 (0.56, 0.78)	None	No DIF	None
Phys 3. Participating in exercise	3/3	0.05 (0.02, 0.13)	0.37 (0.002, 0.57)	0.61 (0.50, 0.72)	None	No DIF	None

Abbreviations: Phys, Physical Functioning; DIF, differential item functioning.
[a] Supplemental analyses excluded items Phys 4-Phys 6 as they were considered not as relevant for this dimension.
[b] Total score = number of subsamples item performed well on all Rasch criteria (out of 5 subsamples).
[c] Mean item level performance calculated using only the subsamples that the item performed well on all Rasch criteria.
[d] *p*-value for the individual item χ^2 statistics.
[e] Fit residuals may be positive or negative, thus absolute value of the fit residuals reported in the table.
[f] None of the subsamples fit the Rasch model.
[g] Insufficient sample size to obtain 5 subsamples so only 3 subsamples were created for supplemental analyses.
[h] 1 out of the 5 subsamples did not fit the Rasch model.
[I] Only 1 subsample performed well on all Rasch criteria, thus no range reported in table.

aTable 3.8 Summary of Psychometric Analysis Results - School Absence

Item Description *Problems with…*	**% Response Ceiling (Never)**	**% Response Floor (Almost always)**	**Correlation**[a]
2-5 years[b]			
SchAbs 1. Missing school because sick	58.2%	0.1%	0.91
SchAbs 2. Missing school to go to doctor	74.3%	0.1%	0.88
6-13 years			
SchAbs 1. Missing school because sick	47.0%	0.3%	0.90
SchAbs 2. Missing school to go to doctor	64.9%	0.3%	0.88
14-17 years			
SchAbs 1. Missing school because sick	38.3%	1.2%	0.91
SchAbs 2. Missing school to go to doctor	56.3%	0.5%	0.88

Abbreviations: SchAbs, School Absence.
[a] Correlation of item score with dimension score.
[b] School Absence items were not asked for children aged 2-3 years in the LSAC, thus results reflect responses for children aged 4-5 years.

aTable 3.9 Summary of Psychometric Analysis Results - School Functioning

Item Description *Problems with…*	**% Response Ceiling (Never)**	**% Response Floor (Almost always)**	**Correlation**[a]
2-5 years			
School 3. Keeping up with schoolwork	62.5%	2.6%	N/A[b]
6-13 years			
School 1. Paying attention in class	32.7%	4.6%	0.90
School 2. Forgetting things	21.3%	1.5%	0.78
School 3. Keeping up with schoolwork	39.2%	5.0%	0.89
14-17 years			
School 1. Paying attention in class	32.5%	4.3%	0.91
School 2. Forgetting things	24.4%	1.2%	0.79
School 3. Keeping up with schoolwork	26.5%	5.4%	0.89

Abbreviations: SchAbs, School Absence; N/A, not applicable.
[a] Correlation of item score with dimension score.
[b] Only School 3 is included in the PedsQL for this dimension for this age group, thus correlation was not applicable.

aTable 3.10 Summary of Psychometric Analysis Results - Social Functioning

Item Description *Problems with…*	**% Response Ceiling (Never)**	**% Response Floor (Almost always)**	**Correlation**[a]
2-5 years			
Soc 1. Getting along with others	48.8%	1.6%	0.74
Soc 2. Others not wanting to be friends	46.8%	0.2%	0.75
Soc 3. Getting teased	60.8%	0.1%	0.68
Soc 4. Unable to do things others can do	63.2%	0.5%	0.70
Soc 5. Keeping up with other children	66.0%	2.4%	0.73
6-13 years			
Soc 1. Getting along with others	39.5%	4.7%	0.79
Soc 2. Others not wanting to be friends	43.4%	0.9%	0.78
Soc 3. Getting teased	42.4%	0.7%	0.76
Soc 4. Unable to do things others can do	59.3%	0.8%	0.69
Soc 5. Keeping up with other children	67.3%	4.0%	0.72
14-17 years			
Soc 1. Getting along with others	40.3%	4.7%	0.79
Soc 2. Others not wanting to be friends	51.4%	0.6%	0.80
Soc 3. Getting teased	54.0%	0.6%	0.77
Soc 4. Unable to do things others can do	61.1%	1.3%	0.75
Soc 5. Keeping up with other children	61.0%	3.2%	0.78

Abbreviations: Soc, Social Functioning.
[a] Correlation of item score with dimension score.

aTable 3.11 Summary of Psychometric Analysis Results - Emotional Functioning

Item Description *Problems with…*	**% Response Ceiling (Never)**	**% Response Floor (Almost always)**	**Correlation**[a]
2-5 years			
Emot 1. Feeling afraid or scared	23.0%	0.3%	0.72
Emot 2. Feeling sad or blue	31.0%	0.1%	0.71
Emot 3. Feeling angry	12.2%	0.2%	0.67
Emot 4. Trouble sleeping	36.7%	1.9%	0.65
Emot 5. Worrying	50.9%	0.2%	0.72
6-13 years			
Emot 1. Feeling afraid or scared	35.8%	0.5%	0.77
Emot 2. Feeling sad or blue	31.2%	0.3%	0.77
Emot 3. Feeling angry	15.4%	0.7%	0.69
Emot 4. Trouble sleeping	42.3%	1.9%	0.71
Emot 5. Worrying	39.7%	0.9%	0.77
14-17 years			
Emot 1. Feeling afraid or scared	53.0%	0.5%	0.78
Emot 2. Feeling sad or blue	31.4%	0.7%	0.84
Emot 3. Feeling angry	19.0%	1.0%	0.75
Emot 4. Trouble sleeping	34.5%	3.1%	0.77
Emot 5. Worrying	36.6%	1.1%	0.81

Abbreviations: Emot, Emotional Functioning.
[a] Correlation of item score with dimension score.

aTable 3.12 Summary of Psychometric Analysis Results - Physical Functioning

Item Description *Problems with…*	**% Response Ceiling (Never)**	**% Response Floor (Almost always)**	**Correlation**[a]
2-5 years			
Phys 1. Walking	93.1%	0.2%	0.58[b]
Phys 2. Running	91.8%	0.3%	0.62[c]
Phys 3. Participating in exercise	80.5%	0.6%	0.67[d]
Phys 4. Lifting something heavy	54.6%	0.5%	0.64
Phys 5. Taking a bath or shower	77.5%	0.8%	0.64
Phys 6. Doing chores	22.1%	3.8%	0.64
6-13 years			
Phys 1. Walking	83.1%	1.5%	0.81[b]
Phys 2. Running	76.0%	1.8%	0.82[c]
Phys 3. Participating in exercise	73.1%	3.1%	0.86[d]
Phys 4. Lifting something heavy	63.2%	1.0%	0.71
Phys 5. Taking a bath or shower	78.9%	5.1%	0.78
Phys 6. Doing chores	36.5%	4.3%	0.65
14-17 years			
Phys 1. Walking	82.3%	2.5%	0.84[b]
Phys 2. Running	70.6%	2.3%	0.82[c]
Phys 3. Participating in exercise	67.7%	3.5%	0.87[d]
Phys 4. Lifting something heavy	69.1%	1.5%	0.80
Phys 5. Taking a bath or shower	87.5%	8.2%	0.79
Phys 6. Doing chores	46.7%	3.9%	0.67

Abbreviations: Phys, Physical Functioning.
[a] Correlation of item score with dimension score.
[b] Correlation of Phys 1 with Physical Functioning dimension if the dimension included only 3 items (Phys 1-Phys 3) was 0.81 for age group 2-5 years, 0.88 for age group 6-13 years, and 0.86 for age group 14-17 years.
[c] Correlation of Phys 2 with Physical Functioning dimension if the dimension included only 3 items (Phys 1-Phys 3) was 0.86 for age group 2-5 years, 0.93 for age group 6-13 years, and 0.92 for age group 14-17 years.
[d] Correlation of Phys 3 with Physical Functioning dimension if the dimension included only 3 items (Phys 1-Phys 3) was 0.83 for age group 2-5 years, 0.91 for age group 6-13 years, and 0.91 for age group 14-17 years.

Chapter 4: Developing a Valuation Protocol to Construct the PedsUtil Scoring System Using a Discrete Choice Experiment with Time Trade-Off

Introduction

As mentioned previously, there are two components to preference-based measures of HRQoL that can be used to generate QALYs. The first is a health state classification system which describes a comprehensive set of health states, and the second is a scoring system that estimates health utility weights for every health state defined by the health state classification system. Chapters 2 and 3 described the first part of deriving a preference-based measure of HRQoL for children, which was the development of the PedsUtil health state classification system. This chapter describes the second part, which is the construction of the PedsUtil scoring system.

Developing the PedsUtil scoring system requires completing a valuation exercise in which respondents are asked to value a sample of health states defined by the PedsUtil health state classification system. Traditionally, standard cardinal methods have been used to directly value health states. The most commonly used cardinal techniques include SG and TTO. The SG approach is based on expected utility theory.[141] It gives respondents a choice between a certain intermediate health state and a gamble consisting of two possible health states, one of which is better than the intermediate state (usually perfect health), and one of which is worse (usually dead). The probability associated with the gamble is varied until the respondent is indifferent between choosing the certain intermediate health state and the gamble.[2,142] Unlike the SG approach, the TTO method asks respondents to choose between two choices of certainty.[34] In the conventional TTO method, respondents are presented with the task of determining what amount of time they are willing to trade in order to be in a better versus poorer health state. The respondent has the choice of being in a worse health state for a longer period of time followed by death, or being in a better health state (usually perfect health) for a shorter period of time followed by death. The time spent in the better health state is varied until the respondent is indifferent between the alternatives.[2,142] SG and TTO, however, have been criticized for being too complex for many respondents to complete. Furthermore, extraneous factors, such as risk aversion and time preference, may bias the elicitation of SG and TTO values.[143-145] Additional issues may arise with these elicitation techniques when valuing health states for children, especially since adults are oftentimes asked to value the health of a child (i.e., proxy valuations). For example, there is evidence to suggest that adults are less willing to trade-off life years when completing TTO tasks to value child health.[146,147] Consequently, there has been increasing interest to use ordinal methods, such as DCEs, to estimate health utility values.[148,149]

DCEs are based on random utility theory.[150] In the context of child health valuations, respondents are usually asked to state their preference between two child health states defined by a health state classification system. This approach assumes that respondents act in a manner that maximizes their utility and make choices based on the dimension levels of the health states.[151] A promising feature of this valuation technique is that it reduces the exposure of respondents to questions about death in children. However, a limitation is that the modeled discrete choice values are on a latent utility scale with arbitrary anchors. Therefore, the estimated preference weights need to be rescaled onto the full 0-1 QALY scale in order to meaningfully incorporate the DCE values into economic evaluations. Existing studies have used various methods to produce data for anchoring, including variations of DCE + duration,[152-154] TTO,[91-99] visual analog scale (VAS),[69,155] and the location-of-dead approach,[69] though it is important to note that each of these anchoring methods also has its own set of limitations.

The aim of this chapter was to develop and apply a valuation protocol to value child health states defined by PedsUtil health state classification system in order to generate preference weights for the preliminary construction of the PedsUtil scoring system. A pilot DCE was administered to value child health states defined by the PedsUtil health state classification system. In the absence of a preferred anchoring method, TTO values for select health states were collected online to rescale the DCE values. The PedsUtil scoring system was constructed to value HRQoL for 4 different pediatric age groups (i.e., 2-4 years, 5-7 years, 8-12 years, 13-18 years).

Methods

Overview of Approach

The valuation protocol included 2 online surveys: a DCE and a TTO survey. The surveys were administered to adult respondents consistent with recommendations from the US Panel on Cost-Effectiveness in Health and Medicine.[1] The valuation protocol was designed to estimate age-specific preference weights for 4 different pediatric age groups (i.e., 2-4 years, 5-7 years, 8-12 years, 13-18 years) in order to accurately value HRQoL for children of different developmental stages. These age groups also reflect the validated age groups of the PedsQL.[29] Survey respondents were randomized to value child health states for 2 out of the 4 pediatric age groups. This study adapted previously published methods that have estimated scoring systems for HRQoL measures[31,53,73] and followed best practices consistent with published guidelines for designing and implementing conjoint analysis research.[151,156,157] This study was granted an exempt determination by the University of Michigan Institutional Review Board (IRBMED # HUM00182088).

Valuation Protocol – DCE and TTO

DCE Survey Content and Development

Each DCE survey was comprised of the following elements (in order): study overview and IRB information, screening questions, instructions, 2 practice DCE tasks, first set of DCE tasks for first age group, 1 debrief question, instructions, second set of DCE tasks for second age group, 1 debrief question, and additional background and sociodemographic questions. The DCE tasks involved respondents choosing between “Health Description A” and “Health Description B” (**Figure 4.1**). A health description was formed by selecting 1 level in each dimension of the PedsUtil health state classification system (**aFigure 4.1**). The median age for each pediatric age group valued was used in the health state descriptions for the DCE tasks (i.e., 3-year-old, 6-year-old, 10-year-old, and 15-year-old).

Figure 4.1 Example DCE Question[a]

Please imagine a **10-YEAR-OLD CHILD** who can live with Health Description A or Health Description B.

Please select the health description that you think a **10-YEAR-OLD CHILD** would prefer to live with.

Please answer the question from the **point of view of a 10-YEAR-OLD CHILD**.

The levels for the dimensions of health are listed below for your reference:

Never — Almost Never — Sometimes — Often — Almost Always

Note: Some parts of the health descriptions are the same between Health Descriptions A and B and some parts are different. The parts that are different are highlighted in magenta below.

Dimension	Health Description A	Health Description B
Physical Functioning	**Never** has problems with participating in sports activity or exercise	**Never** has problems with participating in sports activity or exercise
Pain	**Sometimes** has problems with having hurts or aches	**Often** has problems with having hurts or aches
Fatigue	**Almost always** has problems with low energy level	**Often** has problems with low energy level
Emotional Functioning	**Almost never** has problems with worrying about what will happen to them	**Often** has problems with worrying about what will happen to them
Social Functioning	**Never** has problems with other kids not wanting to be their friend	**Never** has problems with other kids not wanting to be their friend
School Functioning	**Often** has problems with keeping up with schoolwork	**Sometimes** has problems with keeping up with schoolwork
School Absence	**Never** has problems with missing school because of not feeling well	**Never** has problems with missing school because of not feeling well

Which health description would a 10-YEAR-OLD CHILD prefer to live with?

Health Description A ○ Health Description B ○

[a] Example DCE task shown is for valuing the health of a 10-year-old child. Three other ages (3-year-old, 6-year-old, and 15-year-old) were also used in the health state descriptions but are not shown in the figure.

The survey was first pretested with 12 participants to select the perspective and framing of whose health to value in the choice tasks. Based on participant feedback during initial pretesting, it was determined that adopting the other-child perspective, which asked adult respondents to value the health of a hypothetical child, would increase comprehension of choice tasks.[*] Moreover, two different framings of the choice tasks were implemented in response to participant feedback. For age groups 5-7 years, 8-12 years, and 13-18 years, respondents were asked: *"Which health description would a X-year-old child prefer to live with?"*. This framing was selected in order to avoid adult proxies paternalistically deciding what is best for the child without considering what the child may want for themselves. On the other hand, respondents for the youngest age group were asked: *"Which health description do YOU prefer for a 3-year-old child to live with?"*. The choice tasks for the youngest age group were framed differently because pretest participants found it difficult to accurately state the preferences of a 3-year-old child. The two frames were further evaluated in subsequent rounds of pretesting (*n* = 15) and were found to be stable. Wording, layout, and format of the choice tasks were also pretested to validate survey design. The online survey was developed using Qualtrics software Version April 2023.[158] A formal pilot test (*n* = 329) was conducted from April 17, 2023 to April 26, 2023.

[*] Two different perspectives were pretested, including the other-child perspective (described above in text) and the self-child perspective in which respondents were asked to imagine themselves as a child. These perspectives were developed based on previously published literature on the effects of perspective on valuation outcomes.[68-71] The other-child perspective was selected for use in this study because respondents found it difficult to distinguish between current adult preferences and self-child preferences.

DCE Experimental Design

The PedsUtil health state classification system is comprised of 5^7 = 78,125 unique health states for each age group. Since it was not feasible to value all possible health states, 280 pairs of health states or choice sets were selected for valuation for each age group that maximized statistical efficiency while minimizing respondent burden (described in more detail below). The 280 choice sets were divided into 35 blocks with 8 choice sets in each block. Respondents were randomized to a single block of questions, and each respondent was asked to answer the choice sets for 2 different age groups for a total of 16 choice sets. The presentation order of choice sets for each age group was randomized to minimize ordering bias.

Simultaneously comparing 7 dimensions of health between 2 health states was determined to be too difficult of a task, thus the number of dimensions that differed between the health states was constrained to 4 dimensions as done in previous studies.[152,153,159] As illustrated in **Figure 4.1**, the dimensions that differed between “Health Description A” and “Health Description B” were highlighted in magenta[†] to help respondents identify these differences. A forced choice fractional factorial design was first generated that specified the levels of the 4 dimensions that differed between the 2 health states. A restriction was imposed in the design to eliminate any dominated pairs of health states (i.e., all dimension levels are better or worse for one health state compared to the other health state). This resulted in 40 choice sets in the 4-dimension base design (**aTable 4.1**). This experimental design was generated using SAS software

[†] This color was specifically chosen to optimize graphics for color blindness accessibility.

Version 9.4.[160] The next step was to then identify which 4 dimensions (out of 7) to vary between the 2 health states. There are 35 distinct ways[‡] to vary 4 out of 7 dimensions. A subset (7) of the 35 combinations was selected so that each pair of dimensions in a health state was varied equally often (**aTable 4.2**). Consequently, each choice set in the 4-dimension base design was applied to the 7 combinations, resulting in 280 total choice sets.[§] For the 3 dimensions that did not differ between the 2 health states, the mildest severity level ("Never") was assigned to these dimensions as it was presumed to reduce respondent burden. The final experimental design is reported in **aTable 4.3**.

TTO Survey Content and Development

Each TTO survey was comprised of the following elements (in order): study overview and IRB information, screening questions, instructions, 1 practice VAS, 1 practice TTO, first set of valuation tasks for first age group, 1 debrief question, instructions, second set of valuation tasks for second age group, 1 debrief question, and additional background and sociodemographic questions. The valuation tasks included both VAS and TTO tasks for each health state (**Figures 4.2 and 4.3**). The purpose of incorporating VAS into the valuation protocol was to familiarize respondents to each health state before valuing the health state using TTO. Similar to the DCE survey, the median age for each pediatric age group valued was used in the health state descriptions for the VAS and TTO tasks.

---

‡ Selecting 4 out of 7 dimensions to differ in levels between the 2 health states is equal to $C_4(7) = \binom{7}{4}$ = 35 different combinations.

§ 40 choice sets x 7 combinations = 280 choice sets.

Figure 4.2 Example VAS Task[a]

Health Description X:

Dimension	Health Description
Physical Functioning	**Often** has problems with participating in sports activity or exercise
Pain	**Sometimes** has problems with having hurts or aches
Fatigue	**Often** has problems with low energy level
Emotional Functioning	**Sometimes** has problems with worrying about what will happen to them
Social Functioning	**Sometimes** has problems with other kids not wanting to be their friend
School Functioning	**Sometimes** has problems with keeping up with schoolwork
School Absence	**Often** has problems with missing school because of not feeling well

Practice Question 1:

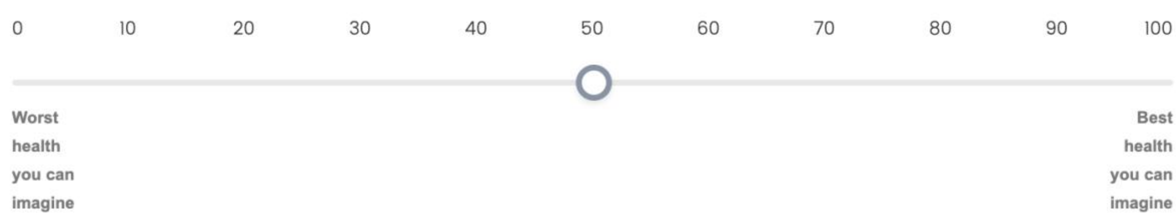
Please rate **Health Description X** on a scale of 0 to 100 from the perspective of a **10-YEAR-OLD CHILD** where

- 0 equals the worst health you can imagine
- 100 equals the best health you can imagine

Please rate the health description from the **point of view of a 10-YEAR-OLD CHILD.**

Click and drag the marker on the scale below to indicate your rating. Please make sure to click on the marker so that your answer is recorded.

Rating for **Health Description X** of a **10-YEAR-OLD CHILD**:



[a] Example VAS task shown is for valuing the health of a 10-year-old child. Three other ages (3-year-old, 6-year-old, and 15-year-old) were also used in the health state descriptions but are not shown in the figure.

Figure 4.3 Example TTO Task[a]

Health Description X:

Dimension	Health Description
Physical Functioning	**Often** has problems with participating in sports activity or exercise
Pain	**Sometimes** has problems with having hurts or aches
Fatigue	**Often** has problems with low energy level
Emotional Functioning	**Sometimes** has problems with worrying about what will happen to them
Social Functioning	**Sometimes** has problems with other kids not wanting to be their friend
School Functioning	**Sometimes** has problems with keeping up with schoolwork
School Absence	**Often** has problems with missing school because of not feeling well

Assume that the **10-YEAR-OLD CHILD** has 10 years left to live (lives until 20 years old) with this health description.

