Childhood health-related quality of life in context of economic evaluation in Malawi

Thesis submitted in accordance with the requirement of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy in Global Health

by

Lucky Gift Chiwiya Ngwira

Student ID: 00146702

Submitted on 18th April 2022

Supervisor: Prof Louis W Niessen, Liverpool School of Tropical Medicine
Co-supervisors: Prof Stavros Petrou, University of Oxford
Assoc Prof Sarah Smith, , London School of Hygiene and Tropical Medicine
Dr. Hendramoorthy Maheswaran, Imperial College London

Declaration

I, Lucky Gift Ngwira, hereby declare that this thesis is based on my original work. To the best of my knowledge it does not contain work previously published or written by another person except where due acknowledgement has been made in the text. I declare that neither the whole nor part of this thesis has been or is to be submitted for another degree.

Signature:

Date: 18th April 2022.

.

Funding

This doctorate was funded by the MRC Norway through the BREATHE trial (NCT02426112) from 2017 to 2020. The IMPALA consortium (NIHR 16/136/35) funded an eight-month stipend support for the write-up period of the doctorate from January-August 2021. This work is also a recipient of a small grant from the EuroQol Foundation (Project 20190200). The Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW) in Blantyre, Malawi, is acknowledged as the host institute.

Potential conflict of interest

As part of the cross-cultural adaptation of the EQ-5D-Y and EQ-5D-Y-5L, I received a small grant from the EuroQol Foundation (non-profit organisation). The EuroQol Foundation oversees the work of the EuroQol Group who developed the EQ-5D-Y and EQ-5D-Y-5L. The Foundation has not been involved in the preparation, formulation, nor write-up of this thesis.

Abstract

Background & aims:

Preference-based measurements (instruments) of childhood health-related quality of life (HRQoL) in sub-Saharan Africa (sSA) are limited. This thesis set out: i) to identify existing childhood preference-based HRQoL instruments; ii) to explore how the existing instruments are adapted for use in sSA and; iii) to establish if the adaptation process aligns with local health perceptions in Malawi.

Methods:

Step 1: Systematic review and selection of childhood preference-based HRQoL instruments

This step involved: i) a literature review to identify existing childhood preference-based HRQoL measures, and their use in sSA settings; and ii) assessing identified instruments in terms of their psychometric properties and adherence to utility theory to inform the selection of measures for cross-cultural adaptation in Malawi.

Step2: Cross-cultural adaptation of childhood preference-based HRQoL instruments

The adaptation process included the following two stages.

i) Translation:

Two independent translators forward and backward translated selected instruments (EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report, and PedsQL™ 4.0 teen self-report) into Chichewa language for Malawi. The translated versions were piloted (cognitive interviews) in a sample of participants (n=38), aged 8-17years from the main referral hospital and primary and secondary schools within Blantyre city, the commercial capital of Malawi.

ii) Cognitive interviews process:

a) Self completion of the questionnaire

For the EQ-5D-Y-5L only, a card exercise was employed to assess the correct hierarchical ranking of response options. All the Chichewa translated instruments were administered to different sets of participants for self-completion. If necessary, children who struggled to understand instruction for self-completion of the questionnaires were given assistance but not on how to respond to them.

b) Cognitive interviews

Each child participating in the interviews then took part in a one-to-one interview about their understanding of the questionnaire. The interviews were open ended, and participants were asked to comment on how they understood each question and response options. Following piloting and making necessary changes, pre-final Chichewa versions were sent to developers for approval.

Step 3: Psychometric validation

The approved Chichewa versions were administered to a larger sample (n=298, age 8-17 years) consisting of healthy (n=95) and sick (chronic and acute, n=194) children to assess psychometric performance. The EQ-5D-Y and EQ-5D-Y-5L have five dimensions each with three and five response options respectively. The PedsQL™ 4.0 has 23 items and five response options each. The EQ-5D-Y and EQ-5D-Y-5L scores were generated in two different ways. HRQoL sum scores were generated by adding numeric values of response levels and utility scores were generated using USA utility values to evaluate psychometric performance and utility performance (empirical validity) respectively. The PedsQL™ 4.0 scores were generated following an algorithm provided by the instrument developers. The instruments performance were then assessed using standardized psychometric criteria: item performance (missing data >5%); internal consistency reliability (Cronbach $\alpha \geq 0.7$); and validity (convergent of EQ-5D-Y sum scores and PedsQL™ 4.0 scores >0.4; discriminant: no association between HRQoL scores with gender & grade; known groups between healthy and sick groups: effect size >0.5). Additionally, since the EQ-5D-Y and EQ-5D-Y-5L are utility instruments, they were assessed for empirical validity (relative efficiency of the EQ-5D-Y-5L relative to the EQ-5D-Y, where >1 means the former performing better than the latter). Content validity was later evaluated by firstly developing a conceptual framework for health from four focus group discussions (FGDs) composed of 6-12 participants, aged 8-17 years. The framework was conceptualized using axial and selective coding FGDs to generate main concepts which were considered important by children and adolescents in a Malawian setting. The concepts from the model were compared against those of the existing EQ-5D-Y dimensions to assess if the EQ-5D-Y contains all the relevant dimensions that children and adolescent consider important in this setting.

Results:

Ten preference-based HRQoL instruments were identified and none of these had been developed in low- and/or middle- income countries. Of these, only two (HUI3 and EQ-5D-Y had been used in a sSA setting. : i) EQ-5D-Y (preference-based), and ii) PedsQL™ 4.0 self-report (non preference-based) were selected for cross-cultural adaptation into Chichewa and subsequently psychometrically validated in a Malawian setting.

i) Translation:

There were several conceptual and linguistic problems identified from the translation process. Most of these were identified and resolved by translators during the consensus process. However, translation issues such as those referring to 'life', 'health', and 'healthy' were problematic. Some translation issues including those pertaining to qualifiers for responses and dimension equivalence were only resolved during cognitive interviews.

ii) Cognitive interviews process:

a) Self-completion of the questionnaires

Children, <12 years, struggled to self-complete questionnaire without interviewer assistance.

b) Cognitive interviews on conceptual issues and questionnaire

Children related the EQ-5D-Y 'looking after oneself' dimension to lack of provision such as soap instead of functional ability. 'Kusamba' in PedsQL™ 4.0 translated 'to bath' was found to imply a menstrual cycle and the translation was subsequently revised. Four rounds of card ranking exercise were necessary to establish appropriate EQ-5D-Y-5L hierarchical ordering.

iii) Psychometric validation:

There was little problem with missing data except in children aged 8-12 years. The Cronbach assessment of internal consistency reliability was acceptable and above the threshold (α>0.7) for all instruments. The correlation of EQ-5D-Y (EQ-5D-Y-5L) sum scores and PedsQL™ 4.0 overall scores was >0.4, providing evidence of convergent validity; but the correlation between some of the EQ-5D-Y dimensions and PedsQL™ 4.0 self-report sub-scales was mixed. There was no association between HRQoL scores with gender and age but the association with school grade ($p<0.05$) showed lack of discriminant validity. For empirical validity, the EQ-5D-Y-5L was 31%-91% less efficient than the EQ-5D-Y at detecting differences in health status using external measures. A new conceptual framework describing health for Malawi children/young people was developed with seven concepts. Some concepts like 'God-given 'and 'acceptance

of illness' have rarely been included in generic health instruments. Only two concepts in the model were at the same level as two EQ-5D-Y dimensions.

Discussion and conclusion:

Health, life and healthy were complex terms to translate, and even with help from cognitive interviews these were found to be problematic. The cognitive interviews were instrumental in isolating translation problems and appropriate hierarchical ordering. The psychometric findings demonstrate that the Chichewa (Malawi) PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report established reliability, but evidence of their validity was mixed. While the PedsQL™ 4.0 child self-report met most psychometric criteria, the missing data and lack of known-groups validity means this should be used with caution in younger children and in those with different health conditions.

The study did not find much difference between the EQ-5D-Y and EQ-5D-Y-5L in terms of missing data, reliability and known-group validity. The EQ-5D-Y had problems with known-groups validity. The EQ-5D-Y-5L demonstrated reliability and validity except discriminant validity with grade. Whereas the EQ-5D-Y-5L performed better than the EQ-5D-Y in terms of discriminant validity, the latter outperformed the EQ-5D-Y-5L in the utilization of response options as well as in terms of empirical validity. The EQ-5D-Y and EQ-5D-Y-5L were found to be reliable for use among children and adolescents in relation to economic evaluation. However, psychometric testing is required for test re-test reliability and responsiveness that could not be carried out in this study due to COVID-19 restrictions. Additionally, the conceptual framework developed demonstrates that HRQoL instruments do not align with local perceptions of health. There are some elements of content validity missing and also concepts that could not translate easily limit the content of these instruments in this setting.

In conclusion, this doctorate research produced Chichewa versions which have potential for use in this setting among children and adolescents but with caveats. This thesis documents the cross-cultural adaptation and validation processes of the EQ-5D-Y, EQ-5D-Y-5L and PedsQL™ 4.0 self-report for use among children and adolescents in Malawi.

Acknowledgements

Firstly, I would like to express my utmost gratitude to a very supportive supervisory team led by Prof Louis Niessen during this doctorate journey that has been completed during the most uncertain months of the ongoing COVID-19 pandemic. Prof Niessen took me under his wings as a doctoral candidate even though he hadn't heard of me before but happily accepted to be my mentor, given the importance of the topic. He has given me overwhelming support and encouragement throughout. My sincere thanks go to co-supervisor Prof Stavros Petrou for his guidance and support throughout this research. A huge thanks goes to Prof Sarah Smith for her expert psychometric and qualitative support. Prof Smith has been an invaluable team asset and provided me with a critical and technical support for me to learn the trade. Dr Hendramoorthy Maheswaran also deserves to be mentioned. He was the link within the whole supervisory team and me and was always available to provide a supportive advisory role.

I would also want to thank the MRC Norway and the IMPALA consortium for funding my studies, and the MLW for being the host institute.

More important are the parents and the children who gave consent for this study to happen.

Last but not the least, I want to express my utmost gratitude to my family, my lovely wife Wendy and son, Shaun. You guys have been wonderful. You gave me support and space required on this doctorate journey.

Personal development portfolio

PhD training

2019 Utility Data for Health Technology Assessment, University of York, UK

2018 Statistics and R short course, MLW and CoM, Malawi

2018 The Application of Psychometrics for Measuring Health Outcomes and Quality of Life, University of York, UK

2017 Decision Analytic Modelling for Economic Evaluation, University of York, UK

2017 Online: Measuring and Valuing Health

2017 Online: Health Technology Assessment

Membership

2020- Current Health Financing Technical Working Group, Ministry of Health, Malawi

2020- Current Special Interest Group: Youth Working Group, EuroQol, Netherlands

2020- Current Executive Member: Health Economics Association of Malawi

2018- Current Member: ISOQOL Psychometrics/ Child Health/ Translation and Cultural adaptation Groups, USA

Awards

2020 EuroQol Foundation: EQ-5D-Y analysis support €14,700

2020 EuroQol Foundation travel scholarship, Prague, Czech Republic.

2020 EuroQol Foundation travel scholarship, Cape Town, South Africa.

2019 EuroQol Foundation: Validation of the EQ-5D-Y in Malawi US$14,200

2019 Best oral presentation (3rd place). LSTM PGR conference. Liverpool. UK

2018 MLW Annual dissemination conference travel scholarship, Mangochi, Malawi.

International conferences

2022 2nd African Regional EuroQol Academy Meeting, Virtual. Oral presentation

Title: Comparing the performance of the EQ-5D-Y-3L and EQ-5D-Y-5L across age groups

2021 5th EuroQol Academy Meeting 2021, Virtual. Poster presentation

Title: Head to head psychometric performance of the Chichewa language (Malawi) EQ-5D-Y and EQ-5D-Y-5L

2020 1st EuroQol Early career researchers meeting, Prague. Oral presentation

Title: Development of the new EQ-5D-Y-5L using the ranking order exercises

2020 1st EuroQol African meeting, Cape Town. Oral presentation

Title: Development of the Malawi (Chichewa) EQ-5D-Y-3L and EQ-5D-Y-5L using ranking exercise.

2019 BREATHE Investigators Meeting, Tromso. Oral presentation

Title: Multi-Attribute Utility Instrument in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa

2019 LSTM PGR Conference, Liverpool. Oral presentation

Title: Validation of the EQ-5D-Y and EQ-5D-Y-5L for use in economic evaluation in children and adolescents in sub-Saharan Africa

2018 LSTM PGR Conference, Liverpool. Poster presentation

Title: Measurement of preference-based health related quality of life in children and adolescents in sub-Saharan Africa

National conferences

2017 MLW Annual Scientific Meeting, Mangochi. Poster presentation

Title: Measurement of preference-based health related quality of life in children and adolescents in sub-Saharan Africa

Publications, reports and policy related documents

Publications

1. **Ngwira, L.G.**; Jelsma, J.; Maheswaran, H.; Kapakasa, F.; Petrou, S.; Niessen, L.; Smith, S.; Cross-cultural adaptation of the beta EQ-5D-Y-5L into Chichewa (Malawi). *Accepted. VIHRI*
2. Verstraete, J.; **Ngwira, L.G.**; Scott, D.; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S.; Comparing the performance of the EQ-5D-Y-3L and EQ-5D-Y-5L across age groups. *(draft)*
3. **Ngwira, L.G.**; Khan, K.; Maheswaran, H.; Sande, L.; Nyondo-Mipando, L.; Smith, S.; Petrou, S.; Niessen, L.; Systematic Review of validity and use of childhood preference-based health-related quality of life measures in sub-Saharan Africa. *VIHRI 2021; 25(C):37–47*
4. Sarah Derrett, Mike Herdman, **Lucky G. Ngwira**, Elizabeth Yohe Moore, Jennifer Jelsma. A new approach to assessing children's interpretation of severity qualifiers in a multi-attribute utility instrument – the EQ-5D-Y: Development and testing. *The Patient 2021*; https://doi.org/10.1007/s40271-021-00496-1

Reports

1. **Ngwira, L.G.**; Jelsma, J.; Maheswaran, H.; Kapakasa, F.; Petrou, S.; Smith, S.; Niessen, L.; **Development and validation of the Malawi (Chichewa) EQ-5D-Y-3L paper self-complete.** 2019. Submitted to the EuroQol Foundation. https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/self-complete-on-paper/
2. **Ngwira, L.G.**; Jelsma, J.; Maheswaran, H.; Kapakasa, F.; Petrou, S.; Smith, S.; Niessen, L.; **Development of the Malawi (Chichewa) EQ-5D-Y-5L paper self-complete.** 2019. Submitted to the EuroQol Foundation. (link unavailable as EQ-5D-Y-5L is a trial version)
3. **Ngwira, L.G.**; Maheswaran, H.; Kapakasa, F.; Petrou, S.; Smith, S.; Niessen, L.; **Development of the Malawi (Chichewa) PedsQL 4.0 (child and Teen) paper self-complete.** 2019. Presented to MAPI Trust. http://www.pedsql.org/PedsQL-Translation-Tables.xlsx

Contribution to policy related documents

1. **The Malawi National Health Financing Strategy 2021. Draft 0. Malawi MoH.** *Under development and undergoing internal review*

Contributions to each chapter and related papers

Author	Chapter	Author contribution								
		Concept and design:	*Acquisition of data:*	*Analysis and interpretation of data:*	*Drafting of the chapter/ manuscript:*	*Critical revision of the chapter/paper for important intellectual content:*	*Provision of study material:*	*Obtaining funding:*	*Administrative, technical, or logistic support:*	*Supervision:*
Ngwira, L.G. (LGN), Khan, K. (KK), Jelsma, J. (JJ), Maheswaran, M. (HM), Sande, L (LS), Nyondo-Mipando, L. (LNM), Kapakasa, F. (FK), Verstraete, J. (JV); Derrett, S. (SD), Smith, S. (SS), Petrou, S. (SP), Niessen, L.W. (LWN)	1	LGN	N/A	N/A	LGN	LGN HM SS SP LWN	LGN	LGN LWN	LWN	HM SS SP LWN
	2	LGN HM SP LWN	LGN, LS	LGN	LGN	LGN KK HM LS LNM SS SP LWN	LGN			
	3	LGN SS	N/A	N/A	LGN SS	LGN HM SS SP LWN	LGN			
	4	LGN HM SS SP LWN	LGN, FK	LGN JJ	LGN	LGN JJ HM FK SD SS SP LWN	LGN SD			
	5	LGN HM SS SP LWN	LGN	LGN SS	LGN	LGN HM SS SP LWN	LGN			
	6	LGN HM SP LWN	LGN	LGN SS	LGN JV	LGN JV HM SS SP LWN	LGN			

	7	LGN HM SS SP LWN	LGN	LGN SS	LGN	LGN HM SS SP LWN	LGN			
	8	LGN HM SS SP LWN	LGN	LGN	LGN	LGN HM SS SP LWN	LGN			

Table of contents

List of Tables

List of Figures

Abbreviations

12D	12-Dimensional
15D	15-Dimensiona
16D	16-Dimensional
17D	17-Dimensional
AHUM	Adolescent Health Utility Measurement
AIDS	Acquired Immune Deficiency Syndrome
AQoL-6D	Assessment of Quality of Life -6 Dimension
AQoL-8D	Assessment of Quality of Life -8 Dimension
ART	Antiretroviral therapy
BWS	Best worst scaling
CEA	Cost effectiveness analysis
CH-6D	Child Health-6 Dimension
CHQ	Child Health Questionnaire
CHOQ-PF50	Child Health Questionnaire – Parent Form 50
CHRS	Children's Health Ratings Scale
CHSCS	Child's Health Self-Concept Scale
CHU-9D	Child Health Utility 9Dimension
COOP	The Dartmouth Cooperative Functional Assessment Charts
COSMIN	COnsensus-based Standards for the selection of health Measurement Instruments
COVID-19	Corona Virus Disease-2019
CUA	Cost utility analysis
DCE	Discrete choice experiments
DUHP	Duke Health Profile
EHP	Essential health package
EQ-5D	EQ-5D family of instruments
EQ-5D-3L	EQ-5D-3L adult version with three response level
EQ-5D-5L	EQ-5D-5L adult version with five response level
EQ-5D-Y	EQ-5D-Youth version with three response level
EQ-5D-Y-5L	EQ-5D-Youth version with five response level
EQ-HWB	EQ- Health and Well Being
EQ VAS	EQ-5D visual analogue scale
FDA	Food and Drug Administration
FGD	Focus group discussion
IHME	Institute for Health Metrics and Evaluation
HEPU	Health Economics Policy Unit
HICs	High income countries
HIV	Human Immunodeficiency Virus
HRQoL	Health related quality of life
HAS.	Health services attendant
HSSP	Health sector strategic plan
HIS	Health status index
HTA	Health technology assessment
HUI	Health Utility Index
HUI1	Health Utilities Index Mark 1
HUI2	Health Utilities Index Mark 2

HUI3	Health Utilities Index Mark 3
ICC	Intraclass correlation coefficient
ICER	Incremental cost-effectiveness ratio
ISOQOL	International Society for Quality of Life
KINDL	Kinder Lebensqualität fragebogen
LMIC	Low- and middle-income countries
MAU	Multi-attribute utility
MAUI	Multi-attribute utility instruments
MAUF	Multi-attribute utility function
MAUT	Multi-attribute utility theory
MDG	Millennium development goals
ME	Magnitude estimates
MLW	Malawi-Liverpool-Wellcome Trust
MoH	Ministry of Health
MTMM	Multitrait-multimethod
NCDs	Non-communicable diseases
NHP	Nottingham Health Profile
NMCP	Malawi National Malaria Control Programme
MOT	Medical outcomes trust
NCDs	Non-communicable diseases
PBM	Preference based measures
PedsQL ™ 4.0	Pediatric Quality of Life Inventory ™ Version 4.0
PMTCT	Prevention of mother to child transmission of HIV
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROM	Patient reported outcome measure
PTO	Person trade off
QALY	Quality adjusted life years
QoL	Quality of life
QWB	Quality of Well Being
QWB-SA	Quality of Well Being-Self Administered
RI	Rosser Index
RS	Rating scale
SLA	Service level agreement
sSA	sub-Saharan Africa
SDG	Sustainable development goals
SEM	Standard error of measurement
SIP	Sickness Index Profile
SRM	Standardized response mean
SG	Standard gamble
SF-36	Short Form 36
SF-6D	Short Form 36 -6Dimension
SRM	Standardized response mean
TB	Tuberculosis
TTO	Time trade off
UK	United Kingdom
UNAIDS	United Nations Programme on HIV/AIDS
UNICEF	United Nations International Children's Emergency Fund

USA	United States of America
VAS	Visual analogue scale
VMC	Version management committee
WHO	World Health Organization

1 Chapter 1-Introduction to thesis and background

1.1 Chapter overview

This chapter introduces the whole thesis and provides a general format of how each chapter will be presented. This first chapter provides an overview of Malawi, including its geographic and demographic profile, the health services and child health issues.