Imagine a scenario where the **10-YEAR-OLD CHILD** could trade off time from the end of their life in order to completely avoid living with **Health Description X** and instead live the remaining years of their life in the best health you can imagine. The purpose of this exercise is to compare quality of life versus length of life.

Practice Question 2 (Part 1 of 3):

Do you think the **10-YEAR-OLD CHILD** would be willing to trade off **8 years** from the end of their life (lives until 12 years old) in order to avoid living with **Health Description X** and instead live their remaining years in the best health you can imagine?

Note: *Please consider* ***the point of view of a 10-YEAR-OLD CHILD*** *when answering the question.*

Yes	No
○	○

Practice Question 2 (Part 2 of 3):

If you think the child is willing to trade off 8 years, do you think the **10-YEAR-OLD CHILD** would be willing to trade off **9 years** from the end of their life (lives until 11 years old) in order to avoid living with **Health Description X** (listed below) and instead live their remaining years in the best health you can imagine?

Note: *Please consider* ***the point of view of a 10-YEAR-OLD CHILD*** *when answering the question.*

Yes	No
○	○

Practice Question 2 (Part 3 of 3):

What is the **most** amount of time (out of 10 years) that you think a **10-YEAR-OLD CHILD** would be willing to trade off from the end of their life in order to avoid living with **Health Description X** (listed below)?

Note: *You may answer in more than one time category. Your final answer will be totaled (i.e., if you fill in 1 year and 6 months, your final answer will be 1½ years). Your answer may range from trading off no time (by entering 0 years, 0 months, 0 weeks, and 0 days) up to trading off 10 years.*

Year(s) [ ]

Month(s) [ ]

Week(s) [ ]

Day(s) [ ]

[a] Example TTO task shown is for valuing the health of a 10-year-old child. Three other ages (3-year-old, 6-year-old, and 15-year-old) were also used in the health state descriptions but are not shown in the figure.

The perspective and framing of whose health is being valued in the VAS and TTO tasks were developed to match those of the DCE tasks. For the VAS questions, respondents were asked to *"Please rate the health description on a scale of 0 to 100 from the perspective of a X-year-old child"* for age groups 5-7 years, 8-12 years, and 13-18 years. For age group 2-4 years, respondents were asked to *"Please rate the health description for a 3-year-old child on a scale of 0 to 100".* For the TTO tasks, a modified approach was utilized in this study which asked respondents to define the amount of time they would be willing to trade off to prevent a health state in a child.[161,162] Specifically, age groups 5-7 years, 8-12 years, and 13-18 years were asked: *"What is the most amount of time that you think a X-year-old child would be willing to trade off from the end of their life in order to avoid living with the health description?".* For age group 2-4 years, respondents were asked: *"What is the most amount of time you would be willing to trade off from the end of the child's life in order for the 3-year-old child to avoid living with the health description?"*. Extensive pretesting and cognitive interviews (n = 30) were conducted to ensure respondents understood the framing and wording of the valuation tasks. The online survey was developed using Qualtrics software Version April 2023.[158] A formal pilot test (n = 94) was conducted on April 25, 2023.

TTO Survey Design

Three health states were selected for inclusion in the online TTO survey (**aFigure 4.2**) – 2 moderate to severe health states (3333333 and 4453354) and the worst (PITS) health state (5555555). No mild health states were included in the survey because previous studies have demonstrated that proxy respondents are especially unwilling to trade off time for mild health states in children.[22,161,162] Only 3 health states were valued in the survey to mitigate respondent fatigue, and the order of health states was presented in increasing severity (i.e., 3333333, 4453354, 5555555) to reduce the difficulty of completing the valuation tasks. Since respondents were randomized to 2 out of 4 different age groups, each respondent was asked to complete 6 VAS and 6 TTO tasks.

All TTO tasks adopted a life expectancy of 10 years (as opposed to the full lifetime of the child) since the health state descriptions were most relevant to the childhood years. Previous child health state valuation studies have also implemented TTO tasks using 10 years as the time horizon.[73,91-99] The TTO algorithm used 3 different starting bids (i.e., 2 years, 5 years, and 8 years), which was followed by an additional bid, and then terminated with a final TTO question about the maximum amount of time respondents would be willing to trade off. A bisection approach was used to create the TTO bids. For example, if a respondent stated that they were willing to trade off 5 years, then the next question asked if the respondent would be willing to trade off 7.5 years (which is halfway between 5 years and 10 years). If a respondent was not willing to trade off 5 years, then the next question asked if the respondent would be willing to trade off 2.5 years. An internet-based approach was employed to administer the TTO survey as has been found to be valid and efficient in previous research.[163,164] One of the advantages of using an online survey is that it allows for various parts of the survey to be randomized with relative ease. Respondents in this survey were randomized to different starting TTO bids in order to minimize anchoring bias. Additionally, the survey was designed so that respondents were able to modify or correct previous answers if they chose to do so.

Data Collection and Sampling

Sampling and survey administration to collect pilot data for both DCE and TTO surveys were undertaken by Cint, a survey research firm. Pilot participants were recruited from the Cint online panel of US adults. Two separate pilot samples were obtained for the DCE and TTO surveys. The target pilot sample sizes were approximately 10% of the estimated sample sizes for when the surveys will be fielded large-scale.[**]

Various criteria were employed to check for respondent attentiveness and data quality. For both DCE and TTO surveys, respondents were asked to indicate their age bracket and were also asked to specify their age in two different sections of the surveys. Respondents whose answers did not match for all age questions were excluded from data analysis. Respondents who input nonsensical ages (i.e., over 100 years old) were also excluded. For the DCE, the second practice question was a dominance test in which one health state (1111111) was considered to logically dominate the other health state (5551151). Respondents who failed the dominance test were omitted. For the TTO survey, respondents who traded off more than 10 years (the maximum TTO amount allowed) were excluded. Lastly, pilot participants with a survey completion time less than 1/3 of the sample median completion time were excluded from data analysis as they were presumed to not have reliably completed the choice tasks.

[**] Responses to the large-scale DCE and TTO surveys will be obtained from population-representative US samples of adults. Quota sampling will be used to ensure that the samples are representative of the general US adult population in terms of age, gender, ethnicity, race, and region. The target sample size for the large-scale DCE will be 1,000 respondents per age group, which is approximately 30 responses per choice set for each age group. Sample size calculations were based on the recommendations proposed by Lancsar and Louviere,[165] which suggested a minimum of 20 observations per choice set, and Hensher et al.,[166] which recommended a minimum of 30 observations per choice set. The target sample size for the large-scale TTO will be 300 respondents per age group, which is consistent with or greater than the number of observations per health state used in other TTO studies that have been found to be valid and fit for task.[73,91-98]

Data Analysis

TTO Valuation

TTO values for the 3 health states were calculated as:

$$H_{ij} = (-x_{ij})/t \qquad \textbf{Eq. 4.1}$$

where H_{ij} represents the health disutility for health state j valued by respondent i, x_{ij} represents the amount of time traded off by respondent i to avoid living with health state j, and t represents the life expectancy or maximum TTO amount (i.e., 10 years). TTO values were structured as disutilities (rather than utilities) to facilitate the ease of data analysis. Health utilities can be calculated by subtracting the disutility from 1 assuming that baseline is perfect health. Mean TTO values for each health state were estimated for all age groups combined and for each age group.

Unanchored DCE Models

DCE data were analyzed using the conditional logistic regression model, which was shown by McFadden to be consistent with random utility theory.[150] Random utility theory assumes that the utility of respondents can be broken down into two parts: 1) a systematic component (observed), which is a function of the dimensions and their levels; and 2) a random component (unobserved), which is the error term that reflects the inability to perfectly measure utility.[151] The general model specification for the linear, additive utility function used in this study was:

$$D_j = X'_j\beta + \varepsilon_j \qquad \textbf{Eq. 4.2}$$

where D_j represents the latent disutility for health state j, X represents a vector of dummy variables for each level λ of dimension ∂ of the PedsUtil health state classification system that defines health state j, β represents a vector of estimated coefficients (i.e., preference weights), and ε represents the random error term. Since "Never" was used as the reference level in the utility function, 28 coefficients (i.e., 7 dimensions x [5-1] levels in each) were estimated in the model. Each coefficient indicates the utility decrement from the reference level to the respective level in each dimension. As the dummy coefficients represent progressively worse levels in each dimension compared to the reference level, the coefficients were expected to be negative and increase in absolute size. If non-monotonicity was observed among model coefficients within a dimension, the non-monotonic levels were combined and models were re-estimated to obtain a parsimonious consistent model. Standard errors were adjusted to allow for intra-individual correlation by using a clustered sandwich estimator. A combined model was estimated that included all age groups, as well as separate models for each age group. Relative importance scores of the dimensions were also calculated by dividing the range of coefficients for each dimension by the sum of all ranges.

Anchoring DCE Data Using TTO

Two different approaches were used to anchor the DCE values onto the full 0-1 QALY scale.[43,45,167,168] The first approach anchored the DCE value of the PITS state to the mean observed TTO value of the PITS state. This was achieved by estimating a rescaling parameter, θ, to normalize the DCE values:

$$\theta = \overline{H}_{PITS}/D_{PITS} \quad \textbf{Eq. 4.3}$$

where $\overline{H}_{PITS}$ is the mean observed TTO disutility value associated with the PITS state and D_{PITS} is the predicted latent DCE disutility value for the PITS state (i.e., sum of the model coefficients for the worst level in each dimension). To anchor the DCE data, the model coefficients estimated in **Eq. 4.2** were multiplied by θ.

The second approach mapped the mean observed TTO disutility values for the 3 health states included in the TTO survey onto the predicted latent DCE disutility values using ordinary least squares (OLS) regression:

$$\overline{H}_j = \gamma_0 + \theta D_j \quad \textbf{Eq. 4.4}$$

where $\overline{H}_j$ is the mean observed TTO disutility value for health state j, D_j is the predicted latent DCE disutility value for the same health state j, γ_0 is the regression intercept, and θ is the rescaling parameter or the slope between the TTO and DCE values. The model was estimated with and without the inclusion of an intercept. The DCE latent scale coefficients were then multiplied by the rescaling parameter, θ, to anchor the DCE values.

The final rescaling approach to construct the PedsUtil scoring system was selected based on theoretical considerations and the degree of agreement between the rescaled DCE values and observed mean TTO values as measured by the mean absolute error (MAE) at the health state level. A lower MAE indicates better accuracy of the model. All statistical analyses were conducted using Stata Version 14.1.[169]

Results

Pilot Sample Characteristics

A pilot sample consisting of 329 participants was collected for the DCE. Of those, 3 participants were removed because they indicated they were <18 years old, thus falling outside of eligibility criteria. An additional 113 participants were removed due to data quality issues (described in more detail below). Therefore, 213 participants were included in the final pilot sample that was used for data analysis. The TTO pilot sample consisted of 94 respondents. Of those, 2 participants were excluded because they indicated they were <18 years old and 11 respondents were excluded based on data quality control criteria. A total of 81 respondents were included in the final TTO pilot sample. **Table 4.1** displays the sociodemographic characteristics of survey respondents by the pediatric age group that respondents were randomized to. Since data were collected as pilot samples, data are not representative of the US population.

Table 4.1 Summary of Pilot Sample Characteristics Across Age Groups

Characteristic	Age Group[a]			
	2-4 years	5-7 years	8-12 years	13-18 years
DCE Sample, *n*	99	116	113	98
Age, y				
Mean (SD)	46.7 (16.7)	48.5 (17.2)	47.9 (16.3)	46.4 (16.2)
Gender [*n* (%)]				
Male	41 (41.4)	50 (43.1)	43 (38.1)	28 (28.6)
Female	58 (58.6)	65 (56.0)	68 (60.2)	69 (70.4)
Transgender	0 (0.0)	1 (0.9)	1 (0.9)	0 (0.0)
Gender-variant or gender-nonconforming	0 (0.0)	0 (0.0)	1 (0.9)	1 (1.0)
Ethnicity [*n* (%)]				
Hispanic	8 (8.1)	4 (3.4)	6 (5.3)	4 (4.1)
Not Hispanic	91 (91.9)	112 (96.6)	107 (94.7)	94 (95.9)
Race [*n* (%)]				
American Indian or Alaskan Native	4 (4.0)	2 (1.7)	1 (0.9)	1 (1.0)
Asian	2 (2.0)	1 (0.9)	1 (0.9)	2 (2.0)
Black or African American	12 (12.1)	15 (12.9)	16 (14.2)	13 (13.3)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
White	78 (78.8)	95 (81.9)	94 (83.2)	81 (82.7)
More than one race	1 (1.0)	0 (0.0)	0 (0.0)	1 (1.0)
Other	2 (2.0)	3 (2.6)	1 (0.9)	0 (0.0)
Region[b] [*n* (%)]				
Northeast	17 (17.2)	25 (21.6)	23 (20.4)	23 (23.5)
Midwest	22 (22.2)	25 (21.6)	31 (27.4)	26 (26.5)
South	38 (38.4)	45 (38.8)	43 (38.1)	32 (32.7)
West	22 (22.2)	21 (18.1)	16 (14.2)	17 (17.3)
Marital status [*n* (%)]				
Married or Living with partner	59 (59.6)	58 (50.0)	60 (53.1)	47 (48.0)
Single	27 (27.3)	38 (32.8)	33 (29.2)	34 (34.7)
Divorced or Separated	10 (10.1)	12 (10.3)	15 (13.3)	11 (11.2)
Widowed	3 (3.0)	8 (6.9)	5 (4.4)	6 (6.1)
Highest level of education [*n* (%)]				
Less than high school	2 (2.0)	0 (0.0)	5 (4.4)	5 (5.1)
High school/GED	43 (43.4)	49 (42.2)	47 (41.6)	39 (39.8)
2-year college/Associate's degree	18 (18.2)	22 (19.0)	16 (14.2)	18 (18.4)
4-year college degree	23 (23.2)	29 (25.0)	33 (29.2)	25 (25.5)
Advanced degree (Master's, Doctorate or Professional)	13 (13.1)	13 (11.2)	9 (8.0)	11 (11.2)
Other	0 (0.0)	3 (2.6)	3 (2.7)	0 (0.0)
Employment [*n* (%)]				
Full time	44 (44.4)	51 (44.0)	48 (42.5)	43 (43.9)
Part time	15 (15.2)	15 (12.9)	13 (11.5)	15 (15.3)
Not employed outside of the home	40 (40.4)	50 (43.1)	52 (46.0)	40 (40.8)
Household income[c] [*n* (%)]				
≤$35,000	36 (36.4)	36 (31.0)	41 (36.3)	35 (35.7)
>$35,000 – $50,000	17 (17.2)	22 (19.0)	14 (12.4)	19 (19.4)
>$50,000 – $75,000	17 (17.2)	22 (19.0)	19 (16.8)	16 (16.3)
>$75,000 – $100,000	13 (13.1)	15 (12.9)	19 (16.8)	13 (13.3)
>$100,000	16 (16.2)	19 (16.4)	16 (14.2)	13 (13.3)
Prefer not to disclose	0 (0.0)	2 (1.7)	4 (3.5)	2 (2.0)
Have children [*n* (%)]				
Yes	59 (59.6)	72 (62.1)	65 (57.5)	56 (57.1)
No	40 (40.4)	44 (37.9)	48 (42.5)	42 (42.9)
TTO Sample, *n*	42	43	38	39
Age, y				
Mean (SD)	45.7 (15.8)	43.4 (14.8)	48.8 (13.8)	45.9 (16.6)
Gender [*n* (%)]				
Male	20 (47.6)	24 (55.8)	15 (39.5)	17 (43.6)

Female	22 (52.4)	19 (44.2)	23 (60.5)	22 (56.4)
Transgender	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Gender-variant or gender-nonconforming	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity [*n* (%)]				
Hispanic	1 (2.4)	4 (9.3)	3 (7.9)	2 (5.1)
Not Hispanic	41 (97.6)	39 (90.7)	35 (92.1)	37 (94.9)
Race [*n* (%)]				
American Indian or Alaskan Native	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Asian	2 (4.8)	2 (4.7)	1 (2.6)	3 (7.7)
Black or African American	5 (11.9)	5 (11.6)	6 (15.8)	2 (5.1)
Native Hawaiian or Other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
White	34 (81.0)	34 (79.1)	30 (78.9)	34 (87.2)
More than one race	1 (2.4)	1 (2.3)	0 (0.0)	0 (0.0)
Other	0 (0.0)	1 (2.3)	1 (2.6)	0 (0.0)
Region[b] [*n* (%)]				
Northeast	8 (19.0)	12 (27.9)	9 (23.7)	11 (28.2)
Midwest	12 (28.6)	5 (11.6)	9 (23.7)	6 (15.4)
South	15 (35.7)	19 (44.2)	15 (39.5)	15 (38.5)
West	7 (16.7)	7 (16.3)	5 (13.2)	7 (17.9)
Marital status [*n* (%)]				
Married or Living with partner	23 (54.8)	24 (55.8)	22 (57.9)	23 (59.0)
Single	11 (26.2)	12 (27.9)	11 (28.9)	6 (15.4)
Divorced or Separated	6 (14.4)	7 (16.3)	3 (7.9)	8 (20.5)
Widowed	2 (4.8)	0 (0.0)	2 (5.3)	2 (5.1)
Highest level of education [*n* (%)]				
Less than high school	1 (2.4)	1 (2.3)	0 (0.0)	0 (0.0)
High school/GED	15 (35.7)	11 (25.6)	16 (42.1)	14 (35.9)
2-year college/Associate’s degree	13 (31.0)	9 (20.9)	8 (21.1)	10 (25.6)
4-year college degree	11 (26.2)	12 (27.9)	9 (23.7)	10 (25.6)
Advanced degree (Master’s, Doctorate or Professional)	2 (4.8)	10 (23.3)	5 (13.2)	5 (12.8)
Other	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Employment [*n* (%)]				
Full time	24 (57.1)	30 (69.8)	20 (52.6)	24 (61.5)
Part time	5 (11.9)	2 (4.7)	1 (2.6)	4 (10.3)
Not employed outside of the home	13 (31.0)	11 (25.6)	17 (44.7)	11 (28.2)
Household income[c] [*n* (%)]				
≤$35,000	11 (26.2)	13 (30.2)	15 (39.5)	11 (28.2)
>$35,000 – $50,000 (4)	10 (23.8)	7 (16.3)	8 (21.1)	7 (17.9)
>$50,000 – $75,000 (5)	8 (19.0)	11 (25.6)	9 (23.7)	8 (20.5)
>$75,000 – $100,000 (6)	6 (14.3)	2 (4.7)	2 (5.3)	4 (10.3)
>$100,000 (7,8)	7 (16.7)	10 (23.3)	4 (10.5)	9 (23.1)
Prefer not to disclose	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Have children [*n* (%)]				
Yes	23 (54.8)	26 (60.5)	21 (55.3)	28 (71.8)
No	19 (45.2)	17 (39.5)	17 (44.7)	11 (28.2)

Abbreviations: DCE, discrete choice experiment; y, years; SD, standard deviation; GED, general educational development; TTO, time trade-off.

[a] Age group refers to the pediatric age group that survey respondents were randomized to for completing valuation tasks.

[b] Northeast includes Main, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, Pennsylvania, and New Jersey; Midwest includes Wisconsin, Michigan, Illinois, Indiana, Ohio, North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, and Missouri; South includes Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Mississippi, Alabama, Oklahoma, Texas, Arkansas, and Louisiana; West includes Idaho, Montana, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico, Alaska, Washington, Oregon, California, and Hawaii.

[c] Household income refers to total household income before taxes in 2022.

Overall, the distribution of demographic characteristics was similar across respondents for the 4 age groups for both the DCE and TTO surveys. In both surveys, the majority of respondents were female, not Hispanic, white, employed (either part time or full time), married or living with a partner, and/or have children.

Valuation Data Quality

aTable 4.4 provides an overview of the data quality indicators for both surveys. There were few respondents who were speeders or had invalid answers to the age questions (i.e., responses to all age questions did not match or were nonsensical). The majority of DCE respondents (80%) who were excluded from data analysis were because they failed the dominance test. Similarly, a large portion of TTO respondents (46%) who were excluded were due to respondents trading off more time than the maximum TTO amount (i.e., 10 years). Inconsistencies in TTO responses, meaning that a respondent traded off more time for a milder health state compared to a more severe health state, were also calculated, though these responses were not excluded from the primary data analysis. A total of 40 respondents (49%) had at least 1 inconsistent TTO response (26% for 2-4 years, 42% for 5-7 years, 50% 8-12 years, and 23% for 13-18 years).[††] Lastly, respondents were asked how confident they felt about their answers to the valuation tasks. Among DCE respondents, less than 8% said that they were not confident and less than 2% said that their answers were total guesses. Among TTO

[††] Inconsistencies in VAS responses, meaning that a milder health state was rated lower than a more severe health state, were also calculated. A total of 55 respondents (68%) had at least 1 inconsistent VAS response (36% for 2-4 years, 56% for 5-7 years, 45% 8-12 years, and 49% for 13-18 years).

respondents, only 12% said that they were not confident and less than 4% said that their answers were total guesses.