Over the last two decades there have been great strides by successive Malawi governments aimed at improving health. This is reflected in the improvement in life expectancy of the general population and reductions in premature mortality amongst children. It is likely that the focus of future interventions will not only be aimed at improving general health but also quality of life (QoL). Measuring health is also important in determining inequalities that exist within a health care system. To measure heath as well as QoL or health-related quality of life (HRQoL), appropriate instruments are needed. These will be discussed in detail in chapter 2.

This chapter introduces Malawi and the different economic and health care challenges she faces, particularly in relation to child health. The chapter then provides a problem statement and the research question for the thesis, including the overall aim and objectives that this research addresses. A brief overview on the need for economic evaluation and policy research issues is also presented. The final part of the chapter provides an overview of the doctoral thesis including a summary of what each chapter presents. Finally, the chapter ends with a summary of what is contained in this chapter as well as introducing chapter 2.

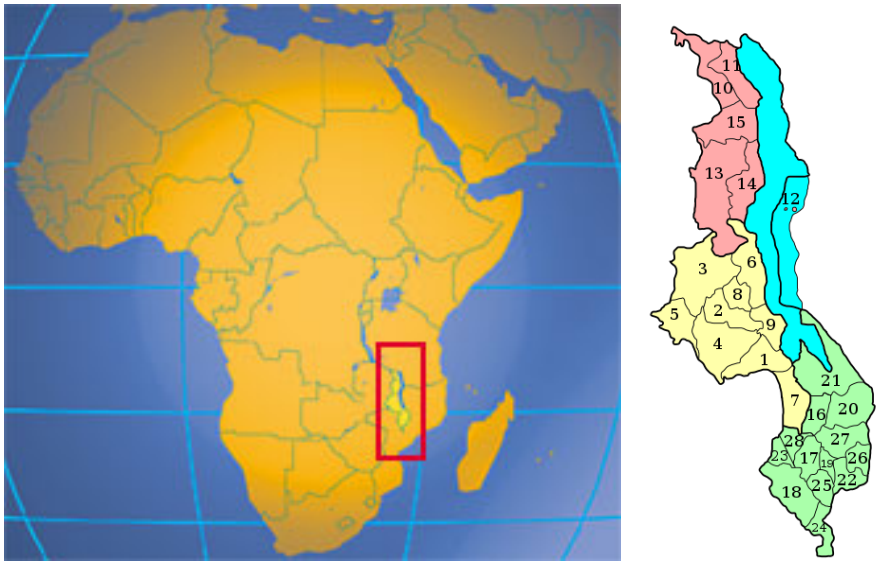
1.2 Introduction

1.2.1 Malawi Country Profile.

1.2.1.1 Geographic information

Malawi (Figure 1.1) is a small land locked country located in Southern Africa. The total land size is 118,484 km^2 and about 25% (29,600km^2) of the land is covered by Lake Malawi. (1) Malawi is bordered by Tanzania to the north and east, Mozambique to the east, south and west, and Zambia to the west.

Figure 1.1: Map of Malawi in relation to Africa

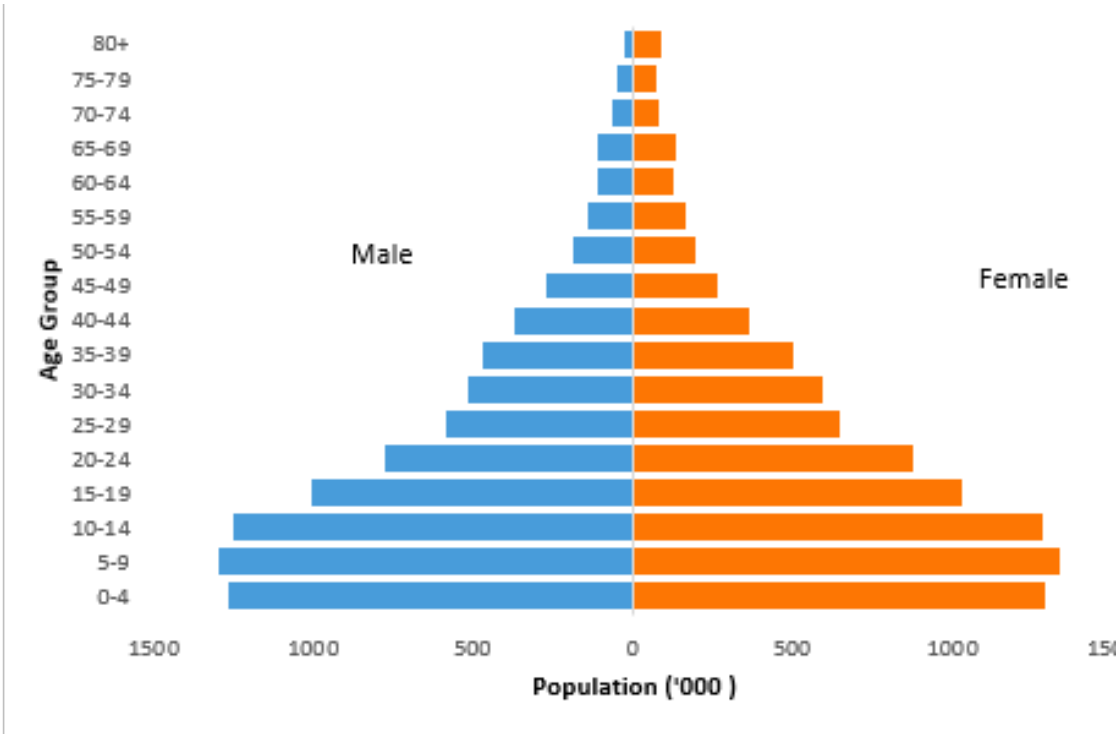


1.2.1.2 Demography

The most recent 2018 Population and Housing Census shows Malawi had a population of 17,600,000 in 2018. (2) The annual population growth rate is approximately 2.9%, and the population is projected to rise to 19,200,000 by 2021. (3) The 2018 census also found that 2.6 million people were aged below the age of 5 years, 6.3 million between 5-17 years and 8.7 million aged ≥18 years. (2) Further, about 3% of the total population are infants aged less than 1 year, 15% are under-five years, 49% are adults aged ≥18 years, and only 4% are aged >65years. The median age is 17 years, highlighting that Malawi has a relatively young

population. The population pyramid in Figure 1.2 highlights that the largest population groups are those aged below 19 years.

Figure 1.2 Population pyramid for Malawi from the 2018 Population and Housing Census



Source: Malawi Population and Housing Census 2018 (2)

1.2.1.3 Economic activity

The most recent estimates of gross domestic product (GDP) in Malawi was $7.7billion. (4) However, the ongoing Corona virus disease (COVID-19) pandemic is expected to have drastic impact on the already fragile local economy. The GDP per capita is estimated at US$411, making Malawi one of the poorest countries in the world. (5) Poverty is described as those earning less than US$1.90 per person per day and the current GDP per capita for Malawi is around US$1.25 per day. (4) It is currently estimated that 51.5% of the population lives below the poverty line. (2) There are higher levels of poverty in the rural areas where 80% of the population live. The high levels of poverty affect people's lives in many ways including lack

of basic provisions such as clean water, food, housing, access to medical care and other daily needs.

While legally one attains adulthood at 18 years of age, the majority of Malawians enter the labour market at 15 years of age. (2) About 52.3% of the population are of this working age group (15-64 years), and of these 72% (6,614,065) are economically active. The economic activity includes both formal and informal work, and a large proportion (85-89%) work in the informal sector. In general, 90% of the labour force are also involved in some form of subsistence farming, growing crops mainly for their household consumption. (2) The 10-15% in formal employment work in several sectors of the economy. (2) The largest employer is the civil service and the official retirement age of a civil servant is 60 years. It is, however, not uncommon for contract extension past the retirement age. Professionally, it is estimated that the formal agriculture sector accounts for slightly over half (51.5%) of all those in employment. (6) The agricultural crops grown are export products and include tobacco, tea and coffee. Figure 1.3 shows the distribution of the total labour force in Malawi.

Figure 1.3 showing Malawi labour force distribution

Source: Malawi Population and Housing Census, 2018 (2)

1.2.1.4 Administrative and political information

Previously called Nyasaland as part of British protectorate, Malawi became independent in 1964 and attained republic status in 1966. There are three arms of government: executive, legislature and judiciary. The head of the state (President) who is also the head of executive arm is elected to office through elections held every five years. The first head of state, Dr. Hastings Kamuzu Banda, ruled Malawi from 1966 to 1994 through a one-party system. Since the abolition of one-party rule in 1994, Malawi has a multi-party system. The President is elected to office for a maximum of two consecutive five-year terms. Previously, the presidential candidate with the largest votes irrespective of the margin would form the government. However, through the landmark fresh presidential elections case of 2020 as sanctioned by the high court, sitting as a constitutional court, a candidate is elected as president if they amass >50%+1 vote. (7) The head of the legislature, the Speaker, is normally elected from a party with the largest number of parliamentarians by all members of parliament. The head of the judiciary, the Chief Justice, is appointed by the President through recommendations from the judicial service commission.

Malawi is politically and culturally divided into three regions; North, Centre and South. These three regions are administratively divided into a total of 28 districts. Each district has an administrative council and each district council is partitioned into traditional authorities (TAs). The TAs are further sub-divided into villages, which have a village headman as the smallest administrative unit. There are four main cities in Malawi: Mzuzu in the North, Lilongwe - the capital city - in the centre, Zomba - the colonial capital, and Blantyre the commercial city in the South.

The majority of people live in rural areas of Malawi and mostly involved in subsistence farming with only 16% being urban dwellers (2). (8) There are at least twelve tribes in Malawi (Chewa

34.4%, Lomwe 18.9%, Yao 13.3%, Ngoni 10.4%, Tumbuka 9.2%, Sena 3.8%, Mang'anja 3.2%, Nyanja 1.9%, Tonga 1.8%, Nkhonde 1.0%, Lambya 0.6%, Sukwa 0.5%) and other tribes make up 1.1% of the population. (2) Languages are spoken on tribal lines so that each of these tribes has its own language. Chichewa (spoken by the Chewa who are the largest tribe in Malawi) is the national language and spoken by the majority of the population. (9) Since Malawi used to be under the British Empire, up to 1964 when she attained independence, English remains the official language. A selection of socio-economic indicators is shown in Table 1.1 below.

Table 1.1 Summary of selected socio-economic indicators for Malawi

Indicator	Value
Population (million)	18.6
Annual growth rate (%)	2.9
Life expectancy at birth (years)	64
Literacy rate (%)	70
School completion rate (%)	51
Population below 15 years (%)	50
GDP per capita (US$)	411
Population below US$1/day (%)	51.5

source: Malawi Population and Housing Census 2018 (2); World Bank (5)

1.2.1.5 Education system

The literacy rate in Malawi, defined as the ability to read or write a sentence in any of the spoken languages in Malawi, is at about 70%. (10) Compulsory education in Malawi runs for a total of 12 years: eight years in primary school and four years in secondary school. In general, there is a high rate of school attendance among both boys and girls in the age group of 6-17 years. While this is greater than 90% for 6-13 year olds attending primary school, the figure drops drastically to an average of 17.5% for 14-17 year olds attending secondary education.

(10) There may be several factors for this but mainly two stand out: limited places in secondary schools and fees. There are more children enrolled into primary schools compared to available places in secondary schools. As a result, only a few out of those who pass exams for entry into secondary school education are selected. The second reason is that whereas primary school is free, fees are a requirement for attendance at secondary school. The fees for secondary education are unaffordable to most households since many Malawians live below the poverty line. Table 1.2 highlights the lower levels of school attendance amongst poorer households in Malawi. School attendance moves from an average of 4.5% for those in the lowest wealth quintile to 41% amongst those in the highest wealth quintile. Further, of those attending secondary education, <10% complete or go beyond secondary education. (10) Several reasons can be attributed to this, including but not limited to fewer places in public universities, unaffordability of university fees, as well as early marriages.

Table 1.2 School attendance in relation to wealth status

Wealth quintile	Gross attendance ratio (%)		Net attendance ratio (%)	
	Primary	Secondary	Primary	Secondary
Lowest	117.9	11.7	90.1	4.4
Second	128.7	16.5	93.2	5.7
Middle	129.5	25.1	94.9	10.6
Fourth	130.7	38.9	96.2	18.7
Highest	127.8	80.1	95.0	41.7
Total	126.9	36.9	93.9	17.7

Source: Malawi National Statistical Office (6)
Gross attendance ratio= % of total number of primary or secondary students/ official primary or secondary school-age group
Net attendance ratio= % of school-age population attending primary or secondary school/ official primary or secondary school-age group

1.2.2 Health service in Malawi

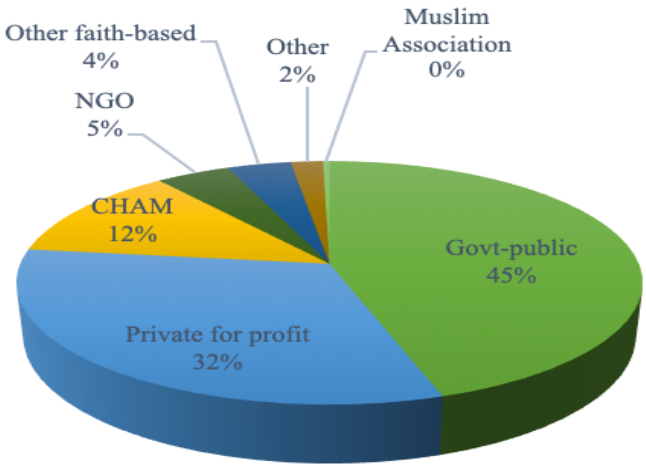
Malawi places high importance on health such that healthcare provision is enshrined within the constitution of the Republic. (11) The health care system is mostly publicly funded, but there is a parallel private system. Table 1.3 and Figure 1.4 below show that the government is the largest single health provider accounting for 45% of all health facilities. The remaining 55% are shared between the private for-profit sector (32%), the Christian Health Association of Malawi (CHAM) 12%, non-governmental organisations (NGOs) 5%, and other faith-based organisations other than CHAM account for the remaining 4%.

Table 1.3 Distribution of health care services in Malawi

Level of health service	Govt-public access	Private for profit	Christian Health Association of Malawi (CHAM)	Muslim Association of Malawi	Faith-based (other than CHAM)	Non-Governmental Organizations (NGO)	Other	TOTALS
Central hospital	4							4
District hospital	24							24
Other hospital	24	16	40		3	3	3	89
Health centre	369	21	115	3	3	9		520
Dispensary	64	94	7	3	1	8	2	179
Clinic	85	363	18		51	49	25	591
Health posts	124	1	12		1			138
Maternity	1							1
TOTALS	695	495	192	6	59	69	30	1546

Source: Malawi Facility Registry, MOHP (12)

Figure 1.4 Distribution of health care services in Malawi



Source: Malawi Facility Registry, MOHP (12)

The government provides health services that are free to Malawians through central hospitals (specialized care), district hospitals (other and middle level specialist care) and primary health centre, health posts or clinics. Health centres offer outpatient and inpatient services as well as conduct minor procedures at primary level. Clinics, dispensaries and health posts on the other hand usually provide health care at community level. The health care services provided at community level are primarily promotive and preventative in nature. Private for-profit organisations require individuals to pay fees at the point of use; or to subscribe to some sort of private health insurance scheme. The private health insurance, which is only about 1.2% of the population, is mostly paid for by the employers at certain large companies and organisations in Malawi. The payment arrangement for CHAM facilities is on a user-fee basis for provision of certain health services. CHAM, previously Private Health Association of Malawi until 1992, is a Christian faith organization that can be traced back to early missionaries who came to settle in the rural communities of Central Africa. Along with their efforts to establish the Christian religion, they also set up education and health systems. (13) Consequently, the majority (80%) of CHAM facilities are in difficult to reach rural communities where some of the government public facilities are not available. (14) CHAM is an important government partner in providing

health care services to the majority of rural Malawians. It is against this background that in 1994 the Malawian government introduced the provision of human resources for health services in CHAM facilities. From 2002, the government approved training of human resources for health services in faith-based health facilities. However, it was not until 2006 that the Malawi government formalized arrangements with CHAM facilities through service level agreements (SLAs), for user-fee cost sharing services. What this means is that the public, in areas where there are no public health facilities within an 8km radius, can access health care services at CHAM facilities. The SLA services between the government and CHAM are not charged for except for those that are not part of the agreement. CHAM facilities consolidate the SLA services, which are in turn reimbursed by the government on a monthly basis. (15) The SLA services are therefore an integral part in the provision of an essential health package (EHP) in Malawi.