VAS and TTO Summary Statistics

VAS summary statistics are presented in **aTable 4.5**. There were no inconsistencies in the observed mean VAS scores. For health state 3333333, the youngest age group (2-4 years) had the highest mean VAS score (70.2) and age group 5-7 years had the lowest mean VAS score (60.9). The mean VAS scores for health state 4453354 were similar across all age groups, which ranged from 56.4 (13-18 years) to 58.5 (2-4 years). The mean VAS scores for the PITS state (5555555) were lowest for age groups 2-4 years and 5-7 years (47.4 and 47.2, respectively) and highest for the oldest age group 13-18 years (51.8).

Table 4.2 shows the summary statistics for the TTO health utilities. Similar to the VAS scores, there were no inconsistencies in the observed mean TTO health utilities for all age groups. On average, respondents were less willing to trade off time for younger age groups, resulting in higher health utilities for younger children for all health states. The mean health utility value assigned to the PITS state (5555555) was 0.70 for all age groups combined, and ranged from 0.66 (8-12 years) to 0.75 (2-4 years) across age groups. The mean observed health utilities for the other 2 health states ranged from 0.72 (13-18 years) to 0.86 (2-4 years).

Table 4.2 TTO Summary Statistics for the 3 Health States Across Age Groups

Health State	*n*[a]	Mean (SD)	Median	Range	% Non-traders[b]
All Ages					
3333333	161	0.80 (0.25)	0.89	0.00-1	33.5
4453354	162	0.76 (0.28)	0.84	0.00-1	27.2
5555555	162	0.70 (0.31)	0.80	0.00-1	22.8
2-4 years					
3333333	42	0.86 (0.20)	0.90	0.16-1	42.9
4453354	42	0.78 (0.26)	0.89	0.10-1	33.3
5555555	42	0.75 (0.29)	0.80	0.00-1	23.8
5-7 years					
3333333	42	0.81 (0.21)	0.87	0.16-1	26.2
4453354	43	0.75 (0.27)	0.80	0.10-1	25.6
5555555	43	0.70 (0.32)	0.79	0.00-1	23.3
8-12 years					
3333333	38	0.78 (0.26)	0.87	0.05-1	31.6
4453354	38	0.77 (0.27)	0.86	0.05-1	23.7
5555555	38	0.66 (0.31)	0.76	0.00-1	21.1
13-18 years					
3333333	39	0.76 (0.32)	0.89	0.00-1	33.3
4453354	39	0.72 (0.32)	0.80	0.00-1	25.6
5555555	39	0.69 (0.34)	0.80	0.00-1	23.1

Abbreviations: SD, standard deviation.
[a] *n* refers to the number of observations.
[b] Non-traders refer to respondents who were unwilling to trade off any time (TTO amount = 0 years).

Unanchored DCE Model

Table 4.3 reports the conditional logistic regression model results. Most of the coefficients at the more severe dimension levels (i.e., levels 4 and 5) were statistically significant compared to reference level 1. Due to the limited pilot sample size, there were some inconsistencies in model coefficients, especially for the age-specific models, meaning that some model coefficients exhibited non-monotonic patterns with increasing (worsening) dimension level. The model for the youngest age group 2-4 years exhibited the greatest number of inconsistencies. Further modeling was undertaken to obtain monotonic models and results are shown in **Table 4.4**. Similar to the previous models, the majority of coefficients at higher levels of severity were found to be statistically significant. The importance of dimensions, as determined by the relative importance scores, was also found to vary by age group. Specifically, Social Functioning was the most important dimension for age groups 2-4 years and 13-18 years, School Functioning for age group 5-7 years, and Emotional Functioning for age group 8-12 years. In terms of least importance, Fatigue was the least important dimension for age groups 2-4 years and 8-12 years and School Absence for age groups 5-7 years and 13-18 years.

Table 4.3 DCE Latent Preference Weights Across Age Groups

Level	All Ages		2-4 years		5-7 years		8-12 years		13-18 years	
	Coefficient[a]	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Phys2	0.048	0.106	-0.013	0.202	0.227	0.169	0.171	0.183	-0.240	0.231
Phys3	-0.089	0.114	-0.456**	0.196	0.135	0.202	-0.027	0.192	-0.049	0.210
Phys4	-0.215**	0.106	-0.409**	0.189	-0.052	0.188	-0.250	0.180	-0.180	0.191
Phys5	-0.446***	0.119	-0.744***	0.197	-0.226	0.196	-0.245	0.205	-0.685***	0.235
Pain2	0.066	0.094	0.152	0.183	-0.083	0.174	-0.009	0.192	0.274	0.199
Pain3	-0.216**	0.103	-0.096	0.210	-0.324*	0.196	-0.201	0.194	-0.231	0.189
Pain4	-0.414***	0.112	-0.304	0.221	-0.498***	0.194	-0.499**	0.201	-0.329*	0.199
Pain5	-0.573***	0.122	-0.614***	0.218	-0.722***	0.206	-0.582***	0.206	-0.377	0.240
Fatigue2	0.113	0.115	0.381*	0.202	0.141	0.188	0.115	0.205	-0.151	0.234
Fatigue3	-0.153	0.111	0.044	0.208	-0.190	0.191	-0.126	0.199	-0.293	0.223
Fatigue4	-0.294***	0.106	0.092	0.191	-0.502**	0.199	-0.262	0.200	-0.415**	0.177
Fatigue5	-0.495***	0.116	-0.269	0.211	-0.605***	0.186	-0.275	0.203	-0.842***	0.219
Emot2	-0.117	0.109	0.217	0.182	-0.114	0.195	-0.331	0.216	-0.175	0.210
Emot3	-0.319***	0.103	-0.419**	0.175	-0.408**	0.179	-0.288	0.195	-0.152	0.202
Emot4	-0.581***	0.108	-0.225	0.198	-0.422**	0.203	-1.026***	0.184	-0.655***	0.187
Emot5	-0.811***	0.119	-0.668***	0.216	-0.685***	0.203	-1.099***	0.212	-0.893***	0.225
Soc2	-0.197*	0.108	-0.314	0.205	-0.084	0.176	-0.186	0.201	-0.252	0.219
Soc3	-0.461***	0.114	-0.481**	0.234	-0.316*	0.191	-0.526**	0.212	-0.559***	0.212
Soc4	-0.713***	0.118	-0.482**	0.221	-0.594***	0.187	-0.857***	0.237	-0.973***	0.198
Soc5	-0.839***	0.139	-0.890***	0.254	-0.729***	0.223	-0.912***	0.250	-0.877***	0.258
School2	0.126	0.113	0.277	0.188	-0.188	0.171	0.209	0.195	0.292	0.226
School3	-0.283***	0.098	-0.363*	0.193	-0.404**	0.167	-0.102	0.160	-0.246	0.200
School4	-0.378***	0.103	-0.519***	0.192	-0.617***	0.170	-0.191	0.186	-0.164	0.207
School5	-0.546***	0.114	-0.471**	0.200	-0.764***	0.191	-0.453**	0.203	-0.513**	0.239
SchAbs2	-0.066	0.096	-0.076	0.170	-0.095	0.179	0.035	0.191	-0.176	0.188
SchAbs3	-0.236**	0.105	-0.238	0.180	-0.147	0.191	-0.370*	0.190	-0.239	0.226
SchAbs4	-0.595***	0.108	-0.818***	0.207	-0.365**	0.183	-0.730***	0.212	-0.560***	0.206
SchAbs5	-0.528***	0.114	-0.687***	0.210	-0.241	0.182	-0.805***	0.204	-0.448*	0.248

Abbreviations: SE, standard error; Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.

[a] Inconsistencies in model coefficients are highlighted in red. *p-value<0.1, **p-value<0.05, ***p-value<0.01.

Table 4.4 DCE Monotonic Latent Preference Weights Across Age Groups

Level	All Ages			2-4 years			5-7 years			8-12 years			13-18 years		
	Coeff.[a]	SE	RI[b] %	Coeff.	SE	RI %	Coeff.	SE	RI %	Coeff.	SE	RI %	Coeff.	SE	RI %
Phys2	0.000	–[c]	10.6	-0.036	0.199	15.5	0.000	–	8.3	0.000	–	7.2	-0.158	0.183	13.4
Phys3	-0.114	0.105		-0.442***	0.164		0.000	–		-0.111	0.174		-0.158	0.183	
Phys4	-0.235**	0.093		-0.442***	0.164		-0.165	0.157		-0.337***	0.129		-0.170	0.187	
Phys5	-0.472***	0.104		-0.759***	0.200		-0.355**	0.166		-0.337***	0.129		-0.671***	0.238	
Pain2	0.000	–	13.6	0.000	–	13.9	-0.098	0.172	17.2	-0.002	0.191	12.4	0.000	–	10.2
Pain3	-0.255***	0.087		-0.171	0.180		-0.349*	0.192		-0.205	0.194		-0.373**	0.156	
Pain4	-0.448***	0.096		-0.383**	0.188		-0.516***	0.192		-0.495**	0.202		-0.474***	0.157	
Pain5	-0.607***	0.108		-0.683***	0.190		-0.732***	0.206		-0.578***	0.206		-0.511***	0.192	
Fatigue2	0.000	–	12.4	0.000	–	7.9	0.000	–	16.1	0.000	–	7.2	-0.172	0.231	17.0
Fatigue3	-0.213**	0.097		0.000	–		-0.269*	0.160		-0.189	0.174		-0.319	0.221	
Fatigue4	-0.349***	0.091		0.000	–		-0.570***	0.168		-0.312*	0.167		-0.420**	0.176	
Fatigue5	-0.554***	0.100		-0.385**	0.164		-0.684***	0.155		-0.333*	0.172		-0.852***	0.217	
Emot2	-0.127	0.109	18.3	0.000	–	15.9	-0.111	0.195	16.2	-0.310*	0.170	23.7	-0.176	0.171	17.7
Emot3	-0.321***	0.103		-0.409***	0.137		-0.406**	0.178		-0.310*	0.170		-0.176	0.171	
Emot4	-0.582***	0.109		-0.409***	0.137		-0.417**	0.205		-1.027***	0.185		-0.643***	0.188	
Emot5	-0.819***	0.118		-0.780***	0.207		-0.691***	0.203		-1.104***	0.211		-0.885***	0.226	
Soc2	-0.204*	0.108	18.8	-0.333	0.205	18.3	-0.086	0.176	17.1	-0.192	0.199	19.6	-0.263	0.219	18.5
Soc3	-0.468***	0.113		-0.495**	0.200		-0.326*	0.190		-0.528**	0.211		-0.571***	0.206	
Soc4	-0.712***	0.117		-0.495**	0.200		-0.592***	0.186		-0.857***	0.237		-0.923***	0.206	
Soc5	-0.841***	0.140		-0.898***	0.259		-0.727***	0.221		-0.910***	0.251		-0.923***	0.206	
School2	0.000	–	13.7	0.000	–	13.0	-0.199	0.171	17.9	0.000	–	12.1	0.000	–	13.3
School3	-0.342***	0.090		-0.502***	0.177		-0.398**	0.168		-0.197	0.151		-0.346**	0.156	
School4	-0.438***	0.097		-0.639***	0.152		-0.617***	0.172		-0.287*	0.168		-0.346**	0.156	
School5	-0.613***	0.105		-0.639***	0.152		-0.764***	0.192		-0.564***	0.175		-0.666***	0.208	
SchAbs2	-0.070	0.095	12.6	-0.063	0.166	15.4	-0.109	0.177	7.2	0.000	–	17.7	-0.177	0.183	9.9
SchAbs3	-0.239**	0.105		-0.246	0.180		-0.154	0.189		-0.385**	0.171		-0.225	0.227	
SchAbs4	-0.565***	0.096		-0.756***	0.179		-0.306*	0.159		-0.749***	0.200		-0.496**	0.199	
SchAbs5	-0.565***	0.096		-0.756***	0.179		-0.306*	0.159		-0.824***	0.185		-0.496**	0.199	

Abbreviations: Coeff., coefficient; SE, standard error; RI, relative importance; Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.

[a] *p*-value<0.1, ***p*-value<0.05, ****p*-value<0.01.

[b] RI scores calculated by dividing the range of coefficients for each dimension by the sum of all ranges and then multiplying by 100.

[c] Standard errors not calculated for dimension levels combined with the reference level (“Never”).

Anchored DCE Values

The latent model coefficients from the monotonic models were used to obtain the anchored DCE values. **Table 4.5** presents the anchored DCE model coefficients using the different anchoring approaches (i.e., PITS anchoring and mapping with and without a constant). The rescaling parameter estimates, θ**,** and results from the OLS mapping models that were used to anchor the DCE values are reported in **aTable 4.6**. On average, anchoring on the PITS state overpredicted the DCE utility values compared to the mean observed TTO utility values across all age groups (**Figure 4.4**). Mapping without a constant also overpredicted the DCE utility values for the moderate health state (3333333), but underpredicted the DCE utility values for the PITS state across all age groups (**Figure 4.5**). Overall, mapping with a constant outperformed the other anchoring methods in terms of prediction accuracy for most age groups (MAE = 0.003 for all age groups combined) (**Figure 4.6**). The only exception was for the youngest age group 2-4 years, which had the lowest MAE (0.023) when DCE values were anchored onto the mean observed TTO value for the PITS state. This discrepancy may be due to the larger number of inconsistencies in model coefficients that was observed for the youngest age group.

Table 4.5 DCE Anchored Preference Weights Across Age Groups

Level	All Ages			2-4 years			5-7 years			8-12 years			13-18 years		
	PITS[a]	Map A[b]	Map B[c]	PITS	Map A	Map B	PITS	Map A	Map B	PITS	Map A	Map B	PITS	Map A	Map B
Phys2	0.000	0.000	0.000	-0.002	-0.002	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	-0.010	-0.012	-0.004
Phys3	-0.008	-0.008	-0.005	-0.023	-0.030	-0.017	0.000	0.000	0.000	-0.008	-0.008	-0.005	-0.010	-0.012	-0.004
Phys4	-0.016	-0.017	-0.010	-0.023	-0.030	-0.017	-0.012	-0.013	-0.008	-0.025	-0.026	-0.014	-0.011	-0.013	-0.005
Phys5	-0.032	-0.035	-0.019	-0.039	-0.051	-0.030	-0.025	-0.027	-0.016	-0.025	-0.026	-0.014	-0.042	-0.050	-0.018
Pain2	0.000	0.000	0.000	0.000	0.000	0.000	-0.007	-0.007	-0.004	0.000	0.000	0.000	0.000	0.000	0.000
Pain3	-0.017	-0.019	-0.010	-0.009	-0.012	-0.007	-0.025	-0.027	-0.016	-0.015	-0.016	-0.009	-0.023	-0.028	-0.010
Pain4	-0.030	-0.033	-0.018	-0.020	-0.026	-0.015	-0.037	-0.040	-0.024	-0.036	-0.038	-0.021	-0.030	-0.035	-0.013
Pain5	-0.041	-0.045	-0.025	-0.035	-0.046	-0.027	-0.052	-0.056	-0.033	-0.042	-0.044	-0.024	-0.032	-0.038	-0.014
Fatigue2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.011	-0.013	-0.005
Fatigue3	-0.014	-0.016	-0.009	0.000	0.000	0.000	-0.019	-0.021	-0.012	-0.014	-0.014	-0.008	-0.020	-0.024	-0.009
Fatigue4	-0.023	-0.026	-0.014	0.000	0.000	0.000	-0.040	-0.044	-0.026	-0.023	-0.024	-0.013	-0.026	-0.031	-0.011
Fatigue5	-0.037	-0.041	-0.023	-0.020	-0.026	-0.015	-0.049	-0.052	-0.031	-0.024	-0.025	-0.014	-0.054	-0.063	-0.023
Emot2	-0.009	-0.009	-0.005	0.000	0.000	0.000	-0.008	-0.009	-0.005	-0.023	-0.023	-0.013	-0.011	-0.013	-0.005
Emot3	-0.022	-0.024	-0.013	-0.021	-0.027	-0.016	-0.029	-0.031	-0.019	-0.023	-0.023	-0.013	-0.011	-0.013	-0.005
Emot4	-0.039	-0.043	-0.024	-0.021	-0.027	-0.016	-0.030	-0.032	-0.019	-0.075	-0.078	-0.043	-0.040	-0.047	-0.017
Emot5	-0.055	-0.061	-0.033	-0.040	-0.052	-0.030	-0.049	-0.053	-0.032	-0.080	-0.084	-0.046	-0.056	-0.065	-0.024
Soc2	-0.014	-0.015	-0.008	-0.017	-0.022	-0.013	-0.006	-0.007	-0.004	-0.014	-0.015	-0.008	-0.017	-0.019	-0.007
Soc3	-0.032	-0.035	-0.019	-0.026	-0.033	-0.019	-0.023	-0.025	-0.015	-0.038	-0.040	-0.022	-0.036	-0.042	-0.015
Soc4	-0.048	-0.053	-0.029	-0.026	-0.033	-0.019	-0.042	-0.045	-0.027	-0.062	-0.065	-0.036	-0.058	-0.068	-0.025
Soc5	-0.057	-0.062	-0.034	-0.046	-0.060	-0.035	-0.052	-0.056	-0.033	-0.066	-0.069	-0.038	-0.058	-0.068	-0.025
School2	0.000	0.000	0.000	0.000	0.000	0.000	-0.014	-0.015	-0.009	0.000	0.000	0.000	0.000	0.000	0.000
School3	-0.023	-0.025	-0.014	-0.026	-0.034	-0.020	-0.028	-0.030	-0.018	-0.014	-0.015	-0.008	-0.022	-0.026	-0.009
School4	-0.029	-0.032	-0.018	-0.033	-0.043	-0.025	-0.044	-0.047	-0.028	-0.021	-0.022	-0.012	-0.022	-0.026	-0.009
School5	-0.041	-0.045	-0.025	-0.033	-0.043	-0.025	-0.054	-0.059	-0.035	-0.041	-0.043	-0.024	-0.042	-0.049	-0.018
SchAbs2	-0.005	-0.005	-0.003	-0.003	-0.004	-0.002	-0.008	-0.008	-0.005	0.000	0.000	0.000	-0.011	-0.013	-0.005
SchAbs3	-0.016	-0.018	-0.010	-0.013	-0.017	-0.010	-0.011	-0.012	-0.007	-0.028	-0.029	-0.016	-0.014	-0.017	-0.006
SchAbs4	-0.038	-0.042	-0.023	-0.039	-0.051	-0.029	-0.022	-0.023	-0.014	-0.055	-0.057	-0.031	-0.031	-0.037	-0.013
SchAbs5	-0.038	-0.042	-0.023	-0.039	-0.051	-0.029	-0.022	-0.023	-0.014	-0.060	-0.062	-0.035	-0.031	-0.037	-0.013
Constant	0	0	-0.117	0	0	-0.109	0	0	-0.106	0	0	-0.124	0	0	-0.183
MAE	0.032	0.030	0.003	0.023	0.032	0.041	0.027	0.025	0.003	0.033	0.039	0.025	0.056	0.054	0.005

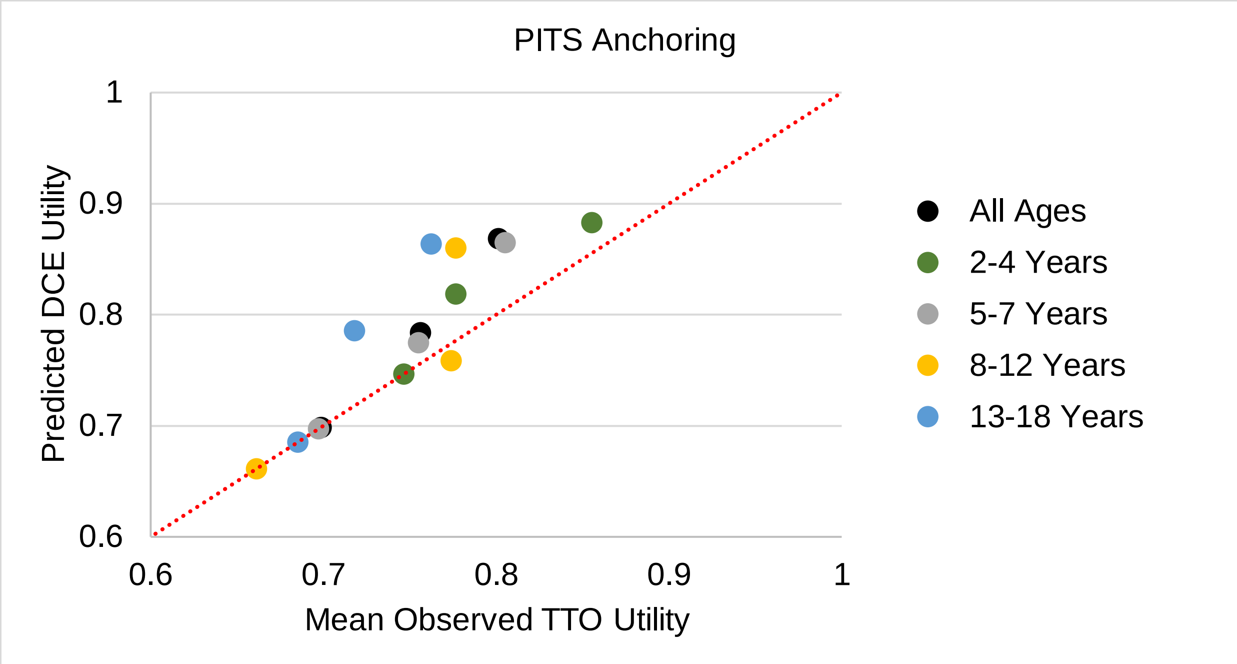
Abbreviations: Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence; MAE, mean absolute error.