EHPs, also known as health benefit packages, have been defined as a "core of interventions that are 'best buy' for the health care system…, and are a critical element to achieving Universal Health Coverage." (16) The provision of EHPs is considered an equity-based 'safety net' that ensures even the poorest attain universal health coverage (UHC). The quest to achieve UHC in Malawi can be traced way back to the beginning of the republic in the 1960s. Soon after attaining independence in 1964, the Malawian government as per the constitution introduced a free health care service. However, the desire to provide a free health care system coupled with an ever increasing population, increases in disease burden and demands for better health care, have affected the quality of services. (17) As a way of overcoming these challenges, most countries with the influence of international organisations adopted the health sector reforms that started in the 1980s and early 1990s. (18, 19) The reforms identified some essential disease areas whose treatment should be provided to all. In Malawi, the reforms did

not start until 1998 when the government introduced decentralisation to enhance healthcare service provision. Under decentralisation, the local district authorities were given semi-autonomy in terms of identifying priority health areas, budgeting and spending. This arrangement did not make significant inroads until 2004 when the government initiated the joint programme of works (POW). The POW was developed as a six year plan (2004-2010) with six pillars (i. human resource, ii. pharmaceutical, medical and laboratory supplies, iii. essential basic equipment, iv. infrastructure and development, v. routine operations and service delivery, and vi. central institutions, policy and systems development) to be implemented by MoH and other stakeholders through the Sector Wide Approach (SWAp). (20)

The SWAp was organized as a health care delivery that had a "common framework for planning, budgeting, and performance monitoring". (21) Through this arrangement, the government and developmental partners could pool resources together in order to effectively deliver common health objectives (sector specific). Further, through the SWAp approach a coordinated and negotiated programme of works (POW) was developed. (21) The POW contained an outline of priority areas, implementation strategies, resource allocation and monitoring and evaluation mechanisms that had to be implemented. This had the advantage of concerted efforts in financial management and delivery instead of many disjointed parallel structures doing the same thing.

The POW covered the period 2004-2010 and contained the very first EHP. The EHP consisted of 11 components that were identified based on bottlenecks in the health system in 2002/03. (22) The package had a focus on reduction of infant and under-five mortality, improving maternal mortality rates, population growth reduction and preventing HIV/AIDS. Despite the systematic way of selecting disease areas, identification of priority areas as well as achieving

EHP objectives has always been challenging. To assist EHP delivery, the government developed the Malawi Growth and Development Strategy (MGDS). The MGDS was designed as a blueprint to achieve the EHPs and was aligned closely to the Millennium Development Goals (MDGs) to be attained by 2015. The MDGs have since been followed with Sustainable Development Goals (SDGs) for nations to attain by 2030. The current EHPs are similarly closely aligned to SDGs.

When the POW time frame expired, it was replaced with the Health Sector Strategic Planning (HSSP) for the period 2011-2016. The EHP was refined at the time in order to include emerging diseases while taking into account the available resources and also utilizing new technologies for resource allocation. (23) The revised EHP within the HSSP therefore consisted of some of the initial disease areas but also non-communicable diseases (NCDs) and other conditions. The NCDs were identified as a major health problem in Malawi through the burden of disease surveys. (24) Additionally, the World Health Organization (WHO) estimates that by 2020 there will be a 60% increase in the burden of NCDs in developing countries like Malawi. (25) As such, burden of disease prioritization as well as cost-effectiveness were the basis for the HSSP EHP design. While taking advantage of a cost-effectiveness framework, the design gave no consideration to the financial affordability to the Malawian health system. (16) This means that EHP implementation within the HSSP was going to be challenging right from the start as it contained some interventions that the country's health budget could not afford.

This raised the need to evaluate the composition of the EHPs to be consistent with spending but also yielding the most population health gains. When the HSSP timeframe expired in 2016, it was replaced by the current HSSP II for the period 2017-2022. The revised EHP within the HSSP II was formulated using a framework that could generate health gains while spending

within the constraints of the health care system. Currently, work is underway to develop HSSP III that will run from 2023-2030. (11) The current proposal aims to use a cost-effectiveness allocation formula to identify disease areas that will be part of the new EHP. (26) Future packaging of EHP will likely rely on cost-effectiveness and other innovative ways. Use of other economic evaluations that rely on preference-based HRQoL instruments to measure effectiveness of interventions in terms of costs and health improvements could be an option. Tables 1.4 and 1.5 below have respectively summaries of chronology of policy changes that shaped and led to the development of the EHPs since 2004. While the first EHP (2004-2010) was based on selected diseases, subsequent EHPs (II and III) have been developed using a cost-effectiveness threshold.

Table 1.4: Malawi essential health package development chronology

Year	Essential Health Package (EHP)	National Plan	Regional and International Plan
1987-2000	Primary health care principles (till 1997). EHP conceptualization (1995 onwards)	• Malawi national health plan 1986-1996 (a 10-year plan since independence) guided by primary health care principles (till 1997) • Malawi development policies (DEVPOL) 1987-1997 • Health policy framework developed in 1995 to shape development of the National health plan since 1996 (but delayed till 1999)	• Poverty reduction strategy papers by World Bank
1999-2004	EHP -costing, refinement & implementation	• 4th National health plan • Malawi Poverty Reduction Strategy Paper (MPRSP) 2002-2005 • Malawi Vision 2020	• World Development Report of 1993 (investing in health) by World Bank • Better Health in Africa, World Bank, 1993-1994
2004-2011	Programme of Works. (Implementation)	• 4th National health plan • Malawi Growth Development Strategy (MGDS) 2006-2011 • Malawi Vision 2020	• Aligned to Millennium Development Goals (MDGs) during implementation
2011-2016	Health Sector Strategic Plan (HSSP)	• MGDS II 2011-2016 • Malawi Vision 2020	• MDGs • 2005 Abuja Declaration • 2008 Ouagadougou Declaration on Primary Health Care

2017-2022	HSSP II	• National Health Policy 2018-2030 • MGDS III 2017-2022 • Malawi Vision 2020	• Sustainable Development Goals (SDGs) • The African Union Agenda 2063
2022-2030	HSSP III	• National Health Policy 2018-2030 • Malawi Vision 2063	• SGDs

Source: (27-29)

Table 1.5 Showing all the EHP for the period 2004-2022

Package	**EHP 2004-2010 under SWAp (POW)***	**EHP 2011-2016 under HSSP[$]**	**EHP 2017-2022 under HSSP II[#]**
1	Vaccine preventable diseases (human resources development)	Vaccine and preventable diseases	Vaccine and preventable diseases
2	Malaria (Pharmaceuticals & Medical supplies)	Malaria	Malaria (diagnosis and treatment)
3	Maternal & Neonatal health conditions including family planning (Essential/basic medical equipment)	Perinatal conditions	Reproductive, Maternal, Newborn and Child Health
4	Acute respiratory tract infections (health infrastructure development)	Acute respiratory infections	Community health (health promotion, disease surveillance, child protection)
5	Acute diarrhea including cholera (routine operations at services delivery level)	Diarrheal diseases	Integrated Management of Childhood illnesses (diarrhea treatment, nutrition)
6	Tuberculosis (central operations, including policy and systems development)	Tuberculosis	Tuberculosis (IPT children, all TB diagnosis and treatment)
7	HIV/AIDS and STI	HIV/AIDS	HIV/AIDS (prevention, testing, treatment)
8	Schistosomiasis	Neglected tropical diseases	Neglected tropical diseases (treatment and drug administration)
9	Malnutrition including micronutrients	Malnutrition	Malnutrition
10	Eye, ear and skin infections	Non-communicable diseases	Non-communicable diseases (mental health, diabetes)
11	Common injuries, accidents including trauma	Eye, ear and skin infection	Oral health

EHP= Essential health package; Green colour indicate what has been maintained over time. EHP 2004-2010 is the only one that was based on selected diseases
*source: Malawi Health Swap Mid-Term Review 2008 (22)
[$] source: Malawi HSSP (30)
[#] source: Malawi HSSP II (11)

In general, the success of UHC through provision of EHPs has been mixed in Africa and associated with considerable challenges. (17) In Malawi, the limited success of EHPs can be attributed to, among other things, level of poverty and health financing. On an individual level, the poverty levels are so high that this hinders most people from utilizing the EHP. The transport costs to get to and from health facilities, especially in rural areas, contribute to a large amount of the out-of-pocket expenditure thereby making it hard to access these services. (31) In addition, even though the EHP is supposed to be without charge, there are charges for medicines and other associated costs that are incurred by the individual. (31) Further, it has also been argued that the EHP was designed from a top-down approach with no consideration for health inequality gaps. (31) As a result of all of this, the EHP is perceived as setting priorities that do not reflect the actual needs of the people on the ground. This makes the implementation of the EHP difficult from an individual perspective.

At a national level, Malawi has been heavily reliant on international donor funding from agencies including the World Bank, Global Fund, International Monetary Fund, African Development Bank, GAVI, United Kingdom (UK) DFID, United States of America (USA) government partners, Germany's Deutsche Gesellschaft fur Internationale Zusammenarbeit (GIZ) and Norway to fuel its national agenda. Donor funding accounts for about 20% of all government funding but this is was high as 75% of all health expenditure as of 2017/2018. (32)

A large percentage of Malawi's GDP (average 12%) is spent on healthcare. (33) While the total annual health expenditure is currently US$ 687.17million, (34) this is only 9.3% of the total Malawi government budget expenditure. (1) This falls short of the Abuja Target, which recommend that health expenditure should be 15% of the total national budget. (17) Further, this health expenditure translates to per capita spending on health of $29 per annum, which is

lower than most countries in the sub-Saharan Africa (sSA) region. (33) The term sSA, as used throughout this thesis, refers to the entire region spanning from south of the Sahara Desert all the way to South Africa. When total health expenditure is put in percentage terms, it translates to the government contributing 23.9%, out of pocket expenditure 12.6% and the private sector account for another 5%. (32) The remaining 58.6%, which account for more than half of total health expenditure, is financed by the donor community. This means that as a country, Malawi is not only unable to finance its own health system but also to provide EHP without assistance. Table 1.6 below provides a summary of health spending in Malawi. The effects of COVID-19 on the global economy will likely present further challenges on donor support reliance. Additionally, for reasons of sustainability there is need to come up with innovative ways of health financing. As an option, financing of UHC through fiscal revenue generation policies has been examined but has only demonstrated to yield minimal gains. (35) One of the recommendations for HSSP II was to improve efficiency in the delivery of health care services. (35) This has not been done but the government recently initiated public reforms to improve the quality of service delivery.

One area that the ongoing government reforms is targeting is the procurement of goods and services. In health care delivery, procurement has been identified as a key driver responsible for wastage and over expenditure. (11) The current draft of the HSSP III is recommending procurement strengthening in order to achieve EHP services. (34) Additionally, government reforms are also focussing at improving human resources - the general attitude of civil servants towards work. The attitude of health workers has been affected by such factors as the remuneration package as well as lack of career advancement opportunities especially in rural areas. (36, 37) It is anticipated that the new reforms will address these issues and subsequently lead to improved health care service delivery. Other than these changes, there is also need to

develop or strengthen tracking tools for resources. (38)

Table 1.6 Table with summary health spending

Indicator	Value
Total national budget (US$ billion)	2.3
Health expenditure per capita (US$)	39.8
Total government health expenditure (US$ million)	687.17
Government health expenditure as percentage of total government expenditure (%)	9.3
Donor funding as percentage of total health expenditure (%)	58.6
Out-of-pocket payment for health (as % of total health expenditure)	12.6
Government expenditure on health as percentage of total health expenditure (%)	23.9
Government per capita health expenditure (US$)	9.5

Source: Health Financing Strategy draft, 2020. (34)

1.2.3 Child health

Childhood is a time when foundation skills are developed, and roles are learned by children that will prepare them for adolescence and adulthood. (39) How a child develops during childhood will affect their future cognitive ability, social, emotional and language skills, physical development, as well as health literacy. (40) However, in most LMICs such as Malawi, poverty and socio-cultural factors that accompany poverty often co-exist. High levels of poverty generally leads to inadequate food, poor hygiene and sanitation which increases children susceptibility to diseases including malnutrition. (41) As discussed below, severe malnutrition can affect physical growth, intelligence as well as school performance.

Some socio-cultural factors also may contribute to children development thereby having effect later in their lives. Culturally, health education in Malawi is largely done in places of seeking health care and schools. While both boys and girls might be in school, it is generally the case

that after school, boys hang out with their friends whereas girls will be responsible for many household chores. It is culturally reasoned that this prepares girls to become responsible women who will be managing homes while the husband is bringing income. This partly contributes to a high school dropout (discussed in detail below) among girls compared to boys.

All these factors can have an effect on how health is perceived between the sexes. These in turn influences school readiness and success in their later life. Further, if children experience inadequate caregiving, environmental stressors, and other negative risk factor, all the milestones could be delayed significantly. Recognizing the important roles the childhood and adolescent periods play in adult health and well-being stresses how vital that this is established early on in their lives. Supportive and safe environments in homes, schools, and communities as well as access to quality health care will shape their understanding of health.

Statistically, Malawi has a relatively young population with nearly half of the population aged <15years and 15% aged <5years. (2) Over recent decades, average life expectancy increased from 37.9 years in 1965, to 45.1 years in 2000 and it now sits at 63.8 years. (5) Future child health interventions will likely be aimed at improving health-related quality of life (HRQoL) outcomes which need associated instruments.

Great strides have been made at reducing under-five childhood mortality in Malawi over the last few decades. For example, Malawi is one of only 62 countries worldwide that met MDG goal 4 aimed at reducing the under-five mortality rate between 1990 and 2015. (17) Under five mortality decreased from 242.8 per 1000 live births in 1990 to 41.6 per 1000 live birth in 2019. (42) The neonatal mortality rate decreased from 36.6 deaths per 1000 live births in 1990 to 17.5 per 1000 live births in 2019. The Infant Mortality Rate (IMR) has similarly shown a

downward trend from 141 per 1000 live births in 1991 to 30.9 per 1000 live births in 2019. (5) The IMR is often seen as an indicator of population health and the downward trend reflects the success of interventions aimed at improving prospects for childhood survival. Table 1.6 provides a summary of the mortality rates for under five children in Malawi and how these compare with the rest of the world.

Table 1.7: Mortality rates for Malawian children in 2019/2020

Indicator	Value	
	Malawi	World
<5-year-old rate (deaths / 1000 live births)	42	39
Infant mortality rate (deaths / 1000 live births)	31	29
Neonatal mortality rate (deaths / 1000 live births)	20	18
Children under 5 years chronically malnourished (%)	37	6.9

source: UNICEF (42); WHO (43) (World values are global rates but these also vary according to WHO region)

Table 1.7 shows that the government of Malawi has worked hard at building on the successes of the MDGs to attain the current SDG goal 3 by 2030. This means that over the next few years, it is expected that there will be more interventions aimed at reducing mortality in children but also improving HRQoL. Despite these advances, child health is still under threat from the following main causes of childhood (for brevity, this refers to children and adolescents in this thesis) death in Malawi.

1.2.3.1 HIV and AIDS

The first HIV case was diagnosed in Malawi in 1985. Since then, HIV/AIDS has had a significant impact on health and socioeconomic outcomes. The prevalence of HIV in Malawi is 8.8% (10.8 % in women and 6.4% in men) and the most affected are those aged 15-49 years.

(44) This means that HIV has affected much of the productive labour force, thereby impacting the economy. Over the years, Malawi has seen a drop in HIV prevalence from 10.6% in 2010 to the current 8.8%. (44) The downward trend can be attributed to among other things, different HIV strategies that the country has implemented. Strategies to combat HIV include behavioural change, incorporation of HIV education into the school curriculum, promotion of safer sexual behaviour, provision of antiretroviral therapy (ART) and other HIV treatment and prevention interventions. (14) The implementation of prevention of mother to child transmission of HIV (PMTCT-Option B+) was very successful in Malawi, leading to universal PMTCT care. (45) The PMTCT-Option B+ entails introduction of lifelong ART for all HIV+ pregnant women irrespective of their CD4 count or clinical staging. Malawi implemented this strategy in 2011, which has contributed to the reduction of new paediatric HIV infection cases from 6,594 in 2011 to 1,655 in 2020. (46)

Malawi has surpassed the 90:90:90 UNAIDS 2020 targets i.e., by 2020 i) 90% of the people with HIV should be diagnosed, ii) 90% of those diagnosed, should start and retained on ART treatment, and iii) 90% of those on ART should have a virally suppressed viral load. The latest data shows that as of June 2020 Malawi had achieved 92:97:94 across the population. (44) Despite these successful initiatives, there are around 74,000 children aged 0-14 years who are HIV+ and only 65% are currently receiving ART. (46, 47) There are several factors affecting various HIV programmes in Malawi including stock-outs of commodities, poor maternal retesting coverage, low uptake of HIV testing among men, and low uptake of ART in children. This only highlights the challenges that remain to improve the HIV situation in Malawi as a whole but particularly amongst children.

1.2.3.2 Tuberculosis

Tuberculosis (TB) is among the top 10 causes of death worldwide and continues to be the leading cause of deaths due to a single infectious agent worldwide. (48) In 2019, 10.0 million fell ill with TB and it claimed the lives of 1.5 million people globally. (48) TB is a disease of poverty and economic distress, with more than two thirds of new cases from the WHO South-East Asia region (44%) and the WHO African region (25%). Nigeria (4%) and South Africa (3.6%) are among eight countries that contributed two thirds to the global burden of TB in 2019. (49) In Malawi, TB is one of the leading causes of mortality and morbidity. Due to co-infection with HIV, it means that Malawi like most other sSA countries has a significant HIV/TB burden. It is estimated that about 45% of people living with HIV have TB globally, and Malawi has reported a similar proportion (41%). (46) In Malawi, the TB incidence rate is 146 per 100,000 population, and mortality rate of 37 per 100,000. (46) Although still high, the figures show that both mortality and incidence rates have gone down over the years.

The success of TB prevention can be attributed to successful strategies that the National TB Control Programme have implemented. The strategies include implementation of various TB diagnosis and treatment monitoring strategies, community TB interventions, and mobilization of adequate resources to support various TB interventions. In terms of reporting, the establishment of electronic medical records has contributed to improvements in notification rates. The high incidence and mortality rates point to demand and supply side challenges to implementing TB interventions. These include, but are not limited to, inadequate infrastructure to expand TB diagnosis and treatment in existing facilities, lack of capacity for TB monitoring, adherence to TB treatment and the over dependence on donor funding to implement TB activities.

1.2.3.3 Malaria

Malaria continues to be a major public health problem worldwide with an estimated 229 million cases globally leading to 409,000 deaths in 2019. (42) Of these deaths, the under five year olds accounted for the largest burden (67%). The Malawi National Malaria Control Programme (NMCP) 1st quarter 2021 report (50) shows that nationally the prevalence in Malawi, by microscopy, is at 24%. The malaria incidence in 2020 was 385 per 1,000 population with a mortality rate of 13 per 100,000 population. It is estimated that malaria accounts for 36% of all outpatient visits and 15% of all hospitalizations in Malawi. (50) The goal of the NMCP is to reduce the incidence of malaria to 193 per 1,000 population by 2022. To achieve this goal, the NMCP promotes use of malaria prevention strategies such as promotion of insecticide-treated bed nets, intermittent preventive treatment in pregnancy, and availability and use of Lumefantrine-Artemether as the first line drug. It is estimated that 85% of children under 5 years of age and pregnant women slept under an insecticide treated net in 2019/2020. (46) Challenges for successful implementation of malaria initiatives include lack of transport to support implementation activities, dependency on donor support for activities, lack of capacity to improve testing services, and non-adherence to treatment regimen. These challenges highlight the need for continued evaluation of malaria intervention programmes.

1.2.3.4 Immunisation coverage

The vaccination programme in Malawi includes the following seven types of vaccines: one dose of Bacillus Calmette-Guerin (BCG) for Tuberculosis protection; two doses for measles-rubella (MR) against measles and rubella; four doses of oral polio vaccine (OPV) against poliomyelitis; three doses of diphtheria, pertussis, tetanus, hepatitis B and haemophilus influenza (PENTA); three doses of pneumococcal conjugate vaccine (PCV) against streptococcus pneumonia; and two doses of rotavirus (ROTA) against gastroenteritis. In general, there is good coverage of the immunization programme in Malawi as shown in Table

1.8 below. The immunization coverage for individual antigen ranges from 80%-94%. In part, the success of the immunization programme in Malawi can be attributed to the outreach clinics and the work of Health Services Attendants (HSAs). (11) HSAs are community-based cadres who play a vital role in running outreach clinics to the most rural communities across the country.

Table 1.8 Malawi vaccination coverage (%) in 2020 per antigen and for the under-fives.

Vaccine	Coverage (%)
PENTA 3	94
OPV 3	93
PCV 3	93
ROTA 2	91
MEASLES 1	89
BCG	87
MEASLES 2	80
Proportion under 5-year-old immunized	90
Proportion under 1-year-old immunized	63

source: HSSP II Mid-term Report, March 2021 draft (46)

A similar successful vaccination programme has been recorded among those under five years of age. Table 1.7 shows this success with 90% of all children under five years of age immunized in 2019/20. As shown above (Table 1.6), this proportion is higher than most countries across the sSA region and in the world. (43) Much as there has been high immunization coverage for individual antigens and among those aged under five years, this is not the case for those under 1 year of age. There has been a steady decline in the rate of immunization for those under 1 year of age from 77% in 2017/2018 to 63% in 2019/2020 partly due to ongoing budget limitations and supply chain leading to a cancellation of outreach vaccine clinics. (46) The COVID-19 pandemic will likely further derail the immunization programme for the 2020/21 period. Despite low immunization rates for under 1-year-olds and possible consequences due to COVID-19, the immunization programme has generally been successful in Malawi. This

has in turn resulted in the prevention of many cases of preventable diseases such as polio.

1.2.3.5 Nutrition

One of the most significant contributors to mortality and morbidity in under-five-years in Malawi is malnutrition. Malnutrition causes among other things stunting, defined as physical growth that is not commensurate with age. Stunting leads to cognitive impairment and lack of productivity when children later attain working age. (51) Globally, stunting ranges from 2% in high income countries (HICs) to as high as 50% in low and middle income countries (LMICs). (49) In sSA, stunting rates in children under five years of age are estimated to be averaging 34%. (51) Similar high prevalence of stunting has been reported in Malawi. In 2017, 37% of all under five year old children were reported as being stunted, 11% severely stunted and 3% suffered from wasting (low weight-for-height). (6) The reduction in prevalence of stunting in Malawi among under five-year-old children has been reported over the last two decades. (46) In 2010, stunting prevalence among all under five year old children was 49% and this has dropped by 16 percent to 33.7% in 2020.

Stunting reduction in Malawi can be attributed to several initiatives targeted at improving preventive and curative health services for under five-year-olds over recent years. These initiatives include growth monitoring, malnutrition diagnosis and treatment, and increase in the number of facilities and health care workers involved in these services. While acknowledging progress made in Malawi in comparison to most countries in the sSA region, the prevalence of stunting remains high compared to other parts of the world. Interventions targeted at nutrition will still need to be supported by the Ministry of Health to reduce the stunting prevalence rates even further. It is important to have interventions targeting stunting early in life as evidence shows that such effects are irreversible later in life. (52)

1.2.3.6 Non-communicable diseases

Table 1.9 below show trends (whole population) of leading non-communicable diseases (NCDs) in Malawi. The past decade has seen a worldwide increase in NCD cases such as cancers, mental disorders, and cardiovascular diseases. (53) Globally, NCDs kill an average of 41 million people every year, equivalent to 71% of all deaths that occur worldwide. (54) About two fifths of these deaths occur among those aged 30 to 69 years, and 85% of these deaths occur in LMICs. It has been estimated that LMICs will continue to see an increase in NCD cases and that Africa will see a threefold increase by 2045. (55) In sSA, mortality due to NCDs is estimated to become the number one cause for all deaths by 2030. (56)

Table 1.9 Trend of leading causes of death in Malawi between 2009-2019

Disease	% change 2009-2019
HIV/AIDS	-67.7
Neonatal disorders	-21.0
Lower respiratory infection	-14.8
Tuberculosis	-4.2
Diarrheal diseases	-22.7
Malaria	-39.5
Stroke	21.4
Ischemic heart disease	23.1
Cirrhosis	9.4
Congenital defects	-12.1

source: IHME (57)

Malawi has seen an increase in NCDs from 2009 to 2019 as per Table 1.8, which currently account for 29% of all deaths in Malawi. (46) The inclusion of NCDs in the EHP is evidence that this problem has been recognised at a policy level. (24) NCDs were not part of the first EHP but due to a survey in 2009 that highlighted the acuteness of this emerging burden, they were included in 2011. (24) The most recent 2017 Malawi NCD survey revealed high prevalence of overweight and obesity, hypertension and diabetes. (58). It is likely that certain

NCDs (e.g. asthma, etc.) that occur in childhood may become more of an issue in Malawian children over the coming years.

1.3 Rationale for research on health-related quality of life in Malawi.

1.3.1 The need for economic evaluation of health packages

As shown above, Malawi has registered success at implementing many health programmes and this has led to a decline in mortality and burden of many infectious diseases. There has been a significant reduction of childhood mortality, especially amongst those aged under five years. At the same time, Malawi and sSA more generally has seen a decline of infectious disease occurrence and a surge of NCDs. (58)

Traditionally NCDs have been associated with the adult population, being affluent and living in urban areas. However, there is evidence that these conditions are becoming prevalent in the young population and in LMICs like Malawi. (58) There could be several reasons for such an increase including the levels of pollution due to cooking methods and eating habits. The fact that most with NCDs are going undiagnosed with suboptimal management of such conditions, highlights the need for strategies to curb this emerging burden of disease. Further, since neither education nor wealth are associated with reduction of NCDs implies that approaches need to be both population as well as individually based. (59) As discussed above, Malawi has a largely young population and there is a likelihood that early exposure to such risk factors could result an in increase in NCDs especially cardiovascular disease and diabetes. (59)

The reduction in mortality due to infectious disease means that more children will progress into adolescence. A surge in NCDs shows that as children grow into adolescence, a number of lifestyle factors including changing eating habits result in new disease burdens that might affect

their health. It is very likely therefore that future health interventions in adolescents focus on understanding and monitoring their health, as well as improving their HRQoL.

1.3.2 Policy and research issues

One of the greatest challenges to attaining UHC can be attributed to a lack of a standardised methodological approach to deciding what to include in the EHP. (60) Most LMICs including Malawi have relied on political promises as opposed to having clear methodologies to achieve this. A resource allocation formula, that uses cost-effectiveness analysis (CEA), has therefore been developed in Malawi to overcome this challenge. (60) The formula ranks interventions in terms of incremental cost-effectiveness, and only interventions that are cost-effective are proposed to be included in the EHP. The cost-effectiveness is determined not by external threshold but by "the detailed calculation of expected service delivery cost based on population size, disease burden, unit cost of treatment and current coverage rates." (61)

The current draft version of health financing strategy for the period 2022-2030 plans to use this same framework to determine the new EHP. (26, 61) The CEA method, however, is informative in providing estimates of incremental cost per health outcome measured in natural units, e.g. deaths averted, number of tests performed, etc. This makes it difficult to draw comparisons between different interventions or programmes that are part of the EHP and that target disparate health goals. A cost-utility analysis (CUA) framework offers an alternative to CEA and measures health outcome in terms of quality-adjusted life years (QALYs) gained or disability-adjusted life years (DALYs) averted. (62) How QALYs are derived is discussed in detail in chapter 2. Briefly, however, QALYs and DALYs are the preferential health metric for CUAs in many jurisdictions because they allow comparison between two or more different disease areas or programmes with disparate health goals. (62) Comparison and future

packaging of the EHP interventions using the CUA framework should be a viable option for Malawi.

Measuring HRQoL is important to richly describe the health of populations beyond measuring death rates, life expectancy and providing a 'baseline' for future health system changes or societal changes. HRQoL measurement can also be used for assessing health differences between groups with certain disease or injuries and the general population, and to identify vulnerable groups at risk of experiencing lower HRQoL. Preference-based HRQoL instruments can be used in different ways including patient reported outcomes, modelling studies, performance measurement/benchmarking and CEA. CEA has become fundamental in resource allocation for trial-based studies but is now becoming increasingly important in HTAs. However, the Paediatric Economic Database Evaluation (PEDE) in 2020 shows that the preferred analytical approach for economic evaluation of childhood health interventions for several HTA agencies is the use of CUA in comparison to CEA (63) and this has been the case since 2009. (64) The emergence of HTA on the African continent means that it will be necessary to have HRQoL instruments which can be used to generate QALYs and DALYs that are used in CUA. Health utility data for QALY are usually generated alongside prospective individual-level studies. (65) The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guideline recommend extraction of health utilities from published sources and these could be synthesized if appropriate. (66) Whether generated alongside individual-level studies or synthesized from published sources, QALYs and DALYs become important unit measure for economic evaluation and resource allocation priority settings.

The current EHP in Malawi has 11 disease areas, based on burden of diseases, identified as priority areas. Maternal and child health is one of these areas. It follows therefore that different

child health interventions will need to be evaluated in order to determine which interventions to include in the EHP. Additionally, instead of only focussing at reducing child mortality, policies will likely have to turn towards achieving better HRQoL outcomes using the necessary tools.

HRQoL instruments (preference and non-preference-based) therefore play a vital role in health measurement whether at individual, group or population level, as well as in informing policy. The difference between preference-based and non-preference-based HRQoL instruments is discussed in detail in Chapter 2. Briefly, though, a unique distinction between the two instruments is that preference-based instruements have associated utility values. These utility values multiplied by the time one is in a health state derives QALYs which are used in CUA. Table 1.10 below provides a general summary of the use of HRQoL instruments.

Table 1.10 Types of HRQoL instruments and how these can be used for policy purposes

HRQoL instruments	**Purpose**
Non preference-based	• Assess patient-reported health outcome • Process of care between patients and clinicians • Evaluation of medical products • Assessment of health care provision • Comparing provider performance • Methodological research in clinical studies and population health surveys • Modelling studies • Burden of disease
Preference based HRQoL	• All of the above listed in non-preference-based instruments • Utility valuation • CEA (trial-based or resource allocation)

Source: (67)

1.4 Problem statement and research question

This research was motivated by improvements made in reducing childhood mortality, discussed above, and the opportunities and challenges that lie ahead. These opportunities and challenges include informing health care policies in Malawi on the quality of life of the surviving children and adolescents. As a result, the population is increasing, about half being below 15 years of age, and at the current growth rate this is projected to double by 2038. (4) A large percentage of this population will be children and adolescents growing into adolescence and adulthood, respectively. At the same time, evidence suggests that young people will also be affected by emergence of NCDs. It is likely that the government will turn its focus at not only measuring health survival but also HRQoL. A CUA framework will be key in determining which interventions are viable for inclusion as part of future EHPs. Measuring HRQoL and performing CUA will require robust preference-based instruments. Some health economists argue that the right terminology is 'preference accompanied measures' (68) Throughout this thesis the former term, i.e., preference-based, is used. Whatever the right terminology is for this, preference-based HRQoL instruments will become important in Malawian settings.

Doctorate Research question:

How to evaluate (adaptation and validation) preference-based HRQoL instruments to support economic evaluations of health interventions targeting children and adolescents in a Malawian setting?

1.5 Purpose of this research

As discussed up to this point, HRQoL instruments are designed to meet different health demands. The thesis will explore the extent to which existing childhood preference-based HRQoL can be adapted and used in a Malawian context. In line with Table 1.10 above, preference-based HRQoL instruments can be used in economic evaluation and priority setting of health interventions. However, before using existing childhood preference-based HRQoL instruments in settings other than where these were developed it is important to establish that they measure the intended construct and perform as robust in new settings. This thesis therefore sets out to explore the following:

i- How health is conceptualized in a Malawian context and whether there is a difference of health concept in comparison to existing instruments

ii- Evaluate the cross-cultural adaptation process of existing instruments in new settings

iii- Examine the implications of using existing preference-based instruments in policy and research

iv- Explore the implications of existing instruments in a wider sSA context.

1.5.1 Overall aims

The broad aim of this thesis is to explore existing preference-based HRQoL instruments used in children and adolescents and assess how they have been adapted for use in sSA settings and whether this process aligns with local perceptions around health.

1.5.2 Objectives

The objectives are:

i. to carry out a literature review of existing childhood preference-based HRQoL instruments (Chapter 2)

ii. to establish whether the identified existing childhood preference-based HRQoL instruments were adequately designed (Chapter 2)

iii. to establish reliability, validity and use of these instruments in sSA (Chapter 2)

iv. to provide justification for a mixed methods approach required for establishing linguistic and conceptual equivalence and validation of existing instruments (Chapter 3)

v. to adapt selected instruments (EQ-5D-Y, EQ-5D-Y-5L, and PedsQL^{TM} 4.0 child and teen self-report) by establishing linguistic and conceptual equivalence for use among children and adolescents in Malawi. (Chapter 4)

vi. to psychometrically validate two of these instruments in Malawian settings (Chapters 5 and 6)

vii. to develop an in-depth understanding of how health is conceptualized in Malawi and establish whether the existing instruments align to perceptions of health in this setting (Chapter 7)

viii. to discuss the overall findings in relation to policy implications and areas for further research (Chapter 8)

1.6 Significance of the study

Health economic evaluations are becoming common practice in LMIC settings, including Malawi. (26, 69-72) Many countries in LMIC settings are setting up health technology agencies (HTAs) for evidenced based information to inform policy makers in resource allocation. Furthermore, researchers in a few countries in Africa, have developed utility tariffs for carrying out CUA. (73, 74)

At a local level, the Malawi Ministry of Health has recently set up a Health Economics Policy Unit (HEPU) in collaboration with Kamuzu University of Health Sciences (formerly Malawi College of Medicine) and UK academic partners. (75) The objective of the HEPU is to be a scientific hub for providing evidence to inform policy makers for resource allocation decisions. Importantly, as Malawi has made much progress in mortality reduction, there will be needed to evaluate health in its broader sense. The packaging of EHPs in pursuit of UHC has included and will continue to include interventions targeted at improving child health considering SDG goal 3. Evaluation of health interventions firstly requires establishing how health is understood in the Malawian context. Further, appropriate tools will be required to undertake CUAs in the context of EHP packaging. There is therefore potential for use of CUA for resource allocation decisions, which requires adequate tools. It is expected that findings from this study will be informative to policy makers as well as the scientific community. This might also assist decision makers in using HRQoL as an outcome measure to evaluate inequalities that exist as well as in shaping future health interventions through EHP provisions. Additionally, the provision of appropriate tools also offers opportunity for objective CUA and health care assessment in children in Malawi.

1.7 Overview of doctorate thesis

While overall mortality rates are declining across sSA, specifically among children, this research set out to examine how to address the quality of survival at early age, yet also setting the scene for lifetime health. As outlined above, the success of child mortality reduction created by health interventions and other strategies in Malawi is likely to translate into an increase focus on the measurement and valuation of HRQoL outcomes for informing resource allocation decisions. Lifestyle habits too might be increasing the potential for NCDs. This will likely

result in need for improved understanding and measurement of HRQoL outcomes. In research, this will require appropriate HRQoL tools to generate evidence that can be used for priority-setting decisions for related health programming. This doctorate contributes to increased understanding of health among children and adolescents in Malawi. Particularly, it focusses on health measurement options relevant to economic evaluations in the broadest sense in both clinical and public health sciences.

The thesis is divided into eight chapters. This chapter has provided an overview of the country of Malawi, its health care system as well as issues around child health. An overview of how HRQoL evaluation is related to child health is then presented. Evaluating HRQoL requires robust instruments (health status and preference-based HRQoL measures) to effectively assess interventions and reliably inform policy makers. The chapter has also outlined the problems for decision makers in the use of HRQOL measures and provided a motivation for this research. Finally, the main research question for this doctorate as well as the main aims and objectives were highlighted.

Chapter 2 presents a review of childhood preference-based HRQoL instruments. The chapter introduces what preference-based HRQoL instruments are and how they are developed and used in health economic studies. This introduces the concept of psychometrics, and psychometric criteria used to assess the performance of these HRQoL instruments. Findings of the systematic literature reviews undertaken are described here. The literature review looks at existing childhood preference-based HRQoL instruments, where and how these have been developed, and how they have been adapted and used in sSA settings. The chapter finishes with the selection of the EQ-5D-Y for adaptation in Malawi using a scoring criterion that was adapted for this purpose. The EQ-5D-Y-5L is a five-response version like the three response

EQ-5D-Y version. It was deemed appropriate that the EQ-5D-Y-5L be adapted and evaluated together with the EQ-5D-Y. Further, in order to evaluate the construct validity of the EQ-5D-Y (and EQ-5D-Y-5L), there was need to draw comparisons with another HRQoL measure. The PedsQL™ 4.0 self-report was chosen for this purpose.