[a] PITS refers to anchoring on the mean observed TTO value of the PITS state (5555555).

[b] Map A refers to mapping without a constant.

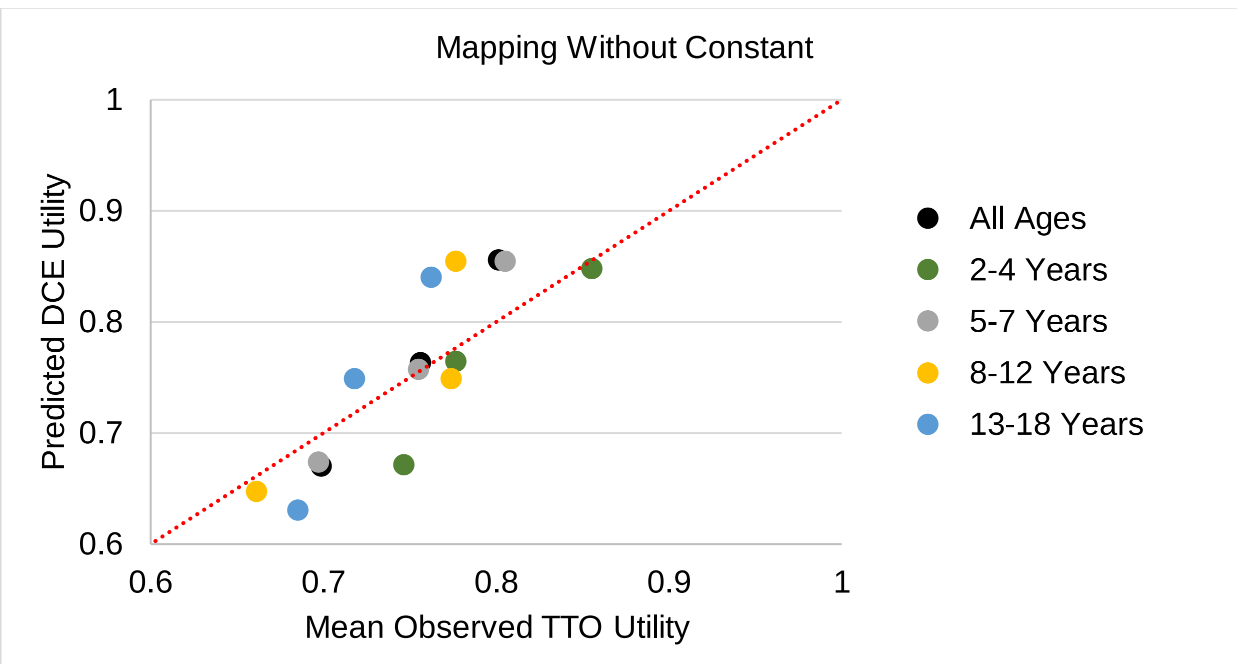
[c] Map B refers to mapping with a constant.

Figure 4.4 Predicted DCE vs. Observed TTO Utilities - PITS Anchoring



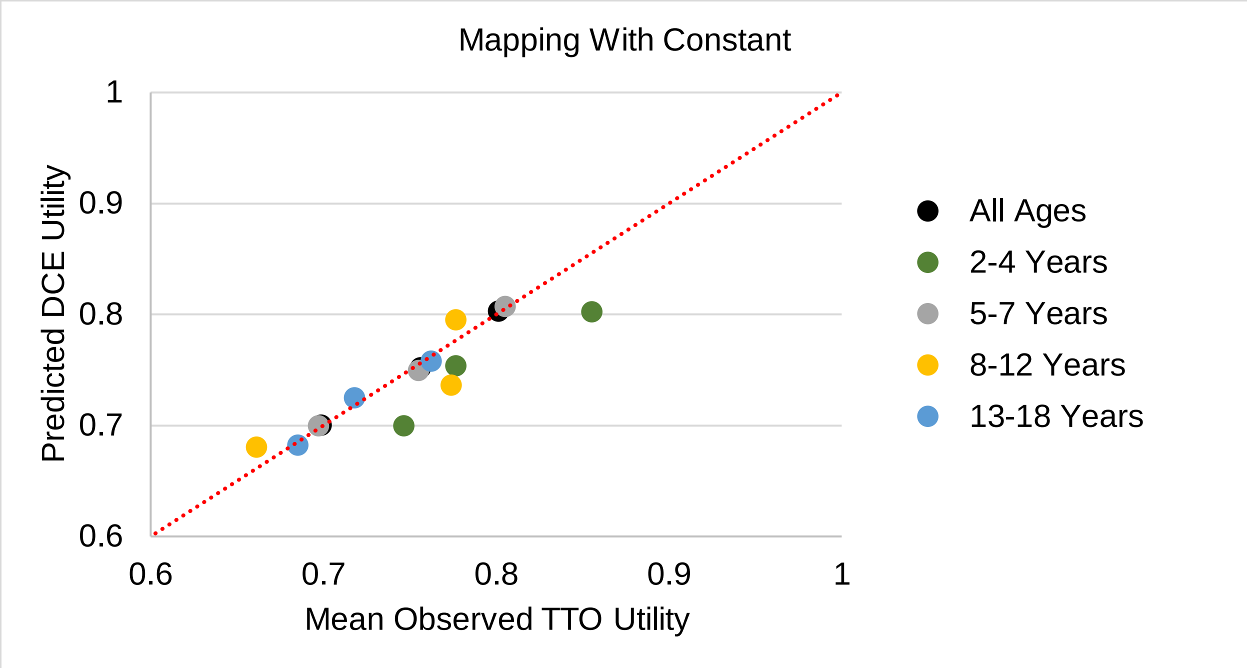
Abbreviations: DCE, discrete choice experiment; TTO, time trade-off.

Figure 4.5 Predicted DCE vs. Observed TTO Utilities - Mapping Without Constant



Abbreviations: DCE, discrete choice experiment; TTO, time trade-off.

Figure 4.6 Predicted DCE vs. Observed TTO Utilities - Mapping With Constant



Abbreviations: DCE, discrete choice experiment; TTO, time trade-off.

Though mapping with a constant was generally the best performing anchoring method in terms of prediction accuracy, using this method generated a scoring system with a compressed QALY scale. Specifically, the inclusion of a constant resulted in health utility values for health state 1111111 (i.e., health state with no health problems) that ranged from 0.82 (13-18 years) to 0.89 (5-7 years). As a result of this issue, mapping with a constant was not used to anchor the DCE values for the value set. The remaining two anchoring approaches had similar MAEs across all age groups, but anchoring on the PITS state resulted in stronger mispredictions for the more moderate health state (3333333). Milder health states are more prevalent in the general population, as well as in patient populations,[29,170] thus obtaining precise estimates for the moderate health state was prioritized over obtaining precise estimates for uncommon severe health states as done in previous studies.[95,96,98] Consequently, mapping without a constant was selected as the anchoring method to construct the PedsUtil scoring system for all age groups.

Table 4.6 presents the provisional PedsUtil scoring system that was developed using pilot DCE and TTO data. Health utilities for all child health states defined by the PedsUtil health state classification system can be estimated by subtracting the relevant disutilities for each level in each dimension from 1. For example, the predicted health utility for a 10-year-old child living with health state 3434343 is:

$$U = 1 - 0.008 - 0.038 - 0.014 - 0.078 - 0.040 - 0.022 - 0.029 = 0.77 \quad \textbf{Eq. 4.5}$$

Table 4.6 PedsUtil Scoring System Across Age Groups

Dimension	Level	Utility Decrement				
		All Ages	2-4 years	5-7 years	8-12 years	13-18 years
Physical Functioning	**Almost never**	0.000	-0.002	0.000	0.000	-0.012
	Sometimes	-0.008	-0.030	0.000	-0.008	-0.012
	Often	-0.017	-0.030	-0.013	-0.026	-0.013
	Almost always	-0.035	-0.051	-0.027	-0.026	-0.050
Pain	**Almost never**	0.000	0.000	-0.007	0.000	0.000
	Sometimes	-0.019	-0.012	-0.027	-0.016	-0.028
	Often	-0.033	-0.026	-0.040	-0.038	-0.035
	Almost always	-0.045	-0.046	-0.056	-0.044	-0.038
Fatigue	**Almost never**	0.000	0.000	0.000	0.000	-0.013
	Sometimes	-0.016	0.000	-0.021	-0.014	-0.024
	Often	-0.026	0.000	-0.044	-0.024	-0.031
	Almost always	-0.041	-0.026	-0.052	-0.025	-0.063
Emotional Functioning	**Almost never**	-0.009	0.000	-0.009	-0.023	-0.013
	Sometimes	-0.024	-0.027	-0.031	-0.023	-0.013
	Often	-0.043	-0.027	-0.032	-0.078	-0.047
	Almost always	-0.061	-0.052	-0.053	-0.084	-0.065
Social Functioning	**Almost never**	-0.015	-0.022	-0.007	-0.015	-0.019
	Sometimes	-0.035	-0.033	-0.025	-0.040	-0.042
	Often	-0.053	-0.033	-0.045	-0.065	-0.068
	Almost always	-0.062	-0.060	-0.056	-0.069	-0.068
School Functioning	**Almost never**	0.000	0.000	-0.015	0.000	0.000
	Sometimes	-0.025	-0.034	-0.030	-0.015	-0.026
	Often	-0.032	-0.043	-0.047	-0.022	-0.026
	Almost always	-0.045	-0.043	-0.059	-0.043	-0.049
School Absence	**Almost never**	-0.005	-0.004	-0.008	0.000	-0.013
	Sometimes	-0.018	-0.017	-0.012	-0.029	-0.017
	Often	-0.042	-0.051	-0.023	-0.057	-0.037
	Almost always	-0.042	-0.051	-0.023	-0.062	-0.037

Discussion

Pilot data on public preferences for child health states defined by the PedsUtil health state classification system were collected in this study in order to construct the provisional PedsUtil scoring system. This study is the first to develop a value set for the PedsQL. The PedsUtil scoring system is the first scoring system that can be used across multiple pediatric age groups and is one of the few scoring systems created from US preferences, thus is more appropriate to use when evaluating US-based policies.

Overall, applying the DCE valuation protocol successfully generated preference weights that exhibited face validity. The valuation protocol reduced the cognitive burden of completing the DCE tasks by limiting the number of dimensions that differed between health states. Allowing only some dimensions to differ across choice sets also has the added benefit of reducing attribute non-attendance. For example, respondents who employ heuristics, such as weighing only a single dimension, are required to trade off between other dimensions. It should be noted, however, that this study observed a slightly higher proportion (28%) of respondents who failed the dominance test compared to previously published child health DCE studies, which have reported violations of dominance up to 20%.[91,93-96,171] The difference in response patterns may be due to differences in methodology and/or study sample. For example, some studies have implemented less stringent dominance tests by incorporating multiple dominance tasks and requiring respondents to fail more than once in order to be excluded from data analysis.[91,93,94,171] Future iterations of this DCE survey protocol should explore implementing alternative dominance tests. Nevertheless, the resulting latent DCE model coefficients in this study demonstrated face validity. The model coefficients had the expected negative sign and were generally monotonic with increasing severity. There were more inconsistencies in model coefficients for age-specific models, but this was expected due to the limited sample size of age subgroups.

On the other hand, applying the TTO valuation protocol produced health utility values that were much higher than expected with a narrow range of values. Other studies that have elicited preferences for child health states have reported mean TTO health utility values of -0.69[91] to 0.20[92] for the PITS state. In comparison, this study found a mean TTO health utility value of 0.70 for the PITS state. There are a multitude of possible explanations for this large discrepancy in values. For example, the other child health valuation studies were all conducted in non-US settings, thus the differences in values may reflect cultural differences. US adults may be much more unwilling to trade off life years from a child's life compared to adults from other countries, though cultural differences most likely do not account for the full magnitude of differences in PITS state values. Moreover, the PITS state as defined by the PedsUtil health state classification system may not be comparable to the PITS state defined by other health descriptive systems, such as the EQ-5D-Y, which may arguably describe a worse PITS state, resulting in lower health utility values. Alternatively, the discrepancy in PITS state values may be an artifact of the study design or method. For example, this study administered the TTO survey online, while many previous child health valuation studies were conducted as face-to-face interviews. However, face-to-face interviews were not feasible to implement in this study, especially since nationally representative samples are required for the large-scale TTO survey. Previous research has shown that mode of administration may influence TTO valuations,[172,173] but more research is needed to determine the magnitude of this effect. Another reason for the discrepancy may be because the EQ-5D-Y valuation studies used composite TTO,[91-99] while this study implemented a modified TTO approach. Respondents may be more reluctant to trade off life years of a child using the modified TTO approach, which would bias the results. In fact, this study found an inconsistency in the pattern of VAS scores and TTO values that illustrates the unwillingness of respondents to trade off time, especially for younger age groups. Specifically, the PITS state had the lowest VAS scores for younger age groups but also had the highest TTO health utility values. Additional valuation surveys that employ different TTO methods are currently being developed and will be administered to determine the validity of the TTO values obtained from this study.

Few studies have measured whether preferences for HRQoL differ by age of the affected individual. The results from this study indicate the possibility that the age of the child considered in valuation tasks impacts resulting health state values. For example, this study found that the relative importance of dimensions varied by age group. Social Functioning was the most important dimension for age groups 2-4 years and 13-18 years, School Functioning for age group 5-7 years, and Emotional Functioning for age group 8-12 years. Interestingly, the findings from this study somewhat align with those of the EQ-5D-Y studies, which found Pain and Emotional Functioning (i.e., worried, sad, or unhappy) to be the most important dimensions when preferences for child health states were elicited in the context of a hypothetical 10-year-old child.[91-99] However, Pain was not weighted heavily in any age group in this study, which is in contrast to these previous child health valuation studies. In terms of TTO values, respondents in this study were generally less willing to trade off time for younger children across all health states. The results from this study, however, should be interpreted with caution due to the limited sample size, which precluded the inclusion of tests of significance for differences in DCE and TTO values across age groups. Further research is planned to field large-scale nationally-representative DCE and TTO surveys, which will provide additional data to better characterize the differences in health state values that respondents may attribute to children of different ages.[‡‡] Of note, the findings from this study are in contrast to some previous studies that have examined the effect of including age as an attribute in valuation exercises. For example, Ramos-Goñi et al.[171] found that varying the age of the child imagined in DCE tasks had minimal impact on elicited values. A TTO study by Prosser et al.[163] on health state preferences for influenza found that respondents were *more* willing to trade off time for younger children. The diverging results between this study and previous studies may be due to numerous methodological differences (e.g., variations in the health descriptive systems used, perspective and framing of choice tasks, valuation protocol, etc.), thus more research is warranted to establish the effect of age in health state descriptions on resulting health state valuations. Nonetheless, the results from this study highlight the potential importance of considering age-related differences when deriving health state

[‡‡] To measure differences in responses across age groups, TTO data from the large-scale national survey will be analyzed using methods such as analysis of variance and the Kruskal-Wallis test. The effect of age on DCE values will be modeled by adding age as a covariate, as well as exploring possible interactions between age and dimension levels.

values, which has significant implications for conducting cost-effectiveness analyses and for making resource allocation decisions.

The provisional PedsUtil scoring system was estimated using a conditional logistic regression model and mapping the latent DCE model coefficients onto the QALY scale. However, there are several limitations to the conditional logistic regression model. For example, the model does not account for scale heterogeneity. Instead, the model assumes that the variance of utility over different choice questions is constant across respondents, which can result in biased preference weights.[151,174] In addition, the conditional logistic regression model does not take into consideration systematic variations in preferences across respondents. Failing to account for preference heterogeneity may also lead to biased estimates.[151] The small sample size of this study rendered the use of alternative modeling techniques infeasible. Therefore, future research will analyze data collected from the large-scale national surveys using additional statistical models that can address some of the shortcomings of the conditional logistic regression model, such as the random-parameters logistic regression model (i.e., mixed logit). The best fitting model will be selected to construct the final PedsUtil scoring system. Future research will also incorporate sensitivity analyses to test the robustness of model results, such as excluding respondents who indicate that they are not confident in their answers to the valuation tasks or excluding non-discriminant TTO responses (i.e., non-zero TTO values that are identical for all health states).

As stated previously, there exist multiple techniques for anchoring latent DCE preference weights. The potential differences in resulting scoring systems based on the anchoring approach selected emphasizes the importance of this decision. However, there are no established guidelines or criteria for choosing between anchoring methods. Published studies have used various criteria such as feasibility, prediction accuracy, theoretical coherence, and empirical consistency.[69,168] This study selected to anchor the latent DCE model coefficients using a linear mapping function without a constant based mostly on judgment and theoretical considerations. In this study, mapping with a constant was found to have the best alignment between the resulting value set and observed TTO data. However, using this anchoring approach generated a value set with a limited range that did not conform to the full 0-1 QALY scale due to the inclusion of a constant. Specifically, the inclusion of a constant resulted in a value set where full health (or health state with no problems in any dimension) was equal to a health utility value of 0.88, on average. A value set with such a compressed QALY scale would significantly impact resource allocation decisions as it may result in children not being able to gain as many QALYs from interventions. An alternative way to interpret the constant is the utility loss associated with any deviation from full health. This would allow for the health utility value for full health to be equal to 1 since the constant is not applicable to this health state. Previously published adult value sets with constants have adopted this interpretation to justify the inclusion of a constant.[95,175-177] However, this can result in value sets with a large gap between full health and the second-best health state.[178] Such concerns around the limited range and distribution of the resulting PedsUtil value set most influenced the choice of anchoring method in this study, but more research is needed to determine how to weigh multiple decision criteria in order to identify the most appropriate method to anchor DCE values for child health states.

There are several other limitations to this study. The DCE was designed so that the 3 overlap dimensions between the 2 health states were equal to the mildest severity level ("Never") as it was presumed to reduce respondent burden. However, this limits the range of severity for the health states included in the DCE. More research is needed to understand how this approach may affect preference weights. In addition, the valuation protocol for the TTO survey included only a small number of health states in order to minimize respondent fatigue. However, using a small design limits the ability to accurately characterize the relationship between DCE and TTO values when mapping the latent DCE model coefficients onto the QALY scale. Future research should explore the impact of increasing the number of health states valued on estimating mapping functions. The TTO valuation protocol also assumed a fixed time horizon of 10 years. This time horizon was chosen to limit valuation tasks to the childhood years, but this decision was somewhat arbitrary as there is no empirical evidence on how time horizon influences child health TTO values. Relatedly, the TTO valuation protocol did not allow for health states to be considered worse than dead as it was assumed that respondents rarely assigned negative values to child health states, and because it would introduce even more complexity to the TTO task. Further research is warranted to test alternative time horizons and the feasibility of including worse than dead states to value child health states. Another potential limitation of this study is that the use of an online panel is subject selection bias. Likewise, there are concerns about whether participants properly engage with and fully understand valuation tasks when surveys are administered online. This study attempted to address this issue by including data quality control criteria, but respondent engagement and understanding are difficult to accurately measure. Nevertheless, other modes of survey administration also face numerous limitations,[46] such as interviewer effects, and using an online survey allows for wide geographic representation. Lastly, this study elicited preferences for child health states from adults, rather than from children/adolescents. Current guidelines favor the use of adult community preferences,[1] but there is increasing interest in eliciting preferences directly from adolescents. Research has shown that there are systematic differences between adult and child preferences for identical health states,[64-66] and there is evidence that adolescents are able to provide valid and reliable preferences.[63] Therefore, future research should derive an adolescent-specific PedsUtil scoring system in order better understand the consequences of choosing one set of preferences over the other.

In conclusion, this paper developed and applied a valuation protocol to value child health states defined by the PedsUtil health state classification system in order to estimate preference weights for the preliminary construction of the PedsUtil scoring system. There remain numerous analytical considerations that require further research, thus the valuation protocol and resulting scoring system are subject to future improvements. Nevertheless, the PedsUtil scoring system is the first preference-based HRQoL measure to be able to estimate health utilities for a full range of ages in children from 2-18 years. The PedsUtil scoring system provides an essential tool for researchers and analysts conducting economic evaluations of child health interventions to accurately and consistently value child health outcomes.