Chapter 3 introduces the mixed method approach adopted and provides a rationale for choosing this approach for cross-cultural evaluation of an existing measure developed elsewhere. The chapter discusses methodological issues that arise in cross-cultural evaluation of existing instruments in general, and for Malawi in particular. The chapter also highlights why psychometric evaluation is important and the different approaches used to do this. Additionally, the chapter discusses the strengths and limitations of the mixed methods approach. Finally, the chapter provides a flow diagram of studies that are contained in this thesis.

Chapter 4 describes the linguistic and conceptual equivalence work done for the EQ-5D-Y and EQ-5D-Y-5L in a Malawian setting. This was deliberately done separately for the two instruments. The EQ-5D-Y was evaluated first because there was ongoing work on the beta EQ-5D-Y-5L by developers. The chapter then provides evidence of elimination of qualifier inversion in establishing linguistic and conceptual equivalence for the EQ-5D-Y-5L. This chapter also presents the linguistic and conceptual equivalence findings for the PedsQL™ 4.0 Child self-report and the PedsQL™ 4.0 Teen self-report in a Malawian setting. The PedsQL™ 4.0 is a generic health instrument and was used alongside the EQ-5D-Y in order to assess the construct validity of the latter.

Chapter 6 presents the psychometric evaluation for the PedsQL™ 4.0 Child self-report and the PedsQL™ 4.0 Teen self-report in a Malawian setting. Before being used for construct validity,

it was first necessary to establish psychometric validity of the Chichewa version of the PedsQL™ 4.0 version. The chapter provides these findings.

Chapter 6 provides the psychometric performance of the EQ-5D-Y and the EQ-5D-Y-5L Chichewa versions in Malawi. The comparison is presented first at an item level and then at a scale level. Item level analysis examines how an instruments' items behave in relation to missing data, maximum endorsement frequency (ceiling or floor effects), as well as examining the inter-relationships and redundancy between items. At a scale level, results are presented of how both the EQ-5D-Y and EQ-5D-Y-5L performed in terms of reliability and validity. The psychometric evaluation is necessary to establish if the instruments are fit for use in Malawian settings. The chapter concludes with the findings and presents future recommendations for use of the EQ-5D-Y and EQ-5D-Y-5L in Malawi.

Chapter 7 presents findings from the qualitative work that sought to understand how Malawian children and adolescents view health. Firstly, the chapter gives a background to the conceptual development of HRQoL scale measurements. The chapter develops a conceptual framework for children in Malawi with concepts that were identified as important in this setting. This developed conceptual framework is then compared with the content of the existing EQ-5D-Y (and EQ-5D-Y-5L) measures to evaluate the content validity of the latter. The comparison is necessary to establish if the EQ-5D-Y contains all relevant dimensions to evaluate health in a Malawian setting. The chapter further comments on the early developed childhood health measures in relation to the conceptual framework.

Finally, chapter 8 presents an overall discussion of this thesis findings. A summary of main findings from each chapter, including strengths and limitations, is presented. The chapter ends

with the policy implications of these findings, how this work adds to the body of knowledge and the avenues for further research in this area.

1.8 Summary of chapter 1.

This chapter started by introducing the setting where this current research took place. The demographics as well as health priority issues pertaining to child health were presented. Childhood population has substantially increased in size and about half of the Malawian population is below adolescence age. This can be attributed to many factors including a decline in mortality rates of the whole population, but especially that of children. The continued population increase means that late childhood illnesses, e.g. asthma and diabetes, will likely become more common. NCDs are expected to continue to increase and this increase will be disproportionately felt by LMICs with a threefold increase in Africa alone. It is very likely that policy makers will need to consider what services are to be provided first to this population in view of health budget constraints. The EHP in Malawi has been adopted to provide leverage in the delivery of UHC. There have been notable challenges around how to package the EHP and one recommendation is to use CUA to identify disease areas and intervention coverage. Subsequently, Malawi has made steady progress to provide evidence to inform policy by establishing HEPU. CUA will become increasingly important for informing intervention coverage and for allocating constrained health care resources. Ultimately, there will be need to also evaluate the health impacts of interventions vis-à-vis provision of UHC. The problem statement as well as motivation for doing this research highlighted the importance of doing this work. The chapter then highlighted the need for generic childhood preference-based HRQoL instruments for EHP packing and for assessing the impacts of health interventions in pursuit of UHC. Finally, the chapter outlines the thesis aim and objectives and provides a link to each chapter that achieves each of the objectives. This chapter ends with a brief summary of all the

subsequent chapters.

The next chapter introduces childhood preference-based HRQoL instruments, how these have been developed and used in health economic studies.

2 Chapter 2-Background and literature review of childhood preference-based health-related quality of life measures in sub-Saharan Africa

With:
Ngwira, L.G.; Khan, K.; Maheswaran, H.; Sande, L.; Mipando-Nyondo, L.; Smith, S.; Petrou, S.; Niessen, L
VIHRI 2021

2.1 Chapter overview

Chapter 1 introduced the topic of child health and the different interventions that have improved child health in Malawi and across the sSA region in light of the SDGs. The focus on attaining SDG goal 3 means that health programmes will not only aim to reduce childhood mortality, but also have a positive impact on children's HRQoL. Economic evaluations are being carried out to provide policy makers additional information to inform decisions around health service delivery. Appropriate and validated HRQoL instruments will therefore be required to generate relevant evidence that can inform these policy goals and actions.

This chapter will review and discuss instruments aimed at measuring HRQoL that can generate outputs for CUA based decision-making. Health economists often refer to such instruments as preference-based HRQoL measures or patient reported outcome measures accompanied by preference-value sets. The chapter will begin with a general description of preference-based childhood and adolescent measures, what they consist of and how they have been developed. The chapter further discusses where these have been developed, their psychometric properties, as well as their use in the sSA region. The chapter ends by presenting a selection process for instruments that were cross-cultural adapted (and validated) among children and adolescents in Malawi for this doctorate research.

2.2 Introduction

There are a plethora of definitions of 'health'. (76) One such definition is given by the WHO as "a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity". (77) From this definition, health is understood to be not just be about specific symptoms related to the disease, but also that it encompasses many other factors including physical, mental and social functioning. Among the different constructs related to health are QoL and, more specifically, HRQoL, as introduced first in chapter 1.

As is the case with health, HRQoL has also taken many definitions, (78-80) often depending on study objectives and related available information. The overarching idea is that HRQoL is a multidimensional construct that incorporates a broad range of health-related outcomes and is focused on the subjective perception of these by the individual. (81) The WHO defines HRQoL as "individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns". (82) Broadly, HRQoL may have similar attributes to QoL but the former principally focuses on how these affects someone's health. The WHO definition summarises HRQoL into five categories: duration of life, impairments, functional states, perceptions and social opportunities. These categories are health-related because all of them can be affected by disease, injury, treatment or policy. (83) Other HRQoL definitions have been constructed on the basis of 'within the skin' concepts, i.e., excluding social interaction for example because it occurs outside the body. (84)

The dawn of the second half of the twentieth century, particularly later in the 1990's, witnessed an increase in measurement of HRQoL among children. (85) This increase in development and use of childhood HRQoL measurements has been due to several factors. One such reason is

that in recent years health outcome goals have shifted from reduction in morbidity and mortality to an attainment of overall higher levels of HRQoL. While QoL has a broad range of constructs that include goals and aspirations, living conditions, spirituality, political, and financial situations (86), HRQoL focusses on how these affect health. (76) In health related outcomes, the interest is assessment of the effects of disease and intervention on the child's well-being. (81) HRQoL therefore moves further than the traditional clinical outcome of efficacy and safety. There has also been a growing need to measure and compare interventions in terms of these wider HRQoL outcomes and their associated costs. (87, 88) This comparison potentially enables policy makers to efficiently allocate resources, and this has particularly been the case in HICs. (89) This can contribute to better impact health policy decisions. However, measuring HRQoL requires design of standardised, valid and generic health outcome instruments.

HRQoL can be measured using preference-based and non-preference-based instruments, measures, scales or tools (these latter terms are used interchangeably in this thesis). It is also important to note at this point that while in most literature, the terms preferences, tariffs, utilities, scores and values are often used interchangeably, these are technically different. (62) Preference is the 'overall umbrella' or process by which people (patients, clinicians or the public) score a certain health state. Utilities and values on the other hand are the resulting outcome scored in this way. (62) Further, utilities and values differ in their elicitation techniques. The latter are generated under certainty conditions as opposed to uncertainty conditions for the former. (90) Certainty conditions implies that the decision-makers may have a reasonable idea about an outcome compared to uncertainty which is purely by chance. In this thesis, as is common with most published literature, the terms preference weights, tariffs, utilities, values and utility values are used interchangeably. Non-preference-based instruments

measure health without drawing upon the preferences of the population for aspects of health. Often, therefore, non-preference-based instruments are used as outcome measures or endpoints to measure effectiveness.

Preference-based HRQoL measures, on the other hand, "assess patient preference across broad areas including symptoms, physical functioning, work and social activities, and mental well-being." (91) and are generally used for health economic evaluations. Preference-based HRQoL measures are unique in that they have both a descriptive system and a scoring system for health states. Health states are generated from combinations of responses (also referred to as severity levels) on the descriptive system and can be quantified on a scale indexed at 1 (representing perfect health) and 0 (state equal to death). How this is done is discussed in detail in section 2.3.2 below. This conversion allows HRQoL and length of time to be combined into a QALY, a concept introduced in chapter 1. However, the QALY is not the only outcome measure for economic evaluation. The DALY, another concept introduced in chapter 1, healthy life equivalent (HYE) and health-adjusted life expectancy (HALE) have similarly been developed for the same purpose. (92) Of these, DALYs have over the years emerged as the most widely used measure for economic evaluation in LMICs. (93) DALYs prominence in these settings can be attributed to the fact that they try to address burden of disease which most LMICs are grappling with.

With this in mind, in the 1990s the WHO and World Bank collaborated to quantify the burden of premature death, disease, and injury particularly in LMICs. (94) The burden was assessed by calculating DALYs which is a sum of the years of life lost (due to premature death) and years lived with a disability. (95) DALYs therefore measures the gap between a hypothetical ideal health achievement and the actual population's health. The goal is to reduce this

difference, since one DALY represents 1 healthy life lost, meaning lower DALYs are better. DALY weights are disease specific and these were initially obtained from healthcare professionals using person-trade off technique (discussed in detail in section 2.3.2.2 below). The approach for calculating DALY weights has changed over the years as the most recent ones were estimated using general population surveys. (96) On the other hand, QALYs are estimated differently and this is also discussed in detail in section 2.3.2 below. The advantage of both DALYs and QALYs, is that they can be used in CUA to compare different health interventions. (97) This comparison of interventions across different disciplines using either QALY or DALY as a single outcome measure positions CUA ahead of other methods of economic evaluation. (62) This aid policy makers in resource allocation within health but also other areas.

Malawi is an LMIC, and it may be expected to use DALYs in economic evaluation and for priority setting resource allocation. This is the case as up to this point the EHPs have been designed from CEA using DALY as a measure of interest. However, as discussed in chapter 1, the decline in childhood mortality and increase in non-communicable diseases such as obesity will likely propel measurement of length and quality of life as the children grow into adulthood. Future childhood health care interventions might lean towards using QALYs in a setting such as Malawi. Preference-based HRQoL instruments plays an important role in generating QALYs, which are used in this way. This is the justification of the interest in preference-based HRQoL measures which are discussed in this chapter. The chapter is outlined to addresses the following four objectives:

i) to describe how the content of generic preference-based HRQoL instruments are developed

ii) to describe the psychometric properties and criteria used for evaluating these instruments

iii) to present a literature review of childhood generic preference-based HRQoL instruments

iv) to assess identified childhood generic preference-based HRQoL in terms of psychometric and other practical properties in order to select instruments for use in the Malawi setting.

2.3 Preference-based health-related quality of life instruments

2.3.1 Types: Generic v disease/condition-specific instruments

Disease or condition-specific instruments have been developed to assess HRQoL for specific health conditions and are used in both observational clinical studies and randomised clinical trials. (98) Generic instruments have been developed to assess HRQoL in healthy and sick populations, with any health condition. These instruments also have value in general population surveys.

There are various advantages and disadvantages of using generic against condition-specific instruments (Table 2.1) but much depends on the objective of the particular research study. (83) Generic instruments have the advantage of evaluating the impact of interventions in populations with different health conditions, making them easy to compare. The downside is that generic instruments are generally less sensitive to health conditions. On the other hand, condition-specific instruments tend to be more sensitive to detecting change(s) in the health-status of people affected by that particular condition. (99) The drawback being that condition-specific instruments cannot be used in those with very different health conditions to what the tool was developed for. (100) For example, a condition-specific tool for back pain, may ask

detailed questions about the impact of pain i.e., lower pain, intensity and frequency of pain, mobility, changes in sensation, urinary incontinence. These types of questions are meant to capture all the consequences of back pain and be able to better detect improvements or worsening over time or with interventions. There would be no benefit therefore in using this back pain specific tool to measure HRQoL in patients with heart disease, as the latter experience different symptoms altogether.

Traditionally it has only been possible to develop QALYs from generic instruments, though there are now several disease specific instruments for which utility scores can also be derived. (101) Additionally, through such methods like mapping condition-specific instruments to generic preference-based measures, utility values can be generated. (102) The use of vignettes to directly derive utilities from patients (or the general public) is yet another option. All this goes to show that it is possible to generate utility values from condition-specific instruments and use these to make comparisons across different health interventions. The points raised above though highlights the complexities of generating utility values from condition-specific instruments. Further, condition-specific instruments can also be expensive to develop as it takes considerable time and resources to design and validate instruments for specific conditions.

Table 2.1 Generic and condition specific measures

Property	Preference-based		Non-Preference-based	
	Generic e.g. HUI2	Condition-specific e.g. EORTC-8D	Generic e.g. SF-36	Condition-specific e.g. SGRQ
Use	Applicable across different disease conditions and healthy	Specific to a disease condition and therefore more clarity	Applicable across different disease conditions and healthy	Specific to a disease condition and therefore more clarity
Administration	Shorter therefore less costly to operationalize.	Tend to shorter and therefore reduced time to administer	A bit longer and therefore costly to operationalize	Tend to be long and may take more time to administer leading to fatigue to responder
Utility values	Has utility values and therefore can be used for cost-utility analysis	Generally, not available although a few instruments have	Does have profile score which can be used for cost-effectiveness analysis	Not intended to have utilities but have profile scores
Sensitivity	Less sensitive to small changes in health conditions	Most do not have utility values although through mapping or bolt-on utility values can be converted from generic measures.	Generally, has more items (therefore has more content validity than) preference based	No utility values available
Relevance	Some questions lack relevance to a specific condition	Costly to operationalize since they are specific to a condition (lacks generalizability)		

Source: Developed by author using data from other sources. (102, 103)

EORTC-8D: The European Organization for Research and Treatment of Cancer; HUI2: Health Utility Index Mark 2; SF-36: Short Form 36 items; SGRQ: St George's Respiratory Questionnaire

2.3.2 How generic preference-based health-related quality of life instruments are constructed

Generic preference-based HRQoL measures, also known as multi-attribute utility instruments, incorporate two processes: firstly, participants complete the descriptive system and second, translating their responses into a health utility score. (104) How the descriptive system and utility values are generated is described in detail below starting with the former first.

2.3.2.1 How descriptive systems are developed

The descriptive system is made up of a wide range of HRQoL dimensions, domains or attributes, and in this thesis these terms are used interchangeably. The dimensions, which capture different aspects of HRQoL in the descriptive system are made up of single or multiple items, also known as questions. The descriptive system is usually developed using two approaches: composed/ holistic and decomposed.

In a composed approach, items are generated in real time. Items or questions are directly constructed through interviews with a sample of the targeted population or vignettes. (104) The idea for this approach is to have an expansive descriptive system. (105, 106) Some researchers argue that this method provides a richer profile compared to one generated from an already structured descriptive system. Despite this advantage, the composed approach is rarely used in practice to formulate the descriptive system. Instead, a decomposed approach to come up with a descriptive system has widely been used.