Appendix

aFigure 4.1 PedsUtil Health State Classification System[a]

Physical Functioning
Never has problems with participating in sports activity or exercise
Almost never has problems with participating in sports activity or exercise
Sometimes has problems with participating in sports activity or exercise
Often has problems with participating in sports activity or exercise
Almost always has problems with participating in sports activity or exercise

Pain
Never has problems with having hurts or aches
Almost never has problems with having hurts or aches
Sometimes has problems with having hurts or aches
Often has problems with having hurts or aches
Almost always has problems with having hurts or aches

Fatigue
Never has problems with low energy level
Almost never has problems with low energy level
Sometimes has problems with low energy level
Often has problems with low energy level
Almost always has problems with low energy level

Emotional Functioning
Never has problems with worrying about what will happen to them
Almost never has problems with worrying about what will happen to them
Sometimes has problems with worrying about what will happen to them
Often has problems with worrying about what will happen to them
Almost always has problems with worrying about what will happen to them

Social Functioning
Never has problems with other kids not wanting to be their friend
Almost never has problems with other kids not wanting to be their friend
Sometimes has problems with other kids not wanting to be their friend
Often has problems with other kids not wanting to be their friend
Almost always has problems with other kids not wanting to be their friend

School Functioning
Never has problems with keeping up with schoolwork
Almost never has problems with keeping up with schoolwork
Sometimes has problems with keeping up with schoolwork
Often has problems with keeping up with schoolwork
Almost always has problems with keeping up with schoolwork

School Absence
Never has problems with missing school because of not feeling well
Almost never has problems with missing school because of not feeling well
Sometimes has problems with missing school because of not feeling well
Often has problems with missing school because of not feeling well
Almost always has problems with missing school because of not feeling well

[a] Wording for PedsUtil health state classification system differs slightly between age groups but items selected are the same across all age groups. PedsUtil health state classification system for age group 8-12 years shown in this table.

aTable 4.1 4-Dimension Base Design[a]

Choice Set	Alternative	Dimensions			
		X1	X2	X3	X4
1	1	Never	Almost never	Almost always	Often
1	2	Almost always	Never	Never	Almost never
2	1	Never	Almost always	Often	Never
2	2	Sometimes	Often	Almost always	Sometimes
3	1	Almost always	Almost always	Often	Almost never
3	2	Never	Never	Almost always	Never
4	1	Never	Sometimes	Sometimes	Sometimes
4	2	Almost never	Almost never	Almost always	Almost always
5	1	Almost never	Often	Sometimes	Often
5	2	Almost always	Sometimes	Almost always	Almost always
6	1	Often	Almost never	Often	Almost never
6	2	Almost never	Almost always	Never	Almost always
7	1	Almost never	Almost always	Almost never	Almost never
7	2	Sometimes	Sometimes	Never	Often
8	1	Almost always	Almost always	Sometimes	Sometimes
8	2	Often	Never	Almost never	Often
9	1	Never	Almost always	Almost always	Almost always
9	2	Almost always	Almost never	Almost never	Often
10	1	Often	Almost never	Almost always	Sometimes
10	2	Almost never	Sometimes	Almost never	Never
11	1	Almost never	Often	Often	Almost never
11	2	Sometimes	Almost always	Sometimes	Sometimes
12	1	Often	Almost always	Almost never	Almost always
12	2	Never	Often	Almost always	Sometimes
13	1	Almost never	Almost never	Almost never	Sometimes
13	2	Almost always	Often	Often	Never
14	1	Almost always	Often	Almost never	Sometimes
14	2	Often	Almost never	Often	Never
15	1	Almost always	Almost never	Sometimes	Never
15	2	Sometimes	Almost always	Often	Almost never
16	1	Almost never	Sometimes	Never	Often
16	2	Sometimes	Often	Almost never	Never
17	1	Often	Almost always	Sometimes	Never
17	2	Almost never	Often	Never	Almost never
18	1	Almost never	Almost never	Never	Never
18	2	Never	Never	Almost never	Almost never
19	1	Almost always	Sometimes	Often	Never
19	2	Sometimes	Almost always	Sometimes	Often
20	1	Almost always	Almost never	Often	Often
20	2	Almost never	Sometimes	Almost always	Never
21	1	Often	Almost never	Never	Almost always
21	2	Almost always	Never	Almost always	Often
22	1	Never	Often	Never	Almost never
22	2	Almost never	Sometimes	Often	Sometimes
23	1	Never	Almost never	Almost always	Almost never
23	2	Sometimes	Often	Often	Almost always
24	1	Almost never	Never	Sometimes	Never
24	2	Never	Sometimes	Almost never	Often
25	1	Often	Never	Often	Often
25	2	Never	Almost never	Almost never	Almost never
26	1	Almost always	Often	Almost always	Never
26	2	Never	Never	Never	Sometimes
27	1	Often	Sometimes	Never	Almost never
27	2	Never	Never	Often	Almost always
28	1	Almost always	Sometimes	Sometimes	Almost always

28	2	Almost never	Never	Often	Sometimes
29	1	Almost never	Almost always	Almost always	Often
29	2	Never	Never	Never	Almost always
30	1	Sometimes	Often	Often	Sometimes
30	2	Almost always	Never	Never	Often
31	1	Almost always	Almost never	Often	Sometimes
31	2	Often	Often	Sometimes	Often
32	1	Sometimes	Almost never	Never	Almost always
32	2	Often	Sometimes	Almost never	Sometimes
33	1	Sometimes	Almost always	Almost never	Never
33	2	Often	Often	Sometimes	Almost always
34	1	Often	Almost always	Almost always	Almost never
34	2	Sometimes	Almost never	Almost never	Often
35	1	Almost always	Almost always	Never	Sometimes
35	2	Sometimes	Sometimes	Sometimes	Almost never
36	1	Often	Never	Almost never	Sometimes
36	2	Never	Almost always	Often	Often
37	1	Sometimes	Never	Sometimes	Almost never
37	2	Often	Often	Almost never	Almost always
38	1	Often	Almost always	Never	Never
38	2	Almost never	Sometimes	Sometimes	Almost always
39	1	Sometimes	Never	Almost always	Almost always
39	2	Never	Often	Sometimes	Sometimes
40	1	Sometimes	Sometimes	Never	Never
40	2	Almost always	Never	Sometimes	Almost always

[a] Specifies the levels of the 4 dimensions that differ between the 2 health states in the DCE tasks.

aTable 4.2 Combinations of Varying Dimensions[a]

Choice Set	Health State 1							Health State 2						
	Phys	Pain	Fatigue	Emot	Soc	School	SchAbs	Phys	Pain	Fatigue	Emot	Soc	School	SchAbs
1	a	b		c			d	w	x		y			z
2	d	a	b		c			z	w	x		y		
3		d	a	b		c			z	w	x		y	
4			d	a	b		c			z	w	x		y
5	c			d	a	b		y			z	w	x	
6		c			d	a	b		y			z	w	x
7	b		c			d	a	x		y			z	w

Abbreviations: Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.

[a] This table identifies which 4 out of the 7 dimensions to vary between the 2 health states in the DCE tasks. 7 choice sets (out of 35 possible combinations) were selected so that each pair of dimensions in a health state was varied twice. For example, Physical Functioning and Pain vary in choice sets 1 and 2, Pain and Fatigue vary in choice sets 2 and 3, etc. Each choice set from aTable 4.1 was applied to all 7 combinations in this table. The empty cells in this table indicate the dimensions that are the same (i.e., “Never”) between the 2 health states.

aTable 4.3 Final DCE Experimental Design[a]

Choice Set	Alt	Phys	Pain	Fatigue	Emot	Soc	School	SchAbs
Block 1[b]								
1	1	Never	Almost never	Never	Almost always	Never	Never	Often
1	2	Almost always	Never	Never	Never	Never	Never	Almost never
2	1	Never	Almost always	Never	Often	Never	Never	Never
2	2	Sometimes	Often	Never	Almost always	Never	Never	Sometimes
3	1	Never	Never	Almost always	Never	Often	Never	Never
3	2	Sometimes	Sometimes	Often	Never	Almost always	Never	Never
4	1	Never	Almost never	Almost always	Almost always	Never	Often	Never
4	2	Never	Never	Never	Never	Never	Almost always	Never
5	1	Never	Never	Sometimes	Never	Sometimes	Never	Sometimes
5	2	Never	Never	Almost always	Almost never	Almost never	Never	Almost always
6	1	Sometimes	Never	Never	Often	Almost never	Often	Never
6	2	Almost always	Never	Never	Almost always	Almost always	Sometimes	Never
7	1	Never	Often	Never	Never	Almost never	Often	Almost never
7	2	Never	Never	Never	Never	Almost always	Almost never	Almost always
8	1	Almost always	Never	Almost never	Never	Never	Almost never	Almost never
8	2	Sometimes	Never	Never	Never	Never	Often	Sometimes
Block 2								
9	1	Often	Never	Almost never	Never	Almost always	Never	Never
9	2	Almost never	Almost always	Never	Never	Never	Never	Never
10	1	Never	Never	Never	Almost always	Never	Often	Never
10	2	Never	Sometimes	Sometimes	Often	Never	Almost always	Never
11	1	Almost never	Almost always	Almost always	Never	Often	Never	Never
11	2	Never	Never	Never	Never	Almost always	Never	Never
12	1	Never	Never	Almost never	Almost always	Almost always	Never	Often
12	2	Never	Never	Never	Never	Never	Never	Almost always
13	1	Sometimes	Never	Never	Sometimes	Never	Sometimes	Never
13	2	Almost always	Never	Never	Almost always	Almost never	Almost never	Never
14	1	Never	Sometimes	Never	Never	Often	Almost never	Often

14	2	Never	Almost always	Never	Never	Almost always	Almost always	Sometimes
15	1	Almost never	Never	Often	Never	Never	Almost never	Often
15	2	Almost always	Never	Never	Never	Never	Almost always	Almost never
16	1	Almost always	Almost always	Never	Sometimes	Never	Never	Sometimes
16	2	Often	Never	Never	Almost never	Never	Never	Often
Block 3								
17	1	Never	Often	Never	Almost never	Never	Almost always	Never
17	2	Never	Almost never	Almost always	Never	Never	Never	Never
18	1	Never	Never	Never	Never	Almost always	Never	Often
18	2	Never	Never	Sometimes	Sometimes	Often	Never	Almost always
19	1	Often	Never	Never	Almost never	Almost always	Almost always	Never
19	2	Almost always	Never	Never	Never	Never	Never	Never
20	1	Never	Sometimes	Never	Sometimes	Never	Sometimes	Never
20	2	Never	Almost always	Almost never	Almost never	Never	Almost always	Never
21	1	Never	Sometimes	Never	Never	Sometimes	Never	Sometimes
21	2	Never	Almost always	Never	Never	Almost always	Almost never	Almost never
22	1	Often	Never	Sometimes	Never	Never	Often	Almost never
22	2	Sometimes	Never	Almost always	Never	Never	Almost always	Almost always
23	1	Almost never	Almost always	Never	Almost never	Never	Never	Almost never
23	2	Sometimes	Sometimes	Never	Never	Never	Never	Often
24	1	Sometimes	Almost always	Almost always	Never	Sometimes	Never	Never
24	2	Often	Often	Never	Never	Almost never	Never	Never
Block 4								
25	1	Never	Never	Often	Never	Almost never	Never	Almost always
25	2	Never	Never	Almost never	Almost always	Never	Never	Never
26	1	Often	Never	Never	Never	Never	Almost always	Never
26	2	Almost always	Never	Never	Sometimes	Sometimes	Often	Never
27	1	Never	Often	Never	Never	Almost never	Almost always	Almost always
27	2	Never	Almost always	Never	Never	Never	Never	Never
28	1	Sometimes	Never	Sometimes	Never	Never	Sometimes	Never
28	2	Almost never	Never	Almost always	Never	Never	Almost always	Almost never
29	1	Never	Never	Often	Almost never	Often	Never	Sometimes

29	2	Never	Never	Almost always	Almost always	Sometimes	Never	Almost always
30	1	Often	Almost never	Never	Often	Never	Never	Almost never
30	2	Almost never	Almost always	Never	Never	Never	Never	Almost always
31	1	Almost never	Almost never	Almost always	Never	Almost never	Never	Never
31	2	Often	Sometimes	Sometimes	Never	Never	Never	Never
32	1	Never	Sometimes	Almost always	Almost always	Never	Sometimes	Never
32	2	Never	Often	Often	Never	Never	Almost never	Never
Block 5								
33	1	Almost always	Never	Never	Often	Never	Almost never	Never
33	2	Never	Never	Never	Almost never	Almost always	Never	Never
34	1	Never	Often	Never	Never	Never	Never	Almost always
34	2	Never	Almost always	Never	Never	Sometimes	Sometimes	Often
35	1	Almost always	Never	Often	Never	Never	Almost never	Almost always
35	2	Never	Never	Almost always	Never	Never	Never	Never
36	1	Almost never	Often	Never	Sometimes	Never	Never	Often
36	2	Almost always	Sometimes	Never	Almost always	Never	Never	Almost always
37	1	Almost never	Often	Almost never	Never	Often	Never	Never
37	2	Almost always	Almost never	Almost always	Never	Never	Never	Never
38	1	Often	Never	Never	Almost never	Often	Almost never	Never
38	2	Never	Never	Never	Almost always	Almost never	Almost always	Never
39	1	Never	Almost never	Almost never	Almost always	Never	Almost never	Never
39	2	Never	Often	Sometimes	Sometimes	Never	Never	Never
40	1	Never	Never	Sometimes	Almost always	Almost always	Never	Sometimes
40	2	Never	Never	Often	Often	Never	Never	Almost never
Block 6								
41	1	Never	Almost always	Never	Never	Often	Never	Almost never
41	2	Never	Never	Never	Never	Almost never	Almost always	Never
42	1	Almost always	Never	Often	Never	Never	Never	Never
42	2	Often	Never	Almost always	Never	Never	Sometimes	Sometimes
43	1	Never	Sometimes	Never	Sometimes	Never	Never	Sometimes
43	2	Almost never	Almost never	Never	Almost always	Never	Never	Almost always

44	1	Often	Almost never	Often	Never	Sometimes	Never	Never
44	2	Almost always	Almost always	Sometimes	Never	Almost always	Never	Never
45	1	Never	Almost never	Often	Almost never	Never	Often	Never
45	2	Never	Almost always	Almost never	Almost always	Never	Never	Never
46	1	Never	Never	Almost never	Almost never	Almost always	Never	Almost never
46	2	Never	Never	Often	Sometimes	Sometimes	Never	Never
47	1	Never	Almost never	Never	Never	Almost never	Almost never	Almost always
47	2	Never	Never	Never	Never	Often	Sometimes	Sometimes
48	1	Sometimes	Never	Never	Sometimes	Almost always	Almost always	Never
48	2	Almost never	Never	Never	Often	Often	Never	Never
Block 7								
49	1	Almost never	Never	Almost always	Never	Never	Often	Never
49	2	Never	Never	Never	Never	Never	Almost never	Almost always
50	1	Almost always	Almost always	Never	Often	Never	Never	Almost never
50	2	Never	Never	Never	Almost always	Never	Never	Never
51	1	Sometimes	Never	Sometimes	Never	Sometimes	Never	Never
51	2	Almost always	Almost never	Almost never	Never	Almost always	Never	Never
52	1	Never	Often	Almost never	Often	Never	Sometimes	Never
52	2	Never	Almost always	Almost always	Sometimes	Never	Almost always	Never
53	1	Never	Never	Almost never	Often	Almost never	Never	Often
53	2	Never	Never	Almost always	Almost never	Almost always	Never	Never
54	1	Almost never	Never	Never	Almost never	Almost never	Almost always	Never
54	2	Never	Never	Never	Often	Sometimes	Sometimes	Never
55	1	Never	Sometimes	Never	Never	Sometimes	Almost always	Almost always
55	2	Never	Almost never	Never	Never	Often	Often	Never
56	1	Almost always	Never	Sometimes	Never	Never	Sometimes	Almost always
56	2	Never	Never	Almost never	Never	Never	Often	Often
Block 8								
57	1	Never	Almost always	Never	Almost always	Never	Never	Almost always
57	2	Almost always	Almost never	Never	Almost never	Never	Never	Often
58	1	Often	Almost never	Never	Almost always	Never	Never	Sometimes
58	2	Almost never	Sometimes	Never	Almost never	Never	Never	Never

59	1	Sometimes	Often	Almost never	Never	Almost always	Never	Never
59	2	Never	Almost never	Sometimes	Never	Almost never	Never	Never
60	1	Never	Almost never	Almost never	Often	Never	Often	Never
60	2	Never	Sometimes	Sometimes	Almost always	Never	Sometimes	Never
61	1	Never	Never	Almost always	Often	Almost always	Never	Almost never
61	2	Never	Never	Sometimes	Never	Often	Never	Almost always
62	1	Almost never	Never	Never	Sometimes	Almost never	Almost never	Never
62	2	Often	Never	Never	Never	Almost always	Often	Never
63	1	Never	Almost never	Never	Never	Sometimes	Almost always	Often
63	2	Never	Often	Never	Never	Never	Often	Almost never
64	1	Almost never	Never	Sometimes	Never	Never	Never	Almost always
64	2	Almost always	Never	Often	Never	Never	Almost never	Sometimes
Block 9								
65	1	Almost always	Never	Almost always	Never	Almost always	Never	Never
65	2	Often	Almost always	Almost never	Never	Almost never	Never	Never
66	1	Never	Sometimes	Often	Almost never	Never	Almost always	Never
66	2	Never	Never	Almost never	Sometimes	Never	Almost never	Never
67	1	Almost never	Almost never	Often	Never	Often	Never	Never
67	2	Sometimes	Sometimes	Almost always	Never	Sometimes	Never	Never
68	1	Never	Never	Almost never	Almost never	Often	Never	Often
68	2	Never	Never	Sometimes	Sometimes	Almost always	Never	Sometimes
69	1	Almost never	Never	Never	Almost always	Often	Almost always	Never
69	2	Almost always	Never	Never	Sometimes	Never	Often	Never
70	1	Never	Almost never	Never	Never	Sometimes	Almost never	Almost never
70	2	Never	Often	Never	Never	Never	Almost always	Often
71	1	Often	Never	Almost never	Never	Never	Sometimes	Almost always
71	2	Almost never	Never	Often	Never	Never	Never	Often
72	1	Almost never	Sometimes	Never	Never	Never	Never	Often
72	2	Sometimes	Often	Never	Almost never	Never	Never	Never
Block 10								
73	1	Never	Almost always	Never	Almost always	Never	Almost always	Never

73	2	Never	Often	Almost always	Almost never	Never	Almost never	Never
74	1	Never	Never	Sometimes	Often	Almost never	Never	Almost always
74	2	Never	Never	Never	Almost never	Sometimes	Never	Almost never
75	1	Often	Never	Never	Almost never	Almost never	Often	Never
75	2	Sometimes	Never	Never	Sometimes	Sometimes	Almost always	Never
76	1	Never	Almost always	Often	Almost always	Never	Almost never	Never
76	2	Never	Sometimes	Never	Often	Never	Almost always	Never
77	1	Never	Almost never	Never	Never	Almost always	Often	Almost always
77	2	Never	Almost always	Never	Never	Sometimes	Never	Often
78	1	Almost never	Never	Almost never	Never	Never	Sometimes	Almost never
78	2	Often	Never	Often	Never	Never	Never	Almost always
79	1	Almost always	Almost never	Never	Sometimes	Never	Never	Never
79	2	Sometimes	Almost always	Never	Often	Never	Never	Almost never
80	1	Often	Almost never	Sometimes	Never	Never	Never	Never
80	2	Never	Sometimes	Often	Never	Almost never	Never	Never
Block 11								
81	1	Never	Never	Almost always	Never	Almost always	Never	Almost always
81	2	Never	Never	Often	Almost always	Almost never	Never	Almost never
82	1	Almost always	Never	Never	Sometimes	Often	Almost never	Never
82	2	Almost never	Never	Never	Never	Almost never	Sometimes	Never
83	1	Never	Often	Never	Never	Almost never	Almost never	Often
83	2	Never	Sometimes	Never	Never	Sometimes	Sometimes	Almost always
84	1	Almost always	Never	Almost never	Never	Never	Almost always	Often
84	2	Often	Never	Almost always	Never	Never	Sometimes	Never
85	1	Never	Never	Sometimes	Almost never	Almost never	Never	Almost never
85	2	Never	Never	Never	Almost always	Often	Never	Often
86	1	Almost always	Often	Never	Almost never	Never	Never	Sometimes
86	2	Often	Almost never	Never	Often	Never	Never	Never
87	1	Never	Almost always	Almost never	Never	Sometimes	Never	Never
87	2	Almost never	Sometimes	Almost always	Never	Often	Never	Never

88	1	Never	Often	Almost never	Sometimes	Never	Never	Never
88	2	Never	Never	Sometimes	Often	Never	Almost never	Never
Block 12								
89	1	Almost always	Never	Never	Almost always	Never	Almost always	Never
89	2	Almost never	Never	Never	Often	Almost always	Almost never	Never
90	1	Never	Almost always	Never	Never	Sometimes	Often	Almost never
90	2	Never	Almost never	Never	Never	Never	Almost never	Sometimes
91	1	Often	Never	Often	Never	Never	Almost never	Almost never
91	2	Almost always	Never	Sometimes	Never	Never	Sometimes	Sometimes
92	1	Almost never	Almost never	Never	Almost never	Never	Never	Sometimes
92	2	Almost always	Often	Never	Often	Never	Never	Never
93	1	Sometimes	Almost always	Often	Never	Almost never	Never	Never
93	2	Never	Often	Almost never	Never	Often	Never	Never
94	1	Almost never	Never	Never	Sometimes	Almost always	Often	Never
94	2	Often	Never	Never	Never	Often	Almost never	Never
95	1	Never	Never	Almost always	Almost never	Never	Sometimes	Never
95	2	Never	Almost never	Sometimes	Almost always	Never	Often	Never
96	1	Never	Never	Often	Almost never	Sometimes	Never	Never
96	2	Never	Never	Never	Sometimes	Often	Never	Almost never
Block 13								
97	1	Never	Almost always	Never	Never	Almost always	Never	Almost always
97	2	Never	Almost never	Never	Never	Often	Almost always	Almost never
98	1	Almost never	Never	Almost always	Never	Never	Sometimes	Often
98	2	Sometimes	Never	Almost never	Never	Never	Never	Almost never
99	1	Often	Almost always	Never	Almost never	Never	Never	Almost always
99	2	Never	Often	Never	Almost always	Never	Never	Sometimes
100	1	Sometimes	Almost never	Almost never	Never	Almost never	Never	Never
100	2	Never	Almost always	Often	Never	Often	Never	Never
101	1	Never	Sometimes	Almost always	Often	Never	Almost never	Never
101	2	Never	Never	Often	Almost never	Never	Often	Never
102	1	Never	Never	Never	Almost always	Almost never	Never	Sometimes