In a decomposed approach, the descriptive system items and dimensions including utility values are already determined by previous development work. This contrasts with a composed approach where the descriptive system formulated in real time from the target population. There are two main methods for formulating the decomposed descriptive system items and

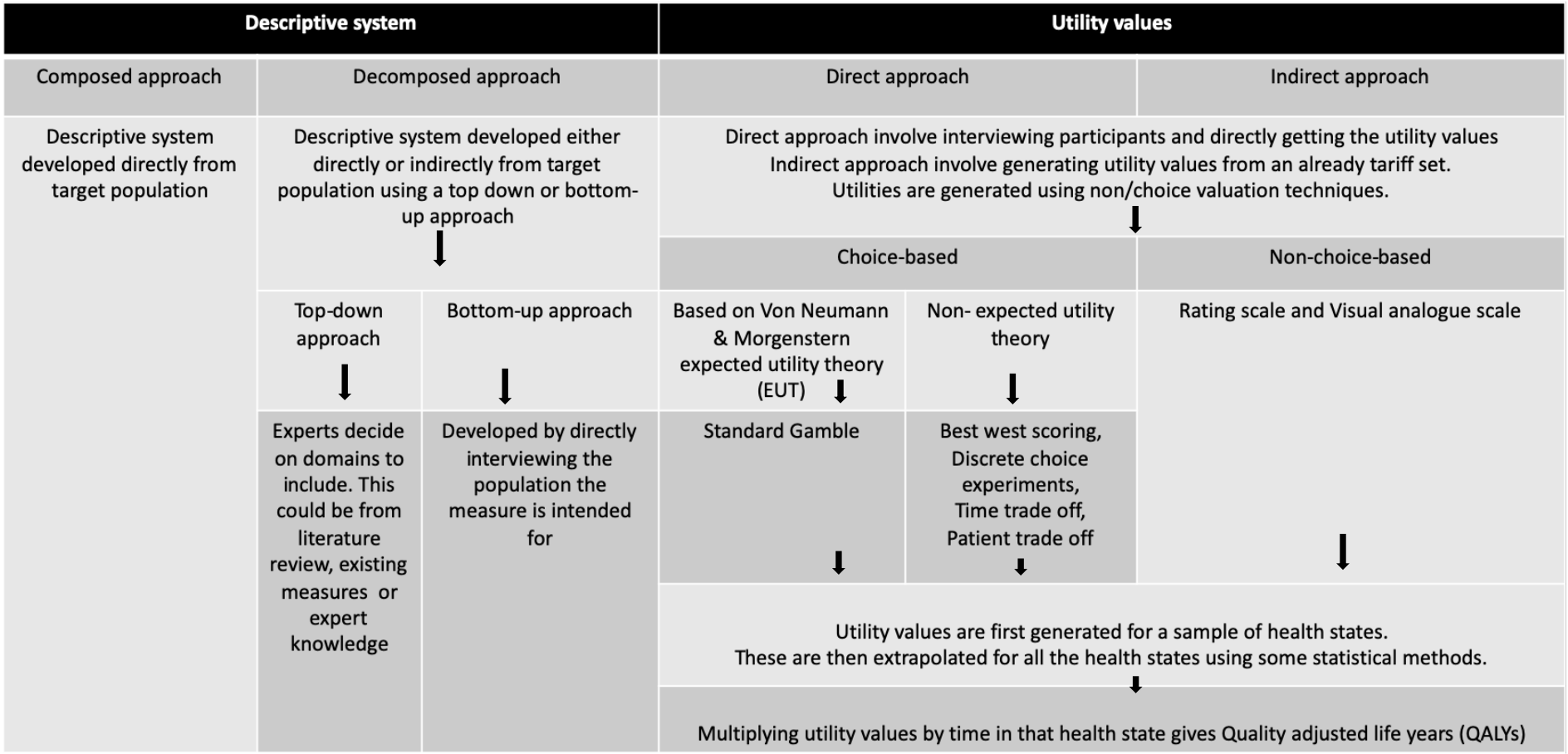
dimensions, top-down and bottom-up approaches. The top-down approach involves developers reviewing the literature, gathering expert opinion and undertaking national surveys to determine the health domains. (91, 107) This may also involve consulting the targeted population to refine or pilot the descriptive system. Generally, the top-down approach is considered to be researcher-driven with little involvement of the target population. Even with their involvement, the final decision on instrument dimensions to retain is with developers.

In comparison, the bottom-up approach starts with qualitative research (e.g., in-depth interviews; focus group discussions) with the target population. In a way this is similar to the composed approach stated above but the difference here is that whereas the former is done in real time, the decomposed bottom-up approach will formulate the descriptive system well in advance. The qualitative analysis of the interviews and focus groups informs the conceptual framework choice of domains to retain. The advantage of the decomposed bottom-up over the top-down approach is that the target population determines the relevant dimensions without 'imposing' it on them as is the case with the latter. Despite this, most childhood preference-based HRQoL instruments have been developed using the top-down approach. (107) Various reasons have been cited for using this approach including conceptual and methodological challenges in developing preference-based instruments for children. (108) While this may be true for childhood preference-based measures, these methods have been used many times in developing both generic and disease specific non-preference measures for children. (109) Further, both the decomposed approaches (top-down or bottom-up) have not escaped their share of criticism. For example, the decomposed approach is said to be restrictive in formulation of the descriptive system compared to the composed approach. (110)

2.3.2.2 Generation of utility values

Through the different combinations of the descriptive system responses, corresponding health states are generated. The generated health states make up what is known as the health state classification system. Each individual health state has a corresponding unique utility value, such that the health state classification system (all possible combinations of the responses from the descriptive system) has corresponding health state utility values. (84) The utility values are usually anchored on a scale indexed at 0 representing a health state equivalent to death and a maximum of 1 equivalent to perfect health. Negative values are often available for states worse than death and this is well known conceptual problem. Utility values are generated either through a direct method or an indirect method. Figure 2.1 below shows a summary of how the descriptive system and the utility values can be generated.

Figure 2.1 Overview of how the descriptive system and utility values are developed



Source: based on own synthesis.

2.3.2.2.1 Direct methods for obtaining utility values

The direct method, also known as the holistic approach, works more or less in real time. This involves asking patients directly not only to describe their health status, but also to value it using a complex scaling technique such as the standard gamble approach or time trade-off approach. In reality, the direct approach to generate utility values has rarely been used. (106) There are several reasons cited for this including ethical issues. It is argued that it would be unethical to directly ask a sick population to undertake valuation tasks that include death to generate utility values. This particularly being the case among patients with a chronic condition. (111)

2.3.2.2.2 Indirect methods for obtaining utility values

The indirect method involves converting the responses from the descriptive system using previously developed utility values. As shown in the descriptive system above, depending on the number of dimensions and items covered, this can produce many health states. It is therefore practically impossible to generate utility values directly for each health state from the population. Instead, utility values are firstly derived directly from the population for a sample of health states using any of the utility valuation techniques in Figure 2.1 and described below. (112) Using statistical methods, utility values of a sample health state are extrapolated for all possible health states. (105) The extrapolated utility values for each health state then become predefined off-the shelf utility values for that instrument.

Valuation techniques

There are two types of utility valuation techniques for generating utility values from a sample of the general population: choice and non-choice based. (62) Examples of the latter include rating scales (e.g., visual analogue scale (VAS)) and ratio scales. The former includes time trade-off (TTO), standard gamble (SG), person trade-off (PTO) and more recently, discrete

choice experiments (DCEs) and best worst scaling (BWS). These different types of valuation techniques are explained in detail below.

Choice-based techniques

There are mainly five common choice-based techniques for generating utilities: SG, TTO, PTO, DCEs and more recently, BWS. The SG valuation technique is rooted in the von Neumann-Morgenstern axioms of expected utility theory. (113) The assumption is that rational people will always choose an option that enables them to maximize expected utility. The SG involve a recipient making a choice between two alternatives- i) one under a certainty condition e.g. living under present conditions for the remainder of one's life, or ii) an intuitive choice between the uncertainty of perfect health (p) or immediate death (1-p) until the recipient is indifferent between the choice of death and perfect health. The probability, p, is then the utility attached to the present state of health. The SG has the advantage of being rooted in uncertainty conditions, which resembles most medical conditions. (114) Further to this, the assumption of certainty conditions is not representative of real-life choices since one cannot say for certain how long they would live. (114, 115) Another major drawback with the SG technique is the complexity required to comprehend and perform probability tasks by recipients. (115)

TTO was developed by Torrance due in part to overcome complexities associated with the SG technique. (116) TTO involves asking a recipient to give up some of the time (could be life expectancy time) for a different health state (usually perfect health). The premise for a TTO is about improvement of health in terms of amount of life expectancy that an individual is willing to trade off to attain perfect health. (117) The duration of the health state is varied until the recipient is indifferent between the two choices, at which point the preference value is determined. For example, a person in present health state (t) who gives up some of their life-expectancy time in order to attain perfect health (x) would derive a utility score of x/t. (118)

While TTO involves choice and is relatively simpler to complete compared to the SG, (116) the outcomes are not rooted in uncertainty (probability) utility conditions.

DCEs, also referred to as conjoint analysis (118), have gained popularity as a valuation technique in the health care context in the last two decades, (119) although their use in economics and psychology market research dates back to early 1960's. (120) Unlike TTO and SG, DCEs do not involve an iterative process to identify at which point a person is indifferent. Rather a participant is presented with two or more choices, each choice made up of different attributes, and is asked to pick one. (119) The technique works under the assumption that any intervention is described by attributes and varying these result in an individual picking a choice. (118) The attributes are varied, and individual choices are based on values individuals attach to those attributes. The choices are then aggregated to inform which attributes are valued more by a targeted population.

The BWS technique is a more recent valuation technique, which in a way is an extension of DCEs. BWS is rooted in random utility theory which takes into account decision makers' tastes and behaviours in respondent choices. The BWS technique, like DCEs, asks participants to choose between two lists of attributes, but this is in a form of an importance continuum. (121) The list of BWS attributes are thus compared between the most (best) and least (worst) attractive option, and respondents choose all the pairwise comparisons. The aggregated difference between the best and worst options represent the maximum utility difference for all the attributes.

The PTO works on the premise that one is not rating their own preference whether personally experiencing or imagining to be in that health state. Rather, PTO involves asking people how

many outcomes they would consider giving up in order to gain outcomes for other people in another kind. (98) This technique is not often used in economic evaluation as it does not involve individuals trading their own preference. However, the PTO has extensively been used in generating DALY weights. The weights (initially from a few health professionals but now from general population) involve asking respondents if they would prefer to save 1 life year for 1000 healthy people or that of 1 life year of more people with a less than perfect health. For example, if 1,000 healthy people claim some resources as that of 7500 with severe pneumonia would give a weight (due to severe pneumonia) of 0.133 (1000/7500). (86) The utility value (equivalent to QALY) can be estimated by subtracting 1 from the DALY weight which would be 0.867 (1-0.133).

Despite these differences, a common thread that runs through choice-based valuation techniques is that they must meet the following three assumptions: i) reflect the preferences of individuals, ii) utility values should be measured on an interval scale, and iii) the scales should be anchored to perfect health and death. (62)

Non-choice-based techniques

Non-choice-based tariff techniques include rating scales (RS) and ratio scales or magnitude estimates (ME). (98) RS involve people marking on scale with or without discrete intervals from the lowest (dead or worst health state) to the highest level (perfect health). An example of this is the VAS of the EQ-5D (EQ VAS), with discrete intervals like a thermometer scale ranging from the lowest (equivalent to worst possible health state) to the highest level (perfect health). A subject picks a point on the scale and draws a line that corresponds to how they feel their health is on that day. ME, on the other hand, is about comparing health states in terms of ratios - such that respondents compare disutility (undesirability) as a function of disutility of

the reference state. (106) For example, if health state B is x times worse than being in state A, being in state B is said to be x times more undesirable as A.

Irrespective of the approach used, direct or indirect, choice or non-choice-based, the different elicitation techniques generate different results. (104) There are therefore several advantages and disadvantages for using one technique against the other. (91) Importantly is the fact that among the most prominent elicitation techniques of RS, TTO and SG, (122) the latter two are preferred. (104) One reason for this is that the RS does not offer respondents sacrifice and hence does not really qualify as a preference-based technique. (104) Furthermore, RS utilities have been found to be substantially different from those derived by two of the choice-based methods (TTO and SG). (116) Of the two, TTO is arguably the preferred elicitation preference technique due to its relative simplicity, (119) while the SG is generally considered the theoretically preferred utility method under uncertainty conditions. (113)

Statistical methods for extrapolation of utility values

As already discussed, due to the large volume of health states generated by descriptive systems, utility values are generally generated for a sample of the health states first. These are extrapolated to the rest of the health states using statistical methods. There are mainly three statistical methods for doing this, (105) although other methods beyond these exist. (123)

Additive methods use statistical regression-based methods like generalized least squares and maximum likelihood estimates. (105) The assumption for these methods is that there is no interaction between dimensions. The EQ-5D is one such example that uses this approach to extrapolate all utility values for all possible health states (243 for the 3 response (or severity) level and 3,125 for the 5 five response level). In the EQ-5D method, for example, the perfect

health state is equal to 1 if all dimensions of health are equal and are scored 1 otherwise for all the other health states.

The second approach is the multiplicative method which assumes some interaction for its multiattribute utility function/theory (MUIF/T). (105) An example of this is the Health Utilities Index Mark 3 (HUI3). The HUI3 perfect health state has a value of 1. All the other health states utility values are obtained by multiplying the values that each health domain takes.

The third method is a multilinear method that assumes a different type of interaction -a combination of the additive and multiplicative methods. (105) The utility for a health state is obtained by adding the values that each domain takes multiplied by the interaction of domains. An example of this is the Assessment Quality of Life (AQoL)-8D.

Whose preference should be used to generate utility values

There are different arguments for or against deriving preferences from patients, clinicians or the general population. (113) The argument for patients is that they are better informed to accurately provide preferences for a particular health state. (124) It has however been observed that patients, for example, tend to accommodate their limitations, goals and expectations to their 'new' condition or state. (98, 125, 126) As such some argue that patients tend to over-rate or under-rate preferences. (127) Others contend that the general population is better placed to provide preferences. (104, 128) It is argued that most health systems are funded by the general population, through taxes, and as potential patients, they should have a say on which interventions should be funded. (129) Ultimately, preference weights obtained from these different groups also lead to different values, higher or lower and at times similar. (98)

Deriving QALYs for cost-utility analysis

Once a participant completes the descriptive system, this produces a health state which has an associated unique utility value. Multiplying the utility value by the time spent in that health state generates a QALY value. QALYs are therefore a product of the length of time in a health state and preference weight(s) to give a single metric. (113) It is this single metric that allows comparability between two or more health interventions or programmes in terms of cost and effectiveness i.e., CEA or, more specifically, CUA- concepts introduced in chapter 1. As an example, someone living 10 years after an intervention with utility of 0.7 will generate 7 undiscounted QALYs. On the other hand, the same person could have lived 10 years without intervention with a utility value of 0.4, which would have generated 4 undiscounted QALYs. The difference of these QALY estimates gives the utility gain due to an intervention. In turn, each of these strategies has associated costs. The differences of costs divided by the difference in QALYs from two or more interventions generates an incremental cost-effectiveness ratio (ICER). ICERs are generated for all non-dominated mutually exclusive and independent interventions. These are then compared to a predefined threshold, which represents the opportunity cost of displacing other treatments, to determine interventions that are said to be cost-effective. (62)

2.3.3 Childhood issues in development of preference-based health-related quality of life instruments

This chapter has introduced how generic preference-based HRQoL instruments and corresponding utility values are developed and generated respectively. The development of childhood preference-based HRQoL instruments has lagged adult instruments (105) due to several conceptual issues. The issues include development of descriptive systems, who completes the descriptive system, direct against indirect utility valuation techniques, and the use of adult utility values in children. (126, 130)

Children particularly of younger age, are generally less able to communicate articulately and express perceptions of their health. (131) The design of instruments targeting this younger age group has therefore been challenging. While some have argued that children (>7 years) are able to properly express their own perceptions of health, others are of the opposite view. (130, 132) Elsewhere it has also been noted that children hide their true emotions and feelings from parents or carers. (133) This raises the question of who is better placed to complete the descriptive system, children or their parent proxies. Even when proxies are used to report on behalf of adults there is sometimes not good agreement. (134) The issue being that self-reports from proxies don't always agree and those from children may be less reliable depending on the age of the child. (135) Further, questions have also been raised regarding children's abilities to express their own preferences. (136) Comprehensibility issues of the utility valuation techniques have been found difficult even among adults and using these techniques would be extra difficult in younger children. For this reason utility values have generally not been developed specifically from younger children, (137) even though this has worked in adolescents. (138) Ethics has also been singled out as an issue since this would mean asking children to gamble death (in SG) for example. (137) This has left users with the option of applying adult proxy utility values in children. However, the utility values generated from children differ from that of their parent proxies. (130) Effort is being made to generate utility values directly from among children and adolescent population using less demanding valuation techniques. Through such valuation techniques as BWS, DCEs and other ranking scales, utility values are being generated directly from children. (139, 140) Still, there are concerns whether the utility values from children would be appropriate for decision making. (141)

Despite the challenges to developing instruments directly from children and generation of utility values, the past two decades have continued to see an increasing interest in the

development of childhood HRQoL instruments. This comes against the background that only the child is best placed to express their own health. (142) In addition, there has also been a need to develop measures in children in the long-term. This has led to development of both preference and non-preference-based instruments to measure HRQoL among children and adolescents. The development of descriptive systems and utility values as discussed above are only two steps required to achieve preference-based HRQoL measures. Another fundamental step in every instrument development is the evaluation of its psychometric properties.

2.4 Psychometric properties: definitions and criteria

Psychometric is made up of two terms: psych which means 'mind' and metric referring to some sort of 'measurement'. The development of psychometrics and the different theories are discussed in detail in chapter 3. Briefly, psychometric properties refer to the instrument's ability to quantify how people self-describe health using existing questionnaires. Most preference-based HRQoL instruments have been criticised for being too simplistic and lacking rigorous psychometric properties for their descriptive system. (143) It is expected that any good instrument should have robust psychometric properties. (83, 144) Additionally, psychometrics provides standards for assessing validity of HRQoL instruments.

Although there are several separate sets of psychometric guidelines, there is a general consensus on psychometric properties and criteria that should be evaluated. In this thesis four guidelines for patient reported outcome measures often used when looking into instruments' psychometric properties were reviewed. (145-148):

i- *The United States Food and Drug Administration (FDA) provides guidance for Industry Patient-Reported Outcome Measures for use in Medical Product Development to Support Labelling Claims*.

ii- *The Scientific Advisory Committee of Medical Outcomes Trust (MOT) - established in 1994 with the aim of 'reviewing instruments an assessing their suitability for broad distribution by the trust'.(146) In order to execute their function, it was reasoned that health status and QoL instruments should establish some principles and review criteria for such.*

iii- *COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). The COSMIN guidelines were developed from a consensus of methodological qualities applied by studies aimed at evaluating measurement properties. (148) A review of those studies by the COSMIN committee led to a consensus of definitions, measurement properties as well as measurement criteria.*

iv- *International Society for Quality of Life (ISOQOL). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centred outcomes and comparative effectiveness research. The ISOQOL guidelines are a product of psychometric properties consolidated through literature review which included the other three guidelines (98, 147).*

For the purposes of this doctorate research, the following list of psychometric properties in Table 2.2 were consolidated from the above four guidelines:

Table 2.2 Psychometric properties consolidated from the four psychometric guidelines

Main Psychometric property	Psychometric property Sub-category	Further sub-category of psychometric property
Conceptual framework		
Reliability	Internal consistency	
	Test re-test	
	Inter-rater	
Validity	Content	
	Construct	Convergent
		Discriminant
		Known-group
	Criterion	Concurrent
		Predictive
	Responsiveness	
Interpretability of scores		
Practicality/burden		
Cross-cultural and language adaptation		

Source: FDA, MOT, ISOQOL, COSMIN (145-148)

These psychometric properties (Table 2.2) and their review criteria are now outlined in detail below.

2.4.1 Conceptual framework

This is the first step in the development of an instrument. The framework provides detailed description of the conceptual basis for development of the construct to be measured. It is assumed that any instrument will be robust and have an existing good conceptual framework. It is for this reason that this is less detailed and evaluated when adapting an existing instrument as is the case when developing one. (149) Evaluation of the conceptual framework is however important especially when adapting the instrument in a setting other than the one it was developed. The conceptual framework is usually assessed by ensuring that there should be conceptual basis for item content, dimension and distinctiveness of scales, intended level of measurement and evidence of defining and describing concepts. It is therefore important that

there should be evidence of involvement of the target population, patient input and/or the general population in the development of an instrument's dimensions. (146, 147)

2.4.2 Reliability

Reliability is the degree to which an instrument is free from random error to give consistent or dependable results each and over time. Reliability takes three forms: internal consistency, test-retest and inter-rater.