102	2	Never	Never	Almost never	Sometimes	Almost always	Never	Often
103	1	Never	Sometimes	Never	Never	Never	Almost always	Almost never
103	2	Never	Often	Never	Never	Almost never	Sometimes	Almost always
104	1	Never	Never	Never	Often	Almost never	Sometimes	Never
104	2	Almost never	Never	Never	Never	Sometimes	Often	Never
Block 14								
105	1	Almost always	Never	Almost always	Never	Never	Almost always	Never
105	2	Almost never	Never	Almost never	Never	Never	Often	Almost always
106	1	Almost never	Often	Never	Often	Never	Never	Almost never
106	2	Sometimes	Almost always	Never	Sometimes	Never	Never	Sometimes
107	1	Almost always	Often	Almost always	Never	Almost never	Never	Never
107	2	Sometimes	Never	Often	Never	Almost always	Never	Never
108	1	Never	Sometimes	Almost never	Almost never	Never	Almost never	Never
108	2	Never	Never	Almost always	Often	Never	Often	Never
109	1	Never	Never	Sometimes	Almost always	Often	Never	Almost never
109	2	Never	Never	Never	Often	Almost never	Never	Often
110	1	Sometimes	Never	Never	Never	Almost always	Almost never	Never
110	2	Often	Never	Never	Almost never	Sometimes	Almost always	Never
111	1	Never	Never	Never	Never	Often	Almost never	Sometimes
111	2	Never	Almost never	Never	Never	Never	Sometimes	Often
112	1	Sometimes	Never	Never	Never	Never	Often	Almost never
112	2	Often	Never	Almost never	Never	Never	Never	Sometimes
Block 15								
113	1	Often	Almost always	Never	Sometimes	Never	Never	Never
113	2	Almost never	Often	Never	Never	Never	Never	Almost never
114	1	Almost never	Almost never	Never	Never	Never	Never	Never
114	2	Never	Never	Never	Almost never	Never	Never	Almost never
115	1	Never	Almost never	Almost never	Never	Never	Never	Never
115	2	Almost never	Never	Never	Never	Almost never	Never	Never
116	1	Never	Never	Almost always	Sometimes	Never	Often	Never
116	2	Never	Often	Sometimes	Almost always	Never	Sometimes	Never

117	1	Never	Never	Often	Almost always	Almost never	Never	Often
117	2	Never	Never	Never	Almost never	Sometimes	Never	Almost always
118	1	Never	Never	Never	Almost always	Often	Almost never	Never
118	2	Almost always	Never	Never	Often	Almost always	Never	Never
119	1	Never	Never	Never	Never	Almost never	Never	Often
119	2	Never	Often	Never	Never	Sometimes	Almost never	Sometimes
120	1	Almost never	Never	Almost always	Never	Never	Almost never	Never
120	2	Often	Never	Often	Never	Never	Almost always	Sometimes
Block 16								
121	1	Never	Often	Almost always	Never	Sometimes	Never	Never
121	2	Almost never	Almost never	Often	Never	Never	Never	Never
122	1	Never	Never	Almost never	Almost never	Never	Never	Never
122	2	Never	Almost never	Never	Never	Never	Almost never	Never
123	1	Never	Almost always	Sometimes	Never	Often	Never	Never
123	2	Often	Sometimes	Almost always	Never	Sometimes	Never	Never
124	1	Never	Never	Never	Almost always	Sometimes	Never	Often
124	2	Never	Never	Often	Sometimes	Almost always	Never	Sometimes
125	1	Often	Never	Never	Often	Almost always	Almost never	Never
125	2	Almost always	Never	Never	Never	Almost never	Sometimes	Never
126	1	Never	Never	Never	Never	Almost always	Often	Almost never
126	2	Never	Almost always	Never	Never	Often	Almost always	Never
127	1	Often	Never	Never	Never	Never	Almost never	Never
127	2	Sometimes	Never	Often	Never	Never	Sometimes	Almost never
128	1	Almost never	Never	Never	Sometimes	Never	Never	Never
128	2	Never	Sometimes	Never	Almost never	Never	Never	Often
Block 17								
129	1	Never	Never	Often	Almost always	Never	Sometimes	Never
129	2	Never	Almost never	Almost never	Often	Never	Never	Never
130	1	Never	Never	Never	Almost never	Almost never	Never	Never
130	2	Never	Never	Almost never	Never	Never	Never	Almost never
131	1	Often	Never	Never	Never	Almost always	Sometimes	Never

131	2	Sometimes	Never	Never	Often	Sometimes	Almost always	Never
132	1	Never	Often	Almost always	Almost never	Never	Often	Never
132	2	Never	Never	Almost never	Sometimes	Never	Almost always	Never
133	1	Never	Often	Never	Never	Often	Almost always	Almost never
133	2	Never	Almost always	Never	Never	Never	Almost never	Sometimes
134	1	Almost never	Never	Never	Never	Never	Almost always	Often
134	2	Never	Never	Almost always	Never	Never	Often	Almost always
135	1	Never	Almost never	Never	Almost always	Never	Never	Almost never
135	2	Sometimes	Often	Never	Often	Never	Never	Almost always
136	1	Never	Almost never	Never	Never	Sometimes	Never	Never
136	2	Often	Never	Sometimes	Never	Almost never	Never	Never
Block 18								
137	1	Never	Never	Never	Often	Almost always	Never	Sometimes
137	2	Never	Never	Almost never	Almost never	Often	Never	Never
138	1	Never	Never	Never	Never	Almost never	Almost never	Never
138	2	Almost never	Never	Never	Almost never	Never	Never	Never
139	1	Never	Often	Never	Never	Never	Almost always	Sometimes
139	2	Never	Sometimes	Never	Never	Often	Sometimes	Almost always
140	1	Almost never	Never	Often	Never	Never	Often	Almost always
140	2	Sometimes	Never	Almost always	Never	Never	Never	Almost never
141	1	Never	Never	Almost always	Often	Almost never	Never	Never
141	2	Never	Never	Often	Almost always	Never	Never	Almost always
142	1	Never	Often	Never	Never	Never	Never	Almost never
142	2	Almost never	Sometimes	Never	Often	Never	Never	Sometimes
143	1	Almost never	Never	Almost never	Never	Almost always	Never	Never
143	2	Almost always	Sometimes	Often	Never	Often	Never	Never
144	1	Never	Never	Almost never	Never	Never	Sometimes	Never
144	2	Never	Often	Never	Sometimes	Never	Almost never	Never
Block 19								
145	1	Sometimes	Never	Never	Never	Often	Almost always	Never
145	2	Never	Never	Never	Almost never	Almost never	Often	Never

146	1	Never	Never	Never	Never	Never	Almost never	Almost never
146	2	Never	Almost never	Never	Never	Almost never	Never	Never
147	1	Sometimes	Never	Often	Never	Never	Never	Almost always
147	2	Almost always	Never	Sometimes	Never	Never	Often	Sometimes
148	1	Often	Almost never	Never	Never	Never	Never	Almost always
148	2	Almost always	Never	Never	Almost always	Never	Never	Often
149	1	Almost never	Never	Often	Never	Never	Never	Never
149	2	Sometimes	Almost never	Sometimes	Never	Often	Never	Never
150	1	Never	Never	Never	Almost never	Never	Often	Never
150	2	Often	Never	Never	Sometimes	Almost never	Sometimes	Never
151	1	Never	Almost never	Never	Almost never	Never	Almost always	Never
151	2	Never	Almost always	Sometimes	Often	Never	Often	Never
152	1	Never	Never	Never	Almost never	Never	Never	Sometimes
152	2	Never	Never	Often	Never	Sometimes	Never	Almost never
Block 20								
153	1	Never	Sometimes	Never	Never	Never	Often	Almost always
153	2	Never	Never	Never	Never	Almost never	Almost never	Often
154	1	Almost never	Never	Never	Never	Never	Never	Almost never
154	2	Never	Never	Almost never	Never	Never	Almost never	Never
155	1	Almost always	Almost never	Never	Often	Never	Never	Often
155	2	Almost never	Sometimes	Never	Almost always	Never	Never	Never
156	1	Almost always	Often	Almost never	Never	Never	Never	Never
156	2	Often	Almost always	Never	Never	Almost always	Never	Never
157	1	Never	Almost never	Never	Often	Never	Never	Never
157	2	Never	Sometimes	Almost never	Sometimes	Never	Often	Never
158	1	Never	Never	Almost never	Never	Almost never	Never	Almost always
158	2	Never	Never	Almost always	Sometimes	Often	Never	Often
159	1	Never	Almost always	Never	Never	Almost never	Never	Almost never
159	2	Never	Often	Never	Never	Almost always	Sometimes	Often
160	1	Sometimes	Never	Never	Never	Almost never	Never	Never

160	2	Almost never	Never	Never	Often	Never	Sometimes	Never
Block 21								
161	1	Almost always	Never	Sometimes	Never	Never	Never	Often
161	2	Often	Never	Never	Never	Never	Almost never	Almost never
162	1	Almost always	Sometimes	Never	Often	Never	Never	Never
162	2	Sometimes	Almost always	Never	Sometimes	Never	Never	Often
163	1	Often	Almost always	Almost never	Never	Often	Never	Never
163	2	Never	Almost never	Sometimes	Never	Almost always	Never	Never
164	1	Never	Almost always	Often	Almost never	Never	Never	Never
164	2	Never	Often	Almost always	Never	Never	Almost always	Never
165	1	Never	Never	Almost never	Never	Often	Never	Never
165	2	Never	Never	Sometimes	Almost never	Sometimes	Never	Often
166	1	Almost always	Never	Never	Almost never	Never	Almost never	Never
166	2	Often	Never	Never	Almost always	Sometimes	Often	Never
167	1	Never	Sometimes	Never	Never	Never	Almost never	Never
167	2	Never	Almost never	Never	Never	Often	Never	Sometimes
168	1	Never	Never	Sometimes	Never	Never	Never	Almost never
168	2	Sometimes	Never	Almost never	Never	Never	Often	Never
Block 22								
169	1	Often	Never	Never	Often	Never	Never	Often
169	2	Never	Almost never	Never	Almost never	Never	Never	Almost never
170	1	Almost always	Often	Never	Almost always	Never	Never	Never
170	2	Never	Never	Never	Never	Never	Never	Sometimes
171	1	Never	Almost always	Often	Never	Almost always	Never	Never
171	2	Sometimes	Never	Never	Never	Never	Never	Never
172	1	Never	Almost never	Often	Sometimes	Never	Never	Never
172	2	Never	Almost always	Never	Never	Never	Often	Never
173	1	Never	Never	Almost always	Almost always	Sometimes	Never	Sometimes
173	2	Never	Never	Sometimes	Almost never	Never	Never	Often
174	1	Almost always	Never	Never	Often	Almost never	Almost always	Never
174	2	Never	Never	Never	Almost always	Never	Never	Never
175	1	Never	Often	Never	Never	Sometimes	Sometimes	Often

175	2	Never	Never	Never	Never	Often	Almost always	Never
176	1	Almost never	Never	Often	Never	Never	Sometimes	Almost always
176	2	Often	Never	Sometimes	Never	Never	Often	Often
Block 23								
177	1	Often	Often	Never	Never	Often	Never	Never
177	2	Almost never	Never	Almost never	Never	Almost never	Never	Never
178	1	Never	Never	Almost always	Often	Never	Almost always	Never
178	2	Never	Sometimes	Never	Never	Never	Never	Never
179	1	Almost never	Often	Sometimes	Never	Never	Never	Never
179	2	Almost always	Never	Never	Never	Often	Never	Never
180	1	Never	Never	Almost never	Often	Sometimes	Never	Never
180	2	Never	Never	Almost always	Never	Never	Never	Often
181	1	Sometimes	Never	Never	Almost always	Almost always	Sometimes	Never
181	2	Often	Never	Never	Sometimes	Almost never	Never	Never
182	1	Never	Almost always	Never	Never	Often	Almost never	Almost always
182	2	Never	Never	Never	Never	Almost always	Never	Never
183	1	Often	Never	Often	Never	Never	Sometimes	Sometimes
183	2	Never	Never	Never	Never	Never	Often	Almost always
184	1	Sometimes	Almost never	Never	Never	Never	Never	Almost always
184	2	Often	Sometimes	Never	Almost never	Never	Never	Sometimes
Block 24								
185	1	Never	Often	Often	Never	Never	Often	Never
185	2	Never	Almost never	Never	Almost never	Never	Almost never	Never
186	1	Never	Never	Never	Almost always	Often	Never	Almost always
186	2	Never	Never	Sometimes	Never	Never	Never	Never
187	1	Never	Never	Never	Almost never	Often	Sometimes	Never
187	2	Often	Never	Never	Almost always	Never	Never	Never
188	1	Never	Almost always	Almost always	Sometimes	Never	Sometimes	Never
188	2	Never	Sometimes	Almost never	Never	Never	Often	Never
189	1	Never	Sometimes	Never	Never	Almost always	Almost always	Sometimes
189	2	Never	Often	Never	Never	Sometimes	Almost never	Never
190	1	Almost always	Never	Almost always	Never	Never	Often	Almost never

190	2	Never	Never	Never	Never	Never	Almost always	Never
191	1	Almost always	Almost never	Never	Often	Never	Never	Sometimes
191	2	Often	Often	Never	Sometimes	Never	Never	Often
192	1	Almost always	Sometimes	Almost never	Never	Never	Never	Never
192	2	Sometimes	Often	Sometimes	Never	Almost never	Never	Never
Block 25								
193	1	Never	Never	Often	Often	Never	Never	Often
193	2	Never	Never	Almost never	Never	Almost never	Never	Almost never
194	1	Almost always	Never	Never	Never	Almost always	Often	Never
194	2	Never	Never	Never	Sometimes	Never	Never	Never
195	1	Never	Never	Never	Never	Almost never	Often	Sometimes
195	2	Never	Often	Never	Never	Almost always	Never	Never
196	1	Sometimes	Never	Sometimes	Never	Never	Almost always	Almost always
196	2	Never	Never	Often	Never	Never	Sometimes	Almost never
197	1	Never	Never	Often	Almost never	Almost always	Never	Almost always
197	2	Never	Never	Almost always	Never	Never	Never	Never
198	1	Sometimes	Often	Never	Often	Never	Never	Sometimes
198	2	Almost always	Never	Never	Never	Never	Never	Often
199	1	Sometimes	Almost always	Almost never	Never	Often	Never	Never
199	2	Often	Often	Often	Never	Sometimes	Never	Never
200	1	Never	Almost always	Sometimes	Almost never	Never	Never	Never
200	2	Never	Sometimes	Often	Sometimes	Never	Almost never	Never
Block 26								
201	1	Often	Never	Never	Often	Often	Never	Never
201	2	Almost never	Never	Never	Almost never	Never	Almost never	Never
202	1	Never	Almost always	Never	Never	Never	Almost always	Often
202	2	Never	Never	Never	Never	Sometimes	Never	Never
203	1	Sometimes	Never	Never	Never	Never	Almost never	Often
203	2	Never	Never	Often	Never	Never	Almost always	Never
204	1	Almost never	Almost always	Never	Almost always	Never	Never	Often
204	2	Never	Never	Never	Never	Never	Never	Almost always
205	1	Sometimes	Sometimes	Often	Never	Often	Never	Never
205	2	Often	Almost always	Never	Never	Never	Never	Never

206	1	Often	Never	Never	Sometimes	Sometimes	Often	Never
206	2	Never	Never	Never	Often	Almost always	Never	Never
207	1	Never	Sometimes	Almost always	Almost never	Never	Often	Never
207	2	Never	Often	Often	Often	Never	Sometimes	Never
208	1	Never	Never	Almost always	Sometimes	Almost never	Never	Never
208	2	Never	Never	Sometimes	Often	Sometimes	Never	Almost never
Block 27								
209	1	Never	Often	Never	Never	Often	Often	Never
209	2	Never	Almost never	Never	Never	Almost never	Never	Almost never
210	1	Often	Never	Almost always	Never	Never	Never	Almost always
210	2	Never	Never	Never	Never	Never	Sometimes	Never
211	1	Almost always	Sometimes	Never	Sometimes	Never	Never	Almost always
211	2	Almost never	Never	Never	Often	Never	Never	Sometimes
212	1	Often	Almost never	Almost always	Never	Almost always	Never	Never
212	2	Almost always	Never	Never	Never	Never	Never	Never
213	1	Never	Sometimes	Sometimes	Often	Never	Often	Never
213	2	Never	Often	Almost always	Never	Never	Never	Never
214	1	Never	Never	Sometimes	Almost always	Almost never	Never	Often
214	2	Never	Never	Often	Often	Often	Never	Sometimes
215	1	Never	Often	Never	Never	Sometimes	Almost always	Almost never
215	2	Never	Sometimes	Never	Never	Often	Often	Often
216	1	Never	Never	Never	Almost always	Sometimes	Almost never	Never
216	2	Almost never	Never	Never	Sometimes	Often	Sometimes	Never
Block 28								
217	1	Never	Never	Often	Never	Never	Often	Often
217	2	Almost never	Never	Almost never	Never	Never	Almost never	Never
218	1	Often	Sometimes	Never	Never	Never	Never	Almost never
218	2	Never	Never	Never	Often	Never	Never	Almost always
219	1	Almost always	Almost always	Sometimes	Never	Sometimes	Never	Never
219	2	Sometimes	Almost never	Never	Never	Often	Never	Never
220	1	Never	Often	Almost never	Almost always	Never	Almost always	Never
220	2	Never	Almost always	Never	Never	Never	Never	Never
221	1	Never	Never	Sometimes	Sometimes	Often	Never	Often

221	2	Never	Never	Often	Almost always	Never	Never	Never
222	1	Often	Never	Never	Sometimes	Almost always	Almost never	Never
222	2	Sometimes	Never	Never	Often	Often	Often	Never
223	1	Never	Never	Never	Never	Almost always	Sometimes	Almost never
223	2	Never	Almost never	Never	Never	Sometimes	Often	Sometimes
224	1	Almost never	Never	Never	Never	Never	Almost always	Sometimes
224	2	Sometimes	Never	Almost never	Never	Never	Sometimes	Often
Block 29								
225	1	Sometimes	Almost always	Never	Almost never	Never	Never	Never
225	2	Often	Often	Never	Sometimes	Never	Never	Almost always
226	1	Often	Almost always	Never	Almost always	Never	Never	Almost never
226	2	Sometimes	Almost never	Never	Almost never	Never	Never	Often
227	1	Almost never	Often	Almost always	Never	Almost always	Never	Never
227	2	Often	Sometimes	Almost never	Never	Almost never	Never	Never
228	1	Never	Sometimes	Almost always	Almost always	Never	Never	Never
228	2	Never	Almost never	Sometimes	Sometimes	Never	Sometimes	Never
229	1	Never	Never	Sometimes	Often	Never	Never	Almost never
229	2	Never	Never	Often	Never	Almost always	Never	Often
230	1	Sometimes	Never	Never	Almost never	Sometimes	Never	Never
230	2	Almost never	Never	Never	Almost always	Often	Often	Never
231	1	Never	Never	Never	Never	Never	Often	Almost always
231	2	Never	Sometimes	Never	Never	Almost always	Almost never	Sometimes
232	1	Never	Never	Almost always	Never	Never	Almost always	Sometimes
232	2	Often	Never	Sometimes	Never	Never	Sometimes	Never
Block 30								
233	1	Never	Sometimes	Almost always	Never	Almost never	Never	Never
233	2	Almost always	Often	Often	Never	Sometimes	Never	Never
234	1	Never	Almost never	Often	Almost always	Never	Almost always	Never
234	2	Never	Often	Sometimes	Almost never	Never	Almost never	Never
235	1	Sometimes	Almost always	Almost always	Never	Never	Never	Never
235	2	Almost never	Sometimes	Sometimes	Never	Sometimes	Never	Never