Internal consistency is the extent to which all the items in a scale are homogenous and reflect the same concept. For instance, a scale measuring social functioning might ask how one relates with friends and others in school/work/home, playing games, chatting and other social aspects of social life. It is expected that these items should be correlated to each other or homogenous since they are measuring the same social domain. This correlation or homogeneity in a scale is usually assessed using Cronbach's alpha, α, or KR-20 for dichotomous scores. (146, 148)

Cronbach's α is calculated using the following formula:

$$\alpha = \frac{k}{k-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right)$$

where k= number of items, σ_X^2 = total score variance, and σ_i^2 = individual score variance. Cronbach's α is usually measured on a scale of 0 to +1. Higher values of α indicate greater homogeneity of an instrument. A reliable instrument will have α ≥0.70 for group level comparisons and ≥0.90 for individual comparisons. (146, 147)

Test-retest reliability is the extent to which a scale is stable over time, assuming no expected change in the interim. It is expected that assuming nothing else has changed, the score from a

participant would remain the same over a short period of time. A clear statement of time interval for test-retest is required- usually assessed within a period of two weeks. Test-retest reliability is assessed using intraclass correlation coefficient (ICC) and kappa coefficient, $\varkappa$. Both ICC and $\varkappa$ are measured on a scale of 0 to +1 with a higher value indicating stronger evidence of the instrument's reproducibility. The criteria for either ICC or $\varkappa$ of ≥ 0.70 may be used for nominal scales. (146, 148)

Inter-rater reliability describes the extent to which a scale produces similar results when the scale has been rated by different people. It is only applied to instruments that are rated by two or more raters, neither of whom are providing a self-report and the results compared. Inter-rater reliability is the degree to which the responses from the different raters agree. The criteria for inter-rater reliability is ICC ≥ 0.70, where a high value indicates high levels of agreement between two assessors. (146, 149)

2.4.3 Validity

Validity refers to an instrument's ability to measure what it claims to measure. As an example, a scale that measures weight might be said to be reliable if it gives consistent readings each time. If that same measuring scale is poorly calibrated, the readings although reliable would be inaccurate (not valid). Validity is thus a way of ensuring that the instrument is not only reliable but that it measures what it purports to measure. This shows the importance of the psychometric property of validity for any good instrument. Validity has various types, but usually include content validity, construct validity, criterion related validity and responsiveness.

Content validity has been said to be probably the most important of all different forms of validity. (125) It is the extent to which the items in the scale represent the conceptual breadth of the construct. All the relevant concepts that make up the construct should be included in a

scale. There is no statistical criterion for assessing content validity of an instrument. Generally, an instrument is expected to be comprehensive and that it should cover all relevant domains. This can be achieved by ensuring that at development stage, there is target population involvement in identifying and selecting items (and domains) of a given construct. (147, 149)

Construct validity refers to the degree to which an instrument confirms an priori hypothesis i.e., whether truthful to measuring what it claims to measure. It is usually assessed by measuring the instrument's convergent validity, discriminant validity and known-group validity.

Convergent validity is the extent to which similar dimensions of two or more instruments are related. Discriminant validity on the other hand measures the extent to which unrelated dimensions between scales should not be similar. Convergent and discriminant validity are both assessed by an a priori hypothesis usually using correlation coefficients. For convergent validity, a coefficient of ≥ 0.50 would indicate moderate to strong correlation. (148) It is expected that similar dimensions on different instruments would have moderate to strong correlation as evidence of convergent validity, the higher the better. Discriminant validity is assessed for lack of correlation between dimensions measuring different concepts which is usually <0.2. The smaller the coefficient the better the discriminant validity. Known groups validity is the capacity for the scale to detect differences between groups known to be clinically different (for example cases and controls). It is usually assessed by an a priori hypothesis that scores for clinically different groups would be different. The statistical significance of the difference between clinically different groups is usually assessed using p-value of <0.5 although this does not indicate the magnitude of the difference. The magnitude is better assessed using Cohen's effect size statistic calculated using the mean difference divided by

standard deviation of the reference (healthy) group as follows:

$$\frac{mean\ of\ healthy\ children - mean\ of\ sick\ children}{standard\ deviation\ of\ healthy\ children}$$

where <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, >0.8 large. (150) The higher the effect size, the better the scale is at distinguishing differences between two or more clinically different groups.

Criterion-related validity is the extent to which scores on the new instrument are correlated with scores of the 'gold standard' that measures the same construct. This can be administered either concurrently (concurrent validity) or by an a priori model of prediction (predictive validity). Criterion-related validity may be evaluated by comparing a 'new' scale and an existing validated gold standard tool measuring the same construct. A correlation ≥ 0.70 between the 'new' scale and an existing validated tool is used as a criterion. (151) Due to lack of 'gold standard' instruments, assessing criterion related validity for most scales is not realistically possible. Also, were there a 'gold standard' there would be no need for developing new measures.

Responsiveness is another part of validity which has been referred to as part of longitudinal construct validity in some quarters. (105) Responsiveness is usually defined as the ability of the scale to show change in relation to a treatment or intervention of known efficacy. For example, a patient receiving treatment is expected to experience improvement of his or her health condition because of the treatment with a known efficacy. A good scale should have the ability to detect these health improvements when the person is administered a health

instrument. Responsiveness is usually assessed by standardised response means (SRM) or effect size. There is no standardized criterion to measure responsiveness since this is very much dependent on the instrument, construct and the intervention/treatment that is causing the change. Generally, responsiveness statistics are evaluated against an a priori hypothesis, indicating the size of change that is expected. (147)

2.4.4 Interpretability of scores (used for minimal important difference)

This is the degree to which one can easily assign meanings to an instrument's quantitative scores. The interpretability of scores therefore refers to a set of instructions of how to use a scale and assign meaning to the values, and as such there is no single assessment for it. A scale is expected to at least have some, if not all, of the following: responder definition, information on how data from the instrument is reported and displayed, representative mean(s) and standard deviation(s) in the reference population, and guidance on minimally important differences in scores. (146, 147, 149, 151)

2.4.5 Practicalities versus work burden to administering a scale

Practicalities and work burden refer to the amount of work involved on those completing and/or how to administer the scale. There are several ways for assessing practicality of a scale including questionnaire length, formatting, size of font (as it affects readability) and completion rate. Completion time, response rate, any indication of coercing for a particular response and resources used to administer the questionnaire are also used to assess the practicality aspect.

2.4.6 Conceptual and linguistic adaptation

This is the extent to which a scale shows linguistic and conceptual equivalence, and adequate measurement properties of the adapted version have been demonstrated. It is usually assessed by all of the following processes: forward translation, backward translation and cognitive

interviews. Forward translation is from the scale source language to the target one and it is recommended that this be done by two independent translators. The forward translators are usually native speakers of the target language but are also conversant with the original source language. The backward translation process involves translating back what was translated in the target language into the original source language. Backward translators usually are native speakers of the source language but are also conversant with the target one. This process ensures that the concepts and their meaning are the same in the target as in the source language. There should be details of translators, how inconsistencies were reconciled and if translators worked independently. (146, 147, 151) A final step involves establishing conceptual equivalence through cognitive interviews. This involves administering the translated version on a sample of the target population and getting feedback about their understanding of the concepts and ideas.

Table 2.3 below presents a summary of all the consolidated psychometric properties and criteria from the four guidelines (FDA, MOT, ISOQOL and COSMIN) as discussed above.

Up to this point, this thesis has only examined properties of a robust HRQoL instrument: descriptive system, utility values, and psychometric properties and evaluation criteria. The next phase is to review available generic childhood preference-based HRQoL instruments and evaluate how health has been captured by these instruments. The available instruments will be assessed in light of the psychometric characteristics and criteria as outlined here and summarized in Table 2.3. In order to achieve this objective, it was necessary to do a systematic review of the published literature. This was to inform what has been published around generic preference-based HRQoL measures in children and adolescent populations. Specifically, the review aimed at identifying and psychometrically appraising the available generic childhood

preference-based HRQoL instruments i) in general ii) their psychometric validation in sSA; and iii) use among children and adolescents in sSA.

Table 2.3 Summary of psychometric properties and criteria consolidated from four guidelines

Psychometric property	Conceptual framework[1-4]	Reliability (dependability/ consistency of an instrument) [1-4]			Validity (accuracy/ represent exactly what is on the ground) [1-4]						Interpretation of scores[1-4]	Practicality/ Burden [1-4]	Cross-cultural adaptation: Conceptual & linguistic equivalence[1-4]
		Internal consistency [1-4]	test re-test (reproducibility)[1-4]	inter-rater [1, 2, 4]	Content [1-4]	Criterion[1-4]		Construct [1-4]			Responsiveness[1-4]			
						concurrent	predictive	Convergent	Discriminant	Known group				
Evaluation criteria from psychometric guidelines	• conceptual basis for item content • target population involved • info on dimension & distinctiveness of scales • intended level of measurement • documentation defining & describing concepts included & intended population • documentation of how concept(s) are organized	**Cronbach's alpha, α** - expected to be α:≥0.70 for group level comparisons ≥0.9 for individual comparisons	**usually assessed using Inter Class Correlation (ICC). At times assessed using kappa** • expected ICC, kappa, κ≥0.70 • kappa may be used for nominal scales (κ≥0.70) • *clear statement of the time interval is required*	**usually assessed using ICC (≥0.70)**: • Inter class correlation coefficient	no single criteria for this but generally that the questionnaire is comprehensive: • Items represent all concepts in the conceptual framework • covers all health-related domains • include qualitative interview to evaluate target population understanding of all included items	It generally relies on the existence of a 'gold' standard instrument. Were the 'gold' standard available, there would be no need for new measures. Where this has been applied: • provide rationale and support for using this. • correlation between the two instruments should be ≥0.70 • for screening instruments, this can be compared between screening results against diagnosis		there should be a priori hypotheses among similar constructs (to merge) and that unrelated constructs should diverge. Assessed by: • **convergent**: correlation with tools measuring same construct ≥0.50 • **discriminant**: no correlation between tools measuring different constructs		**compare scores** for groups known to clinically differ on relevant variables. Assessed by Cohen's effect size: • <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, >0.8 large	usually assessed by: • standardized response means (SRM) = mean diff/ SD of score change • SRM>0	• logic and methods used • responder definition • Info on how data from measure be reported and displayed • representative mean(s) and std dev(s) in the reference population • guidance on minimally important diff in scores	Several ways to assess this and could be a mixture of: • instructions for each item • completion rate • completion time • literacy level • response rate	• process used to translate & culturally adapt instrument • linguistic-≥2 forward translation; ≥1 backward translation; expert and lay review; field test • details of translators • how inconsistencies (if any) were reconciled • translators worked independently

Colours represent the main psychometric properties and sub-categories; 1 United States Food and Drug Administration (145); 2 Scientific Advisory Committee of Medical Outcomes Trust; 3 International Society for Quality of Life; 4 COnsensus-based Standards for the selection of health Measurement INstruments

2.5 Systematic review of childhood preference-based health-related quality of life instruments applied and validated for use in sub-Saharan Africa

2.5.1 Aims and objectives

The broad objective was to investigate the methods and application of generic childhood preference-based HRQoL measures in sSA, thereby informing their future use, especially in economic evaluations of childhood interventions in the sSA region. The objectives of the review are:

Objective 1: to identify existing generic childhood preference-based HRQoL measures and summarise their psychometric quality

Objective 2: to explore which of these generic childhood preference-based HRQoL measures have been cross-culturally validated in sSA

Objective 3: to assess the extent to which the cross-culturally validated generic childhood preference-based HRQoL measures have been used in sSA.

2.5.2 Methods

The literature searches were conducted using a two-stage process compatible with the preferred reporting items for literature reviews and meta-analyses (PRISMA) guidelines. (152) For the first stage, the search aimed to identify existing childhood preference-based HRQoL measures (Objective 1). In the second stage, to address Objectives 2 and 3, the search strategy was refined to include names of instruments identified under Objective 1 and adding the sSA terms. This aimed to identify all published studies that applied or cross-culturally validated any one of these childhood preference-based HRQoL measures in the sSA region. Cross-cultural validation in this literature review and as used throughout this thesis was defined as i) linguistic equivalence (achieved through translation), ii) conceptual equivalence (achieved through focus group discussion or cognitive interviews) and iii) psychometric validation. The peer reviewed literature published between 1st January 1990 and 2nd August 2017 was examined for both

stages of the review. The year 1990 was chosen since most childhood health measures have had their origin around this time.

2.5.2.1 Eligibility criteria

To address Objective 1, studies reporting development of measures were included if: (1) they reported the development of at least one childhood preference-based HRQoL measures (this was limited to the original work by the developers only); (2) the target age(s) for completion of the measure included children between 6-17 years; and (3) were published in the English language. Studies were excluded if they reported measures designed for adults (≥18 years) or young children (< 6 years) as the review focussed on measures with the potential for self-completion by children. (153) Additionally, included studies were restricted to the psychometric development work by the primary developers of each identified instrument.

To address Objectives 2 and 3, studies were included if they had applied and/or cross-culturally validated any measure identified at stage one of the review in a sSA setting. Studies using generic childhood preference-based HRQoL measures that were not carried out either in part or whole in a sSA setting were excluded.

Only peer-reviewed articles published in English were included as there were no resources for translating articles from other languages.

2.5.2.2 Information sources

The initial search was conducted on Ovid platform MEDLINE databases to determine if the review question has been previously addressed. After establishing the facts, the next step was a search for title, abstract and keywords in the following electronic databases:

- MEDLINE (through OVID Sp)

- EMBASE (OVID Sp)
- EconLit (EBSCOhost)
- PsycINFO
- Web of Science
- PubMed

2.5.2.3 Search strategy

In addition to psychometric properties, Brazier et al (154) recommends that preference-based instruments should also be judged for their underlying utility theory: valuation techniques, assumptions, quality of data, whose preference as well as empirical preference evidence. A similar approach was applied to examine each childhood preference-based HRQoL instruments for: i) general description of the instrument, ii) underpinning utility theory, and iii) psychometric validation.

In stage one, the literature searches were done using an intersection of four sub-groups of search terms: 1) HRQoL terms; 2) childhood terms; 3) instrument terms; and 4) psychometric terms. The psychometric terms were included in order to identify validated measures only. A full list of search terms (see Table 2.4) was carried out using the following key words:

[“quality of life” OR “quality adjusted life years” OR “QALY” OR “health state utilities” OR “functional status” OR “health status”] AND [“child [MeSH Terms]” OR ”adolescent” OR “child* OR adolesc*” OR “paediatric*” OR “pediatri*”] AND [“measure” OR “measurement” OR “scale” OR “index” OR “measures”] OR [“valid*” OR “reliab*” OR “assess*” OR “psychometrics [MeSH Terms]” OR “psychometrics”].

Table 2.4 Search terms for identifying available generic childhood preference-based HRQoL measures (objective 1)

Search categories		Search terms
Question 1 apply date filter	#30	29 AND (01/01/1990-02/08/2016)
Question 1	#29	8 AND 16 AND 22 AND 28
Psychometric terms	#28	OR (23 to 27)
	#27	psychometrics
	#26	psychometrics [MeSH Terms]
	#25	assess*
	#24	reliab*
	#23	valid*
instrument terms	#22	OR (17 to 21)
	#21	measures
	#20	index
	#19	scale
	#18	measurement
	#17	measure
child terms	#16	OR (9 to 15)
	#15	pediatri*
	#14	paediatri*
	#13	child* OR adolesc*
	#12	adolescent
	#11	adolescent [MeSH Terms]
	#10	child
	#9	child [MeSH Terms]
QoL terms	#8	OR (1 to 7)
	#7	health status
	#6	functional status
	#5	Health state utilities
	#4	QALY
	#3	quality adjusted life years
	#2	Quality of life
	#1	Quality of life [MeSH Terms]

In stage two, the same databases were searched using terms relating to the identified generic childhood preference-based HRQoL instruments and combined them with sSA terms. This was to identify the application of the identified instruments and/or cross-cultural validation in sSA settings. The full list of identified instruments and sSA search terms used to address the second and third study objectives is outlined below as per Table 2.5:

[(((((EQ-5D or “EQ 5D” or EQ5D or Euroqol or “Euro qol” or EQ-5D-Y or "EQ 5D Y")) OR ((16D Health-Related Quality of Life or 16D HRQoL or 17D Health-Related Quality of Life or 17D HRQoL))) OR ((AQoL-6D or Assessment of Quality of Life-6D))) OR ((“Child Health Utility 9 Dimension” or CHU9D or CHU-9D or “CHU 9D”))) OR ((“quality of well-being” or “quality of well-being” or QWB))) OR “health utility index”) AND sub-Saharan Africa]

Table 2.5 Search terms for identifying applications and cross-cultural validations of generic childhood preference-based HRQoL instruments (objectives 2&3)

Tool/ Measure		Search terms
	#10	8 AND 9
	#9	sub-Saharan Africa
	#8	((((((EQ-5D or "EQ 5D" or EQ5D or Euroqol or "Euro qol" or EQ-5D-Y or "EQ 5D Y")) OR ((16D Health-Related Quality of Life or 16D HRQoL or 17D Health-Related Quality of Life or 17D HRQoL))) OR ((AQoL-6D or Assessment of Quality of Life-6D))) OR (("Child Health Utility 9 Dimension" or CHU9D or CHU-9D or "CHU 9D"))) OR (("quality of well-being" or "quality of well-being" or QWB))) OR "health utility index"
16D & 17D	#7	16D Health-Related Quality of Life or 16D HRQoL or 17D Health-Related Quality of Life or 17D HRQoL
AHUM	#6	Adolescent Health Utility Measure or AHUM
AQoL-6D	#5	AQoL-6D or Assessment of Quality of Life-6D
CHU-9D	#4	"Child Health Utility 9 Dimension" or CHU9D or CHU-9D or "CHU 9D"
EQ 5D-Y	#3	EQ-5D or "EQ 5D" or EQ5D or Euroqol or "Euro qol" or EQ-5D-Y or "EQ 5D Y"
HUI2 & HUI3	#2	"health utilities index"
QWB	#1	"quality of well-being" or "quality of well-being" or QWB

2.5.2.4 Study selection

All the identified abstracts were screened with assistance of a second and third reviewer for Objective 1. The full texts of relevant studies were obtained and independently assessed by the three reviewers against the inclusion and exclusion criteria. To achieve Objectives 2 and 3, all identified abstracts were reviewed with assistance of the second reviewer only. For quality control purposes, a senior academician reviewed disagreements. A final list of studies included in analysis was arrived as consensus of the reviewers.

2.5.2.5 Procedures

Along with other reviewers, the titles and abstracts were independently assessed, to select potential full-text articles for screening. Study selection was agreed upon at each stage.

2.5.2.6 Data extraction and analysis

To meet Objective 1, for each identified generic childhood preference-based HRQoL instrument the following was extracted:

a) General description of the measure

The descriptive system consisting of the classification system; where, how and who developed the classification system; target population; number of dimensions; response levels; and time taken to respond.

b) Underpinning utility theory

This included the number of health of states generated by the classification system, the valuation technique used, statistical method used to extrapolate the utility values to all possible health states, range of utility values for each measure and when and how these were generated.

c) Psychometric validation

All the instruments identified in the literature review were evaluated for psychometric properties outlined in section 2.4 above. The following criteria was applied to all the identified instruments.

i) Conceptual framework. This was classified as: (+) if this met the criteria; (-) if not; and (NR) if not reported. The criterion was said to have been met if there was a clear framework of how dimensions were developed but also if the target population were involved in determining items for an instrument.

ii) Cross-cultural language and conceptual equivalence/adaptation. This was classified as: (+) if met the criteria; (-) if not; and (NR) if not reported. The criteria were met if any of the forward translation, backward translation or cognitive interviews were done.

iii) Reliability. Internal consistency was classified as: (+) if either group Cronbach's α, ≥0.90 or individual comparisons Cronbach's α, ≥0.70 was met (-) if not; and (NR) if not reported. Test retest reliability was classified as: (+) for ICC or k≥ 0.70; (-) if not; and (NR) if not reported. Interrater reliability was classified as: (+) for ICC ≥ 0.70; (-) if not; and (NR) if not reported.

iv) Validity. Content validity was classified as: (+) if met the criteria; (-) if not; and (NR) if not reported. Convergent validity was classified as: (+) if correlation ≥ 0.50; (-) if not; and (NR) if not reported. Discriminant validity was classified as: (+) if no correlation between tools measuring same constructs; (-) if not; and (NR) if not reported. Known-group validity was classified as: (+) if scores for clinically known different groups were different; (-) if not; and (NR) if not reported. Criterion validity was classified as: (+) if correlation ≥ 0.70;

(-) if not; and (NR) if not reported.

v) Responsiveness. Responsiveness was classified as: (+) if there was observed change in scores; (-) if not; and (NR) if not reported.

vi) Interpretability of scores. It was classified as: (+) if criteria for interpreting scores was available; (-) if not; and (NR) if not reported.

vii) Practicality/ burden. Practicality of a scale was classified as: (✓) if there was at least one criterion used; (-) if not; and (NR) if not reported.

The evidence for psychometric properties was assessed as a whole rather than relying on a single statistic to indicate psychometric quality.

To achieve Objective 2, each identified generic childhood preference-based HRQoL instrument was assessed according to whether there had been: i) cross-cultural translation into any native language spoken in sSA; ii) any conceptual equivalence through focus group discussion or cognitive interviews on word meanings and changes to concepts to fit what would be locally equivalent into a sSA setting and iii) psychometric validation. Cross-cultural validation was classified as (+) if it was appraised and performed; (-) if appraised but not performed; and (NR) if not reported.

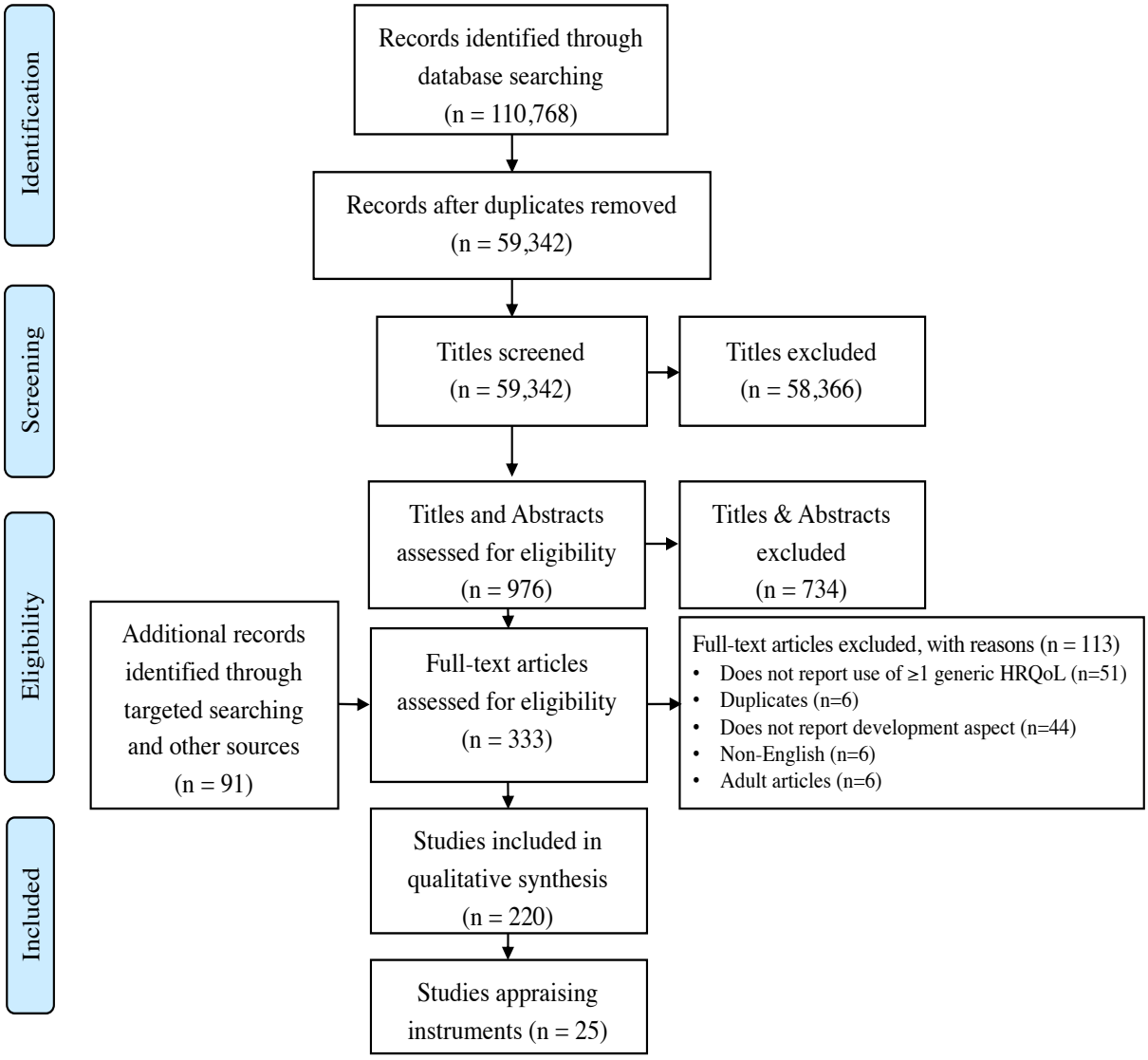
In order to achieve Objective 3, each instrument was assessed to the extent it was used in a sSA setting. Additionally, all reported studies using identified instruments were assessed for their quality using the Critical Appraisal Skills Programme (CASP) toolkit for case-control studies. (155) The CASP tool consists of ten questions that address issues such as study aim and design, data collection and analysis, and ethical issues raised by the published study.

2.6 Results

2.6.1 Identifying childhood preference-based HRQoL measures (Objective 1)

The first stage of the search identified 59,342 unique articles. 58,366 articles were excluded after screening the title and a further 734 articles excluded after screening the abstract. In addition to the remaining 242 articles, an additional 91 articles were identified through snowball searching. The review therefore focussed on the remaining 333 full texts articles for eligibility. Of these, 113 articles were excluded as they did not meet eligibility criteria; a large number (n=51) of these articles were excluded as they did not report use of a generic childhood preference-based HRQoL measures. In total, 220 were included in qualitative synthesis from which 25 articles reported development of at least one generic childhood preference-based HRQoL measures. Ten unique childhood preference-based HRQoL measures were identified from the 25 reviewed articles as per Figure 2.2 below.

Figure 2.2 PRISMA Flow-diagram for children and adolescent preference-based HRQoL instruments to address Objective 1: generic preference-based HRQoL that have been developed for use in children and adolescents



2.6.2 Identified generic childhood preference-based health-related quality of life measures

In summary there was a variety of identified measures in terms of general description, utility measurement and psychometric properties. Below is a summary description and a detailed description for each instrument follows.

i) General description of the measure

The review identified ten existing childhood preference-based HRQoL measures outlined in Table 2.6: 16-Dimensional Health-Related Measure (16D), 17-Dimensional Health-Related Measure (17D), Adolescent Health Utility Measurement (AHUM), Assessment of Quality of Life-6Dimension (AQoL-6D), Child Health Utility 9Dimension (CHU-9D), EQ-5D-Youth (EQ-5D-Y), Health Utilities Index Mark 2 (HUI2), Health Utilities Index Mark 3 (HUI3), Quality of Well Being-Self Administered (QWB-SA) and, Child Health-6 Dimension (CH-6D). The CH-6D is a childhood preference-based HRQoL measure recently developed in South Korea for children aged 7-12 years. (156) All ten measures were developed in HICs, adapted predominantly from existing adult versions with little or no involvement of children at the development stage (107) with the exception of the CHU-9D. (157)

ii) Underpinning utility theory of identified measures (Table 2.7):

Utility values had been generated using TTO for the AQoL-6D, AHUM and CH-6D; combining TTO with a VAS for the EQ-5D-Y; a RS for 16D, 17D, and QWB-SA; and combining a RS and the SG approach for the HUI2 and HUI3. The utilities were elicited using adult preferences - whether own or proxy – for the HUI2, HUI3, QWB-SA, 17D, CHU-9D, EQ-D-Y, AHUM and CH-6D. The AQoL-6D and 16D were the only two measures that elicited the preferences of adolescents. Through either additive or multiplicative statistical methods,

the utility values generated for a sample health states were then extrapolated for all health states of the measures. For the 16D, 17D, AHUM, QWB-SA, EQ-5D-Y, CHU-9D this was through an additive statistical method and through a multiplicative statistical method for the HUI2, HUI3 and AQoL-6D.

iii) Psychometric validation:

Table 2.8 presents the psychometric appraisal of the 10 identified instruments. The psychometric evidence was mixed. Development of a conceptual framework, content validity, and convergent validity were most frequently evaluated. Only six measures (HUI2, HUI3, 16D, 17D, CHU-9D and EQ-5D-Y) had at least one piece of adequate evidence for both reliability and validity. This is summarised below for each instrument and a detailed description of the psychometric evidence can be found in Appendices (Appendix 1 and 2).

These ten childhood preference-based HRQoL instruments are now described in detail below- arranged in order of most recently developed first.

2.6.2.1 Child Health-6 Dimension (CH-6D)

This instrument was developed in 2016 by Kang (156) in South Korea for children aged 7-12 years. No information is available on how the descriptive system was developed. It consists of six self-completed items each with 3 or 4 levels to measure HRQoL in Korean children. The CH-6D describes 2,304 health states and its utility values were derived from the general population using TTO. It is said to be among the first if not the first generic preference-based HRQoL measure developed in the Asian region for children. (156) Although, originally developed for use in children the developer notes that it can also be used in adolescents.

Despite considerable effort to contact the developer, this was unable to get more information about items covered, underlying utility theory and psychometric performance. The only available data was from an abstract the author presented at a conference in 2016. The aim of that paper was to validate CH-6D among adolescents and it showed evidence of convergent validity with HUI2 (correlation=0.658). Evidence of known-group validity is in its ability to distinguish health status of two clinically different groups (score range: 0.595-0.947).

2.6.2.2 Assessment of Health Utility Measurement (AHUM)

AHUM descriptive system was developed using a decomposed top-down approach in 2012 (UK) by a team of experts. The team selected both health attributes and response options from EQ-5D, Short Form-36 (SF-36) and literature review focussing on impacts of childhood conditions. (158) It is self-completed questionnaire for older children, adolescents (11-18years) and adults. The authors do not specifically indicate what the AHUM measures. The targeted population were not involved in the initial and final selection of attributes. However, 11 subjects (12-18 years) with hunter syndrome and 27 caregivers (parents) were involved in domain concept refinements. The instrument has six dimensions: self-care, pain, mobility, perceptions of strenuous activities, self-image and health perceptions, with 4-7 response levels.

Utility values for a sample of health states were generated using choice-based TTO from 312 adults (≥ 18 years) of the UK general population. A statistical additive method was then applied to estimate scores for each of the instruments' 16,800 unique health states.

Evidence of content validity is from the fact that AHUM was derived from EQ-5D and SF-6D. (158) Other than this, there is no evidence of any psychometric performance for this measure

and this could partly be due to the instrument being relatively new.

2.6.2.3 EQ-5D-Y

This is the youth version of the generic preference-based instruments developed by the EuroQol Group. The EQ-5D-3L adult version is the most widely used adult generic preference-based HRQoL measure. (122) The descriptive system of the adult version was developed using a decomposed top-down approach around 1990s by a group of researchers. The group from across five countries reviewed generic health status instruments at the time and made a final selection of dimensions. HRQoL in children was estimated using the adult EQ-5D-3L version until development of the child-friendly version in 2010. (159) The childhood version is distinguishable from the adult one with a Y abbreviation for youth at the end (EQ-5D-Y) and was developed for those aged 7-11 years. The EQ-5D-Y was developed by modifying the contents of EQ-5D adult version, again, using a decomposed top-down approach. A task force from seven countries (on behalf of EuroQol Group) with expertise in child psychology, paediatrics, health economics, statistics, sports sciences and rehabilitation sciences modified the adult version into a child-friendly one. The modification process included revision of dimension definitions, revision of wording and layout, translation, cognitive interviews, integration and decision-making on harmonization, and comparing the two versions. Although, target subjects were not involved in selection of dimensions (as was a carry-over from the adult version), they were involved in dimension refinement and pilot stages. The descriptive system five dimensions was modified into child-friendly language: mobility, looking after myself, usual activities, pain or discomfort and worried, sad or unhappy, so that children and adolescents can easily relate to.

The EQ-5D-Y does not have its own utility values but use those of the adult version. Challenges for obtaining utility values from children are discussed in section 2.3.2.2 of this chapter and elsewhere. (126, 159) The adult utility values were generated from a sample of 42 health states from 3,395 (2,997 respondents) UK adult general population. Using a statistical additive method, scores for each of the 243 unique health states were generated from the 42 valued health states.

There is evidence of content validity, though mixed as developers admit the need to add dimensions like social functioning and school. (159) The instrument has been found to be feasible in several countries with either nothing missing or very few inappropriate responses. Both the descriptive system and EQ VAS system missing data ranged from 0-2%, and time taken to complete <5 minutes. (160) There is also evidence of a range of psychometric properties in cultural contexts from multiple studies globally. (161) The findings show that the EQ-5D-Y is well understood and practical to use by respondents across most countries. (160) Evidence across all contexts show that test-retest reliability ranged from 69.8%-99.7% (160), with high inter-rater agreement (87.3%-100%) between children and proxies. (160) As evidence of convergent validity, kappa coefficient and correlation coefficient were up to 67%, and 0.56 respectively. The known group validity was evident in the EQ-5D-Y ability to produce different HRQoL scores between healthy and ill groups.

2.6.2.4 CHU-9D

Child Health Utilities 9 Dimensional (CHU-9D) was developed in 2008 making it is one of the newer generic childhood preference-based HRQoL measures. (157) The CHU-9D descriptive system is currently the only childhood preference-based HRQoL measure that was developed

using the decomposed bottom-up approach. (157) The developers undertook qualitative work with children during all developmental stages: domain formulation, refinements and piloting. In the final analysis, the instrument was refined to contain nine dimensions. The CHU-9D is a self-completed questionnaire, designed for children aged 7-11years, and has been validated in those aged 6-7yrs and 11-17yrs. (162) The CHU-9D instruments' nine dimensions: worried, sad, pain, tired, annoyed, schoolwork/homework, sleep, daily routine and activities, have five response levels each.

Utility values for the CHU-9D health state classification system have been derived using an indirect valuation approach. Firstly, utility values were obtained through choice-based SG from a sample of 300 UK adult (mean age, 49 years) general population. An additive statistical method was then applied to generate scores for each of the instruments 1.9×10^6 unique health states. (163)

There is evidence of content validity as the domains were elicited directly from target population using qualitative research. (157, 164) Even though about 11% felt that something was missing from the instruments descriptive system (165), 'within the skin' and 'handicap' dimensions are all included in CHU-9D. (157) Response rates ranged between 64% - 98% in the psychometric evaluation study showing evidence of the instruments practicality. (165) Completion rate for each question and mean completion time ranged from 98.7% - 100% and 3.8 - 5.3 minutes respectively. Additionally, the CHU-9D was found to be quick and easy to administer. Most children (95.5%) in general population were able to read and write without difficulty, and 88% of these were able to understand instructions. However, in clinic paediatric sample only 51% were able to read and write, and yet 95% were able to understand

instructions. (165) The internal consistency Cronbach's α from the original instrument development was 0.781. Child-rated level of health highly correlated ($p<0.05$) with descriptive system demonstrating construct validity. (165)

2.6.2.5 Assessment of Quality of Life 6 Dimension (AQoL-6D) adolescent

AQoL-6D has its origin from the AQoL-4D, the latter being developed in 1997. (105) The AQoL-4D descriptive system was developed by a team of Australian researchers using a decomposed top-down approach. Later, the AQoL-4D descriptive system was modified to give birth to AQoL-6D for use in adolescents 12-18yrs in 2004. (143) The four-step modification process included: 1). selection of a HRQoL concept; 2). item selection and describing dimensions using expert input, literature review and focus groups; 3). a 'construction survey' which involved comments from respondents who had experienced the relevant health states and 4). derivation of final items using statistical analysis. (166) The AQoL-6D has six dimensions with four to six response levels generating $7.8x10^{13}$ health states. The AQoL-6D six dimensions (independent living, pain, senses, mental health, relationships and coping) takes 5-10 minutes to complete and the recall period is over the past week. After developing the AQoL-6D in Australia, it was then adapted in four pacific countries