236	1	Never	Never	Sometimes	Almost always	Almost always	Never	Never
236	2	Never	Never	Almost never	Sometimes	Sometimes	Never	Sometimes
237	1	Almost never	Never	Never	Sometimes	Often	Never	Never
237	2	Often	Never	Never	Often	Never	Almost always	Never
238	1	Never	Sometimes	Never	Never	Almost never	Sometimes	Never
238	2	Never	Almost never	Never	Never	Almost always	Often	Often
239	1	Almost always	Never	Never	Never	Never	Never	Often
239	2	Sometimes	Never	Sometimes	Never	Never	Almost always	Almost never
240	1	Sometimes	Sometimes	Never	Never	Never	Never	Never
240	2	Almost always	Never	Never	Sometimes	Never	Never	Almost always
Block 31								
241	1	Never	Never	Sometimes	Almost always	Never	Almost never	Never
241	2	Never	Almost always	Often	Often	Never	Sometimes	Never
242	1	Never	Never	Almost never	Often	Almost always	Never	Almost always
242	2	Never	Never	Often	Sometimes	Almost never	Never	Almost never
243	1	Never	Never	Never	Sometimes	Almost always	Almost always	Never
243	2	Sometimes	Never	Never	Almost never	Sometimes	Sometimes	Never
244	1	Never	Sometimes	Often	Never	Never	Almost never	Never
244	2	Never	Often	Never	Almost always	Never	Often	Never
245	1	Never	Almost never	Never	Never	Sometimes	Often	Never
245	2	Never	Often	Never	Never	Often	Never	Almost always
246	1	Never	Never	Sometimes	Never	Never	Almost never	Sometimes
246	2	Often	Never	Almost never	Never	Never	Almost always	Often
247	1	Sometimes	Never	Never	Almost always	Never	Never	Almost always
247	2	Never	Often	Never	Sometimes	Never	Never	Sometimes
248	1	Never	Sometimes	Sometimes	Never	Never	Never	Never
248	2	Almost always	Almost always	Never	Never	Sometimes	Never	Never
Block 32								
249	1	Never	Never	Never	Sometimes	Almost always	Never	Almost never
249	2	Never	Never	Almost always	Often	Often	Never	Sometimes
250	1	Almost always	Never	Never	Almost never	Often	Almost always	Never

250	2	Almost never	Never	Never	Often	Sometimes	Almost never	Never
251	1	Never	Never	Never	Never	Sometimes	Almost always	Almost always
251	2	Never	Sometimes	Never	Never	Almost never	Sometimes	Sometimes
252	1	Never	Never	Almost never	Never	Never	Sometimes	Often
252	2	Almost always	Never	Often	Never	Never	Often	Never
253	1	Never	Never	Almost never	Sometimes	Never	Never	Sometimes
253	2	Never	Never	Almost always	Often	Often	Never	Almost never
254	1	Often	Almost always	Never	Never	Never	Never	Never
254	2	Almost never	Sometimes	Never	Sometimes	Never	Never	Almost always
255	1	Almost always	Sometimes	Never	Never	Almost always	Never	Never
255	2	Sometimes	Never	Often	Never	Sometimes	Never	Never
256	1	Never	Never	Sometimes	Sometimes	Never	Never	Never
256	2	Never	Almost always	Almost always	Never	Never	Sometimes	Never
Block 33								
257	1	Almost never	Never	Never	Never	Sometimes	Almost always	Never
257	2	Sometimes	Never	Never	Almost always	Often	Often	Never
258	1	Never	Almost always	Never	Never	Almost never	Often	Almost always
258	2	Never	Almost never	Never	Never	Often	Sometimes	Almost never
259	1	Almost always	Never	Never	Never	Never	Sometimes	Almost always
259	2	Sometimes	Never	Sometimes	Never	Never	Almost never	Sometimes
260	1	Sometimes	Never	Never	Sometimes	Never	Never	Almost never
260	2	Often	Often	Never	Almost never	Never	Never	Almost always
261	1	Never	Often	Almost always	Never	Never	Never	Never
261	2	Almost always	Almost never	Sometimes	Never	Sometimes	Never	Never
262	1	Never	Never	Never	Never	Often	Almost always	Never
262	2	Sometimes	Never	Never	Almost always	Almost never	Sometimes	Never
263	1	Never	Almost always	Sometimes	Never	Never	Almost always	Never
263	2	Never	Sometimes	Never	Often	Never	Sometimes	Never
264	1	Never	Never	Never	Sometimes	Sometimes	Never	Never
264	2	Never	Never	Almost always	Almost always	Never	Never	Sometimes
Block 34								
265	1	Never	Almost never	Never	Never	Never	Sometimes	Almost always

265	2	Never	Sometimes	Never	Never	Almost always	Often	Often
266	1	Almost always	Never	Almost always	Never	Never	Almost never	Often
266	2	Almost never	Never	Almost never	Never	Never	Often	Sometimes
267	1	Often	Never	Never	Almost never	Never	Never	Sometimes
267	2	Never	Almost always	Never	Often	Never	Never	Often
268	1	Almost never	Sometimes	Never	Never	Sometimes	Never	Never
268	2	Almost always	Often	Often	Never	Almost never	Never	Never
269	1	Never	Never	Often	Almost always	Never	Never	Never
269	2	Never	Almost always	Almost never	Sometimes	Never	Sometimes	Never
270	1	Never	Never	Almost always	Sometimes	Never	Never	Almost always
270	2	Never	Never	Sometimes	Never	Often	Never	Sometimes
271	1	Never	Almost always	Never	Never	Almost always	Sometimes	Never
271	2	Never	Sometimes	Never	Never	Sometimes	Never	Often
272	1	Never	Never	Never	Never	Sometimes	Sometimes	Never
272	2	Sometimes	Never	Never	Almost always	Almost always	Never	Never
Block 35								
273	1	Almost always	Never	Almost never	Never	Never	Never	Sometimes
273	2	Often	Never	Sometimes	Never	Never	Almost always	Often
274	1	Almost always	Almost always	Never	Never	Never	Never	Sometimes
274	2	Sometimes	Sometimes	Never	Sometimes	Never	Never	Almost never
275	1	Sometimes	Often	Never	Never	Almost never	Never	Never
275	2	Often	Never	Almost always	Never	Often	Never	Never
276	1	Never	Almost never	Sometimes	Never	Never	Sometimes	Never
276	2	Never	Almost always	Often	Often	Never	Almost never	Never
277	1	Never	Never	Never	Often	Almost always	Never	Never
277	2	Never	Never	Almost always	Almost never	Sometimes	Never	Sometimes
278	1	Almost always	Never	Never	Almost always	Sometimes	Never	Never
278	2	Sometimes	Never	Never	Sometimes	Never	Often	Never
279	1	Never	Never	Never	Never	Never	Sometimes	Sometimes
279	2	Never	Sometimes	Never	Never	Almost always	Almost always	Never
280	1	Sometimes	Never	Never	Never	Never	Never	Sometimes
280	2	Never	Never	Sometimes	Never	Never	Almost always	Almost always

Abbreviations: Alt, alternative; Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.
[a] Final DCE experimental design was a forced choice fractional factorial design with overlap in 3 dimensions.
[b] Blocks were designed so that each respondent saw different choice sets from the 4-dimension base design (aTable 4.1), as well as every combination in aTable 4.2.

aFigure 4.2 Health State Descriptions of Health States Valued in TTO Survey[a]

Health Description A:

Dimension	Health Description
Phys	**Sometimes** has problems with participating in sports activity or exercise
Pain	**Sometimes** has problems with having hurts or aches
Fatigue	**Sometimes** has problems with low energy level
Emot	**Sometimes** has problems with worrying about what will happen to them
Soc	**Sometimes** has problems with other kids not wanting to be their friend
School	**Sometimes** has problems with keeping up with schoolwork
SchAbs	**Sometimes** has problems with missing school because of not feeling well

Health Description B:

Dimension	Health Description
Phys	**Often** has problems with participating in sports activity or exercise
Pain	**Often** has problems with having hurts or aches
Fatigue	**Almost always** has problems with low energy level
Emot	**Sometimes** has problems with worrying about what will happen to them
Soc	**Sometimes** has problems with other kids not wanting to be their friend
School	**Almost always** has problems with keeping up with schoolwork
SchAbs	**Often** has problems with missing school because of not feeling well

Health Description C:

Dimension	Health Description
Phys	**Almost always** has problems with participating in sports activity or exercise
Pain	**Almost always** has problems with having hurts or aches
Fatigue	**Almost always** has problems with low energy level
Emot	**Almost always** has problems with worrying about what will happen to them
Soc	**Almost always** has problems with other kids not wanting to be their friend
School	**Almost always** has problems with keeping up with schoolwork
SchAbs	**Almost always** has problems with missing school because of not feeling well

Abbreviations: Phys, Physical Functioning; Emot, Emotional Functioning; Soc, Social Functioning; School, School Functioning; SchAbs, School Absence.

[a] Wording for health state descriptions differs slightly between age groups but dimensions and items are the same across all age groups. Health state descriptions for age group 8-12 years shown in this table.

aTable 4.4 Data Quality Indicators for DCE and TTO Surveys

Quality Indicator	DCE	TTO
Survey completion time		
Median	11.7 min	12.4 min
Mean	16.1 min	23.7 min
Speeders[a]		
n (%)[b]	6 (1.8)	0 (0)
Invalid age[c]		
n (%)[b]	8 (2.4)	4 (4.3)
Speeders & Invalid age		
n (%)[b]	6 (1.8)	1 (1.1)
Dominance violation[d]		
n (%)[b]	93 (28.3)	N/A
Invalid TTO amount[e]		
n (%)[b]	N/A	6 (6.4)
Inconsistent TTO[f]		
n (%)[g]	N/A	40 (49.4)
Level of confidence[h] [*n* (%)[g]]		
Very confident	83 (39.0)	42 (51.9)
Somewhat confident	110 (51.6)	26 (32.1)
Not confident	16 (7.5)	10 (12.3)
They were total guesses	4 (1.9)	3 (3.7)

Abbreviations: DCE, discrete choice experiment; TTO, time trade-off; min, minutes.

[a] Respondents who completed the survey in less than 1/3 median sample completion time.

[b] Denominator used to calculate % was total pilot sample size (*n* = 329 DCE; *n* = 94 TTO).

[c] Respondents whose answers did not match for all age questions and/or input nonsensical ages.

[d] Respondents who failed the DCE dominance test.

[e] Respondents who traded off more than 10 years (i.e., the maximum TTO amount).

[f] Respondents who traded off more time for a milder health state compared to a more severe health state.

[g] Denominator used to calculate % was final pilot sample size included in data analysis (*n* = 213 DCE; *n* = 81 TTO).

[h] Confidence question in surveys asked respondents: "How confident are you in your answers to the [DCE/TTO] questions?"

aTable 4.5 VAS Summary Statistics for the 3 Health State Across Age Groups

Health State	*n*[a]	Mean (SD)	Median	Range	% High-raters[b]
All Ages					
3333333	162	66.8 (22.7)	70.0	0-100	8.0
4453354	162	57.3 (27.3)	59.5	6-100	7.4
5555555	162	49.3 (32.0)	50.0	0-100	6.2
2-4 years					
3333333	42	70.2 (23.2)	73.5	0-100	11.9
4453354	42	58.5 (27.9)	67.0	6-100	4.8
5555555	42	47.4 (31.6)	41.0	0-100	4.8
5-7 years					
3333333	43	60.9 (23.3)	66.0	0-100	4.7
4453354	43	57.3 (26.9)	57.0	18-100	7.0
5555555	43	47.2 (31.6)	42.0	0-100	4.7
8-12 years					
3333333	38	69.2 (20.5)	70.5	32-100	10.5
4453354	38	57.1 (26.9)	50.0	6-100	10.5
5555555	38	51.4 (31.8)	50.0	0-100	7.9
13-18 years					
3333333	39	67.4 (23.2)	71.0	20-100	5.1
4453354	39	56.4 (28.4)	50.0	12-100	7.7
5555555	39	51.8 (33.8)	61.0	0-100	7.7

Abbreviations: SD, standard deviation.
[a] *n* refers to the number of observations.
[b] High-raters refer to respondents who rated the health state as 100.

aTable 4.6 Estimation Results for Anchoring on PITS and Mapping

	All Ages			2-4 years			5-7 years			8-12 years			13-18 years		
	PITS[a]	**Map A**[b]	**Map B**[c]	**PITS**	**Map A**	**Map B**	**PITS**	**Map A**	**Map B**	**PITS**	**Map A**	**Map B**	**PITS**	**Map A**	**Map B**
$\boldsymbol{\theta}$[d]	0.067	0.074	0.041	0.052	0.067	0.039	0.071	0.077	0.046	0.073	0.076	0.042	0.063	0.074	0.027
SE	–[e]	0.008	0.003	–	0.006	0.001	–	0.007	0.004	–	0.010	0.024	–	0.011	0.004
Inter.	0	0	-0.117	0	0	-0.109	0	0	-0.106	0	0	-0.124	0	0	-0.183
***p*-value**	–[e]	0.010	0.042	–	0.008	0.024	–	0.008	0.052	–	0.016	0.328	–	0.021	0.101
R^2	–[e]	0.980	0.996	–	0.984	0.999	–	0.984	0.993	–	0.968	0.758	–	0.958	0.975

Abbreviations: SE, standard error; Inter., intercept.

[a] PITS refers to anchoring on the mean observed TTO value of the PITS state (5555555).

[b] Map A refers to mapping without a constant.

[c] Map B refers to mapping with a constant.

[d] $\boldsymbol{\theta}$ Refers to the rescaling coefficient.

[e] Anchoring on the mean observed TTO value of the PITS state does not produce standard errors, *p*-values, and R^2 statistics.

Chapter 5: Conclusion

Summary

Understanding the economic value of health interventions is critical to their appropriate implementation. However, applying decision analytic and economic evaluation methods to value child health interventions presents challenges that are unique to pediatric populations. One key limitation is the lack of a child-specific preference-based HRQoL measure that encompasses multiple pediatric age groups. Current methods to value child HRQoL are highly variable, which has hindered decision-makers from making consistent and evidence-based policy recommendations. Therefore, the overall aim of this dissertation was to create a new preference-based HRQoL measure for children that can be applied across multiple pediatric age groups: the PedsUtil scoring system.

Chapter 2 described the first step of deriving the PedsUtil scoring system, which was to identify the core dimension structure of the PedsUtil health state classification system. Using data from the LSAC, 4 competing dimension structures were evaluated using confirmatory factor analysis. The findings from this study supported a 7-dimension structure (i.e., Physical Functioning, Pain, Fatigue, Emotional Functioning, Social Functioning, School Functioning, and School Absence) of the PedsUtil health state classification system. Furthermore, this analysis found that the 7-dimension structure was applicable across diverse pediatric populations, including both children with special healthcare needs and typically functioning children, as well as children aged 2-17 years.

Chapter 3 described the second step of deriving the PedsUtil scoring system, which was to use Rasch analysis alongside other psychometric testing and qualitative work in order to select the most representative item within each dimension of the PedsUtil health state classification system. Data for all secondary analyses came from the LSAC, and a convenience sample of pediatricians, clinical trialists, and parents were recruited to participate in expert panels and key informant interviews. After considering all Rasch and psychometric criteria, as well as input from expert panelists and key informant interviewees, this study reduced the full set of PedsQL items to a core set of 7 items in order to finalize the design of the PedsUtil health state classification system.

Finally, Chapter 4 developed and applied a valuation protocol to value child health states defined by the PedsUtil health state classification system using a DCE and a TTO survey. DCE pilot data were analyzed using a conditional logistic regression model and the resulting latent DCE model coefficients were mapped onto the mean observed TTO values using OLS regression to construct the PedsUtil scoring system. Separate preference weights were derived for 4 different pediatric age groups (i.e., 2-4 years, 5-7 years, 8-12 years, 13-18 years). This study produced the first provisional value set for the PedsQL, which provides an efficient approach to valuing health benefits across multiple pediatric age groups. The development of the PedsQL valuation protocol also allows other researchers to replicate this study in order to generate country-specific value sets. The PedsUtil scoring system ensures that children's experiences with disease and treatment are consistently and accurately represented in healthcare value assessments so that decision-makers are able to make strong, evidence-based decisions for healthcare priority setting.

Future Directions for Research

In addition to the suggestions for future research discussed in the previous chapters, there are several other key areas of research to expand upon in order to further develop the findings from this dissertation. For example, this dissertation found that health state valuations may differ by age of the affected child. Future research should explore possible explanations for why these differences may exist. One method to investigate possible reasons for these age-differences is to conduct a qualitative study that uses the think-aloud method while participants are completing child health valuation tasks for children of different ages. Such research would help interpret the quantitative results of child health valuation studies, as well as contribute to the development of recommendations for best practices in valuing child health.

Future research should also compare health utilities estimated from the PedsUtil scoring system to health utilities calculated from existing mapping algorithms, such as those developed to estimate CHU-9D health utilities[179-182] and EQ-5D-Y health utilities[183] using responses to the PedsQL. Assessing the comparability of health utilities obtained through different measurement processes would offer more insight into potential measurement biases and the implications of using specific measures to estimate health utilities for use in cost-effectiveness analyses.

Another important area for future research is to determine the smallest change in health utility values estimated from the PedsUtil scoring system that can be considered as important and meaningful. This minimally important difference (MID) should be compared across age groups, and further work should determine whether the MID is constant across various conditions and pediatric populations. This information will be helpful in establishing the clinical significance of changes in PedsUtil health utility values, and will be especially useful in interpreting treatment effects in pediatric clinical trials.

Lastly, this dissertation constructed the PedsUtil scoring system for children 2-18 years. Future research could expand upon this scoring system to include infants. Specifically, a value set for the PedsQL Infant Scales[184] could be developed in order to estimate health utilities for infants 0-2 years. The development of a scoring system for very young children would improve the quality of child health economic evidence, as well as help ensure equitable decision making for children of all ages.

Collectively, these future studies will advance the field of child health research in order to better inform decision-makers on the efficient and equitable implementation of child health interventions.

Bibliography

1. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. *Cost effectiveness in health and medicine.* Second edition. ed. Oxford; New York: Oxford University Press; 2017.

2. Drummond M. *Methods for the economic evaluation of health care programmes.* Fourth edition. ed. Oxford; New York: Oxford University Press; 2015.

3. Ungar WJ. *Economic evaluation in child health.* Oxford; New York: Oxford University Press; 2010.

4. Keren R, Pati S, Feudtner C. The generation gap: differences between children and adults pertinent to economic evaluations of health interventions. *Pharmacoeconomics.* 2004;22(2):71-81.

5. Simpson LA. The Adolescence of Child Health Services Research. *JAMA Pediatrics.* 2013;167(6):509-510.

6. Forrest CB, Simpson L, Clancy C. Child health services research. Challenges and opportunities. *Jama.* 1997;277(22):1787-1793.

7. Lipstein EA, Brinkman WB, Fiks AG, et al. An emerging field of research: challenges in pediatric decision making. *Med Decis Making.* 2015;35(3):403-408.

8. Sameroff AJ, Chandler MJ. Reproductive risk and the continuum of caretaking casualty. *Review of child development research.*4:187-244.

9. Conti G, Heckman JJ. The developmental approach to child and adult health. *Pediatrics.* 2013;131 Suppl 2(Suppl 2):S133-141.

10. Heck KE, Parker JD. Family structure, socioeconomic status, and access to health care for children. *Health Serv Res.* 2002;37(1):173-186.

11. Cafferata GL, Kasper JD. Family structure and children's use of ambulatory physician services. *Med Care.* 1985;23(4):350-360.

12. Porterfield SL, McBride TD. The effect of poverty and caregiver education on perceived need and access to health services among children with special health care needs. *Am J Public Health.* 2007;97(2):323-329.

13. Lebrun-Harris LA, Ghandour RM, Kogan MD, Warren MD. Five-Year Trends in US Children's Health and Well-being, 2016-2020. *JAMA Pediatrics.* 2022;176(7):e220056-e220056.

14. Dharmage SC, Perret JL, Custovic A. Epidemiology of Asthma in Children and Adults. *Front Pediatr.* 2019;7:246.

15. Centers for Disease Control and Prevention. Leading causes of death and injury Centers for Disease Control and Prevention. https://www.cdc.gov/injury/wisqars/LeadingCauses.html. Accessed February 10, 2023.

16. Sweeting H. Reversals of fortune? Sex differences in health in childhood and adolescence. *Soc Sci Med.* 1995;40(1):77-90.

17. Kogan MD, Alexander GR, Teitelbaum MA, Jack BW, Kotelchuck M, Pappas G. The effect of gaps in health insurance on continuity of a regular source of care among preschool-aged children in the United States. *Jama.* 1995;274(18):1429-1435.

18. Schor EL. Use of health care services by children and diagnoses received during presumably stressful life transitions. *Pediatrics.* 1986;77(6):834-841.

19. Prosser LA. Current challenges and future research in measuring preferences for pediatric health outcomes. *J Pediatr.* 2009;155(1):7-9.

20. Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *Pharmacoeconomics.* 2011;29(8):641-652.

21. Petrou S, Gray R. Methodological challenges posed by economic evaluations of early childhood intervention programmes. *Appl Health Econ Health Policy.* 2005;4(3):175-181.

22. Prosser LA, Hammit JK, Keren R. Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations. *Pharmacoeconomics.* 2007;25(9):713-726.

23. Basu A, Meltzer D. Implications of spillover effects within the family for medical cost-effectiveness analysis. *Journal of Health Economics.* 2005;24(4):751-773.

24. Prosser LA, Lamarand K, Gebremariam A, Wittenberg E. Measuring Family HRQoL Spillover Effects Using Direct Health Utility Assessment. *Medical Decision Making.* 2014;35(1):81-93.

25. Lavelle TA, D'Cruz BN, Mohit B, et al. Family Spillover Effects in Pediatric Cost-Utility Analyses. *Applied Health Economics and Health Policy.* 2019;17(2):163-174.

26. Wittenberg E, James LP, Prosser LA. Spillover Effects on Caregivers' and Family Members' Utility: A Systematic Review of the Literature. *PharmacoEconomics.* 2019;37(4):475-499.

27. Patrick DL, Erickson P. *Health status and health policy : quality of life in health care evaluation and resource allocation.* New York: Oxford University Press; 1993.

28. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30(6):473-483.

29. 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):800-812.

30. Ruperto N, Ravelli A, Pistorio A, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. *Clin Exp Rheumatol.* 2001;19(4 Suppl 23):S1-9.

31. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation.* Second edition. ed. Oxford: Oxford University Press; 2017.

32. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health.* 2000;21:587-611.

33. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Research.* 1980;28:1.

34. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res.* 1972;7(2):118-133.

35. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health.* 2009;12 Suppl 1:S5-9.

36. Weinstein MC, Stason WB. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. *New England Journal of Medicine.* 1977;296(13):716-721.

37. Johannesson M, Pliskin JS, Weinstein MC. A note on QALYs, time tradeoff, and discounting. *Med Decis Making.* 1994;14(2):188-193.

38. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by Risk Neutrality. *Journal of Risk and Uncertainty.* 1997;15(2):107-114.

39. Wakker P, Stiggelbout A. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Med Decis Making.* 1995;15(2):180-186.

40. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making.* 1989;9(2):142-149.

41. Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ. Are health states "timeless"? The case of the standard gamble method. *J Clin Epidemiol.* 1999;52(11):1047-1053.

42. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ.* 1986;5(1):1-30.

43. Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Population Health Metrics.* 2003;1(1):1-12.

44. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ.* 2006;25(3):418-431.

45. Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, Brown M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Econ.* 2009;18(11):1261-1276.

46. Ramos-Goni JM, Pinto-Prades JL, Oppe M, Cabases JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and Modeling of EQ-5D-5L Health States Using a Hybrid Approach. *Med Care.* 2017;55(7):e51-e58.

47. Pickard AS, Law EH, Jiang R, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value Health.* 2019;22(8):931-941.

48. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making.* 2006;26(4):410-420.

49. Bentley TG, Palta M, Paulsen AJ, et al. Race and gender associations between obesity and nine health-related quality-of-life measures. *Qual Life Res.* 2011;20(5):665-674.

50. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care.* 2005;43(3):203-220.

51. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics.* 2012;30(8):729-747.

52. Brazier JE, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess.* 2012;16(32):1-114.

53. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002;21(2):271-292.

54. Pal DK. Quality of life assessment in children: a review of conceptual and methodological issues in multidimensional health status measures. *J Epidemiol Community Health.* 1996;50(4):391-396.

55. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ.* 2003;12(8):697-702.

56. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *Pharmacoeconomics.* 2006;24(12):1199-1220.

57. Juniper EF, Guyatt GH, Feeny DH, Griffith LE, Ferrie PJ. Minimum skills required by children to complete health-related quality of life instruments for asthma: comparison of measurement properties. *Eur Respir J.* 1997;10(10):2285-2294.

58. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess.* 2001;5(4):1-157.

59. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res.* 2001;10(4):347-357.

60. Sung L, Young NL, Greenberg ML, et al. Health-related quality of life (HRQL) scores reported from parents and their children with chronic illness differed depending on utility elicitation method. *J Clin Epidemiol.* 2004;57(11):1161-1166.

61. Saigal S, Stoskopf BL, Feeny D, et al. Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. *JAMA.* 1999;281(21):1991-1997.

62. Saigal S, Rosenbaum PL, Feeny D, et al. Parental perspectives of the health status and health-related quality of life of teen-aged children who were extremely low birth weight and term controls. *Pediatrics.* 2000;105(3 Pt 1):569-574.

63. Dalziel K, Catchpool M, Garcia-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. Feasibility, Validity and Differences in Adolescent and Adult EQ-5D-Y Health State Valuation in Australia and Spain: An Application of Best-Worst Scaling. *Pharmacoeconomics.* 2020.

64. Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. Nothing About Us Without Us? A Comparison of Adolescent and Adult Health-State Values for the Child Health Utility-9D Using Profile Case Best-Worst Scaling. *Health Econ.* 2016;25(4):486-496.

65. Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y-3L Health States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ? *Med Decis Making.* 2021;41(5):584-596.

66. 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.

67. Meade T, Dowswell E. Adolescents' health-related quality of life (HRQoL) changes over time: a three year longitudinal study. *Health and Quality of Life Outcomes.* 2016;14(1):14.

68. Kreimeier S, Oppe M, Ramos-Goñi JM, et al. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) Health States: The Impact of Wording and Perspective. *Value Health.* 2018;21(11):1291-1298.

69. Shah KK, Ramos-Goñi JM, Kreimeier S, Devlin NJ. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. *Eur J Health Econ.* 2020;21(7):1091-1103.

70. Kind P, Klose K, Gusi N, Olivares PR, Greiner W. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Qual Life Res.* 2015;24(10):2519-2539.

71. Lipman SA, Reckers-Droog VT, Karimi M, Jakubczyk M, Attema AE. Self vs. other, child vs. adult. An experimental comparison of valuation perspectives for valuation of EQ-5D-Y-3L health states. *Eur J Health Econ.* 2021;22(9):1507-1518.

72. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care.* 2005;43(5):493-499.

73. Ramos-Goñi JM, Oppe M, Stolk E, et al. International Valuation Protocol for the EQ-5D-Y-3L. *PharmacoEconomics.* 2020;38(7):653-663.

74. Lipman SA, Essers BAB, Finch AP, Sajjad A, Stalmeier PFM, Roudijk B. In a Child's Shoes: Composite Time Trade-Off Valuations for EQ-5D-Y-3L with Different Proxy Perspectives. *Pharmacoeconomics.* 2022;40(Suppl 2):181-192.

75. Chen G, Ratcliffe J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics.* 2015;33(10):1013-1028.

76. Janssens A, Thompson Coon J, Rogers M, et al. A systematic review of generic multidimensional patient-reported outcome measures for children, part I: descriptive characteristics. *Value Health.* 2015;18(2):315-333.

77. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PLoS One.* 2015;10(8):1-21.

78. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going? *Pharmacoeconomics.* 2020;38(4):325-340.

79. Kwon J, Freijser L, Huynh E, et al. Systematic Review of Conceptual, Age, Measurement and Valuation Considerations for Generic Multidimensional Childhood Patient-Reported Outcome Measures. *PharmacoEconomics.* 2022;40(4):379-431.

80. Bailey C, Howell M, Raghunandan R, et al. Preference Elicitation Techniques Used in Valuing Children's Health-Related Quality-of-Life: A Systematic Review. *PharmacoEconomics.* 2022;40(7):663-698.

81. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ.* 1997;16(2):155-175.

82. Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *Journal of Clinical Epidemiology.* 1992;45(5):505-512.

83. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes. Comparison of assessment methods. *Med Decis Making.* 1984;4(3):315-329.

84. Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). *Qual Life Res.* 1996;5(2):205-211.

85. Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res.* 1996;5(6):532-538.

86. Beusterien KM, Yeung JE, Pang F, Brazier J. Development of the multi-attribute Adolescent Health Utility Measure (AHUM). *Health Qual Life Outcomes.* 2012;10:102.

87. Moodie M, Richardson J, Rankin B, Iezzi A, Sinha K. Predicting time trade-off health state valuations of adolescents in four Pacific countries using the Assessment of Quality-of-Life (AQoL-6D) instrument. *Value Health.* 2010;13(8):1014-1027.

88. Ratcliffe J, Huynh E, Chen G, et al. Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Soc Sci Med.* 2016;157:48-59.

89. 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):1234-1242.

90. Chen G, Xu F, Huynh E, Wang Z, Stevens K, Ratcliffe J. Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. *Quality of Life Research.* 2019;28(1):163-176.

91. Prevolnik Rupel V, Ogorevc M. EQ-5D-Y Value Set for Slovenia. *Pharmacoeconomics.* 2021;39(4):463-471.

92. Shiroiwa T, Ikeda S, Noto S, Fukuda T, Stolk E. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. *Med Decis Making.* 2021;41(5):597-606.

93. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, et al. Accounting for Unobservable Preference Heterogeneity and Evaluating Alternative Anchoring Approaches to Estimate Country-Specific EQ-5D-Y Value Sets: A Case Study Using Spanish Preference Data. *Value in Health.* 2022;25(5):835-843.

94. Kreimeier S, Mott D, Ludwig K, et al. EQ-5D-Y Value Set for Germany. *PharmacoEconomics.* 2022;40(2):217-229.

95. Rencz F, Ruzsa G, Bató A, Yang Z, Finch AP, Brodszky V. Value Set for the EQ-5D-Y-3L in Hungary. *PharmacoEconomics.* 2022;40(2):205-215.

96. Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A Value Set for the EQ-5D-Y-3L in the Netherlands. *PharmacoEconomics.* 2022;40(2):193-203.

97. Dewilde S, Roudijk B, Tollenaar NH, Ramos-Goñi JM. An EQ-5D-Y-3L Value Set for Belgium. *Pharmacoeconomics.* 2022;40(Suppl 2):169-180.

98. Fitriana TS, Roudijk B, Purba FD, Busschbach JJV, Stolk E. Estimating an EQ-5D-Y-3L Value Set for Indonesia by Mapping the DCE onto TTO Values. *Pharmacoeconomics.* 2022;40(Suppl 2):157-167.

99. Yang Z, Jiang J, Wang P, et al. Estimating an EQ-5D-Y-3L Value Set for China. *Pharmacoeconomics.* 2022;40(Suppl 2):147-155.

100. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care.* 1996;34(7):702-722.

101. McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the HUI 2 descriptive system: results from a UK survey. *Health Econ.* 2005;14(3):231-244.

102. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care.* 2002;40(2):113-128.

103. Krabbe PFM, Jabrayilov R, Detzel P, Dainelli L, Vermeulen KM, van Asselt ADI. A two-step procedure to generate utilities for the Infant health-related Quality of life Instrument (IQI). *PLoS One.* 2020;15(4):e0230852.

104. Seiber W, Groessl EJ, David KM, Ganiats TG, Kaplan RM. *Quality of well being self-administered (QWB-SA) scale: user's manual.* San Diego: Health Services Research Center, University of California, San Diego;2008.

105. 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):875-886.

106. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. *Med Care.* 1999;37(2):126-139.

107. Growing Up in Australia. Australian Institute of Family Studies. https://growingupinaustralia.gov.au/. Published 2020. Accessed April 6, 2020.

108. Australian Institute of Family Studies. Longitudinal Study of Australian Children Data User Guide. In. Melbourne: Australian Institute of Family Studies; 2018.

109. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Qual Life Res.* 2009;18(2):253-265.

110. Costa DS, Aaronson NK, Fayers PM, et al. Deriving a preference-based utility measure for cancer patients from the European Organisation for the Research and Treatment of Cancer's Quality of Life Questionnaire C30: a confirmatory versus exploratory approach. *Patient Relat Outcome Meas.* 2014;5:119-129.

111. Brazier JE, Mulhern BJ, Bjorner JB, et al. Developing a New Version of the SF-6D Health State Classification System From the SF-36v2: SF-6Dv2. *Med Care.* 2020;58(6):557-565.

112. Hoffman S, Lambert MC, Nelson TD, Trout AL, Epstein MH, Pick R. Confirmatory factor analysis of the PedsQL among youth in a residential treatment setting. *Qual Life Res.* 2013;22(8):2151-2157.

113. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-Attribute Health Status Classification Systems. *PharmacoEconomics.* 1995;7(6):490-502.

114. Muthen LK, Muthen BO. *Mplus User's Guide. Eighth Edition.* Los Angeles, CA2017.

115. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107(2):238-246.

116. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psycometrika.* 1973;38(1):1-10.

117. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behav Res.* 1990;25(2):173-180.

118. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6(1):1-55.

119. Limbers CA, Newman DA, Varni JW. Factorial invariance of child self-report across age subgroups: a confirmatory factor analysis of ages 5 to 16 years utilizing the PedsQL 4.0 Generic Core Scales. *Value Health.* 2008;11(4):659-668.

120. Limbers CA, Newman DA, Varni JW. Factorial invariance of child self-report across healthy and chronic health condition groups: a confirmatory factor analysis utilizing the PedsQLTM 4.0 Generic Core Scales. *J Pediatr Psychol.* 2008;33(6):630-639.

121. Limbers CA, Newman DA, Varni JW. Factorial Invariance of Child Self-Report Across Race/Ethnicity Groups: A Multigroup Confirmatory Factor Analysis Approach Utilizing the PedsQL™∗ 4.0 Generic Core Scales. *Annals of Epidemiology.* 2009;19(8):575-581.

122. Varni JW, Limbers CA, Newman DA. Factorial Invariance of the PedsQL™ 4.0 Generic Core Scales Child Self-Report Across Gender: A Multigroup Confirmatory Factor Analysis with 11,356 Children Ages 5 to 18. *Applied Research in Quality of Life.* 2008;3(2):137-148.

123. Limbers CA, Newman DA, Varni JW. Factorial invariance of child self-report across socioeconomic status groups: a multigroup confirmatory factor analysis utilizing the PedsQLTM 4.0 Generic Core Scales. *Journal of Behavioral Medicine.* 2008;31(5):401-411.

124. Young T, Yang Y, Brazier J, Tsuchiya A. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Med Decis Making.* 2011;31(1):195-210.

125. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health.* 2011;14(5):721-731.

126. King MT, Costa DS, Aaronson NK, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Qual Life Res.* 2016;25(3):625-636.

127. Rasch G. *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press; 1980.

128. Stouffer SA, Guttman LA, Suchman FA, Lazarsfeld PF, Star SA, Clausen JA. *Studies in social psychology in World War II: Vol IV Measurement and prediction.* Princeton: Princeton University Press; 1950.

129. Tennant A, McKenna SP, Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health.* 2004;7:S22-S26.

130. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007;57(8):1358-1362.

131. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol.* 2007;46(Pt 1):1-18.

132. Boone WJ. Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sci Educ.* 2016;15(4).

133. Masters GN. A rasch model for partial credit scoring. *Psychometrika.* 1982;47(2):149-174.

134. Andrich D, Lyne A, Sheridan B, Luo G. RUMM2030. In. Perth, Australia: RUMM Laboratory; 2010.

135. Jörngården A, Wettergen L, von Essen L. Measuring health-related quality of life in adolescents and young adults: Swedish normative data for the SF-36 and the HADS, and the influence of age, gender, and method of administration. *Health Qual Life Outcomes.* 2006;4:91.

136. Arrington-Sanders R, Yi MS, Tsevat J, Wilmott RW, Mrus JM, Britto MT. Gender differences in health-related quality of life of adolescents with cystic fibrosis. *Health Qual Life Outcomes.* 2006;4:5.

137. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes.* 2017;15(1):181.

138. Linacre J. Sample size and item calibration stability. *Rasch Meas Trans.* 1994;7:328.

139. Hanmer J, Cella D, Feeny D, et al. Selection of key health domains from PROMIS(®) for a generic preference-based scoring system. *Qual Life Res.* 2017;26(12):3377-3385.

140. National Research Council Panel to Review the Status of Basic Research on School-Age C. In: Collins WA, ed. *Development During Middle Childhood: The Years From Six to Twelve.* Washington (DC): National Academies Press (US) Copyright © National Academy of Sciences.; 1984.

141. Von Neumann J, Morgenstern O. *Theory of games and economic behavior.* New York: Oxford University Press; 1944.

142. Hunink MGM. *Decision making in health and medicine : integrating evidence and values.* Second edition. ed. Cambridge: Cambridge University Press; 2014.

143. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ.* 2002;11(5):447-456.

144. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ.* 1995;4(4):289-299.

145. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes toward quality of survival. The concept of "maximal endurable time". *Med Decis Making.* 1982;2(3):299-309.

146. Dewilde S, Janssen MF, Lloyd AJ, Shah K. Exploration of the Reasons Why Health State Valuation Differs for Children Compared With Adults: A Mixed Methods Approach. *Value in Health.* 2022;25(7):1185-1195.

147. Åström M, Conte H, Berg J, Burström K. 'Like holding the axe on who should live or not': adolescents' and adults' perceptions of valuing children's health states using a standardised valuation protocol for the EQ-5D-Y-3L. *Qual Life Res.* 2022;31(7):2133-2142.

148. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics.* 2014;32(9):883-902.

149. Soekhai V, de Bekker-Grob E, Ellis A, Vass C. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics.* 2019;37(2):201-226.

150. McFadden D. Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics.* New York: Academic Press; 1974:105-142.

151. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CG, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint

Analysis Good Research Practices Task Force. *Value Health.* 2016;19(4):300-315.

152. King MT, Viney R, Simon Pickard A, et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. *Pharmacoeconomics.* 2018;36(2):225-238.

153. Norman R, Mercieca-Bebber R, Rowen D, et al. U.K. utility weights for the EORTC QLU-C10D. *Health Econ.* 2019;28(12):1385-1401.

154. Mulhern BJ, Bansback N, Norman R, Brazier J. Valuing the SF-6Dv2 Classification System in the United Kingdom Using a Discrete-choice Experiment With Duration. *Med Care.* 2020;58(6):566-573.

155. Webb EJD, O'Dwyer J, Meads D, Kind P, Wright P. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Eur J Health Econ.* 2020;21(5):787-800.

156. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health.* 2011;14(4):403-413.

157. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health.* 2013;16:3-13.

158. Qualtrics. In. Provo, UT: Qualtrics; 2023.

159. Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Qual Life Res.* 2016;25(3):637-649.

160. SAS. In. Cary, NC: SAS Institute Inc.; 2023.

161. Prosser LA, Ray GT, O'Brien M, Kleinman K, Santoli J, Lieu TA. Preferences and willingness to pay for health states prevented by pneumococcal conjugate vaccine. *Pediatrics.* 2004;113(2):283-290.

162. Prosser LA, Bridges CB, Uyeki TM, et al. Values for preventing influenza-related morbidity and vaccine adverse events in children. *Health Qual Life Outcomes.* 2005;3:18.

163. Prosser LA, Payne K, Rusinak D, Shi P, Uyeki T, Messonnier M. Valuing health across the lifespan: health state preferences for seasonal influenza illnesses in patients of different ages. *Value Health.* 2011;14(1):135-143.

164. Simon NJ, Richardson J, Ahmad A, et al. Health utilities and parental quality of life effects for three rare conditions tested in newborns. *J Patient Rep Outcomes.* 2019;3(1):4.

165. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics.* 2008;26(8):661-677.

166. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis: A Primer.* Cambridge: Cambridge University Press; 2005.

167. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making.* 2015;35(3):328-340.

168. Mott DJ, Devlin NJ, Kreimeier S, Norman R, Shah KK, Rivero-Arias O. Analytical Considerations When Anchoring Discrete Choice Experiment Values Using Composite Time Trade-Off Data: The Case of EQ-5D-Y-3L. *PharmacoEconomics.* 2022;40(2):129-137.

169. StataCorp. Stata Statistical Software: Release 14. In. College Station, TX: StataCorp LP; 2015.

170. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL™* 4.0 as a Pediatric Population Health Measure: Feasibility, Reliability, and Validity. *Ambulatory Pediatrics.* 2003;3(6):329-341.

171. Ramos-Goñi JM, Estévez-Carrillo A, Rivero-Arias O, et al. Does Changing the Age of a Child to be Considered in 3-Level Version of EQ-5D-Y Discrete Choice Experiment-Based Valuation Studies Affect Health Preferences? *Value Health.* 2022;25(7):1196-1204.

172. Norman R, King MT, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Qual Life Res.* 2010;19(4):499-508.

173. Jiang R, Shaw J, Mühlbacher A, et al. Comparison of online and face-to-face valuation of the EQ-5D-5L using composite time trade-off. *Qual Life Res.* 2021;30(5):1433-1444.

174. Greene WH, Hensher DA. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation.* 2010;37(3):413-428.

175. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making.* 2001;21(1):7-16.

176. Bouckaert N, Cleemput I, Devriese S, Gerkens S. An EQ-5D-5L Value Set for Belgium. *Pharmacoecon Open.* 2022;6(6):823-836.

177. Rencz F, Brodszky V, Gulácsi L, et al. Parallel Valuation of the EQ-5D-3L and EQ-5D-5L by Time Trade-Off in Hungary. *Value Health.* 2020;23(9):1235-1245.

178. Devlin N, Parkin D, Janssen B. *Methods for Analysing and Reporting EQ-5D Data.* Springer Cham; 2020.

179. 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):453-467.

180. Lambe T, Frew E, Ives NJ, et al. Mapping the Paediatric Quality of Life Inventory (PedsQL™) Generic Core Scales onto the Child Health Utility Index-9 Dimension (CHU-9D) Score for Economic Evaluation in Children. *Pharmacoeconomics.* 2018;36(4):451-465.

181. Sweeney R, Chen G, Gold L, Mensah F, Wake M. Mapping PedsQL(TM) scores onto CHU9D utility scores: estimation, validation and a comparison of alternative instrument versions. *Qual Life Res.* 2020;29(3):639-652.

182. Kelly CB, Soley-Bori M, Lingam R, et al. Mapping PedsQL™ scores to CHU9D utility weights for children with chronic conditions in a multi-ethnic and deprived metropolitan population. *Quality of Life Research.* 2023;32(7):1909-1923.

183. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL™ generic core scales. *Pharmacoeconomics.* 2014;32(7):693-706.

184. Varni JW, Limbers CA, Neighbors K, et al. The PedsQL Infant Scales: feasibility, internal consistency reliability, and validity in healthy and ill infants. *Qual Life Res.* 2011;20(1):45-55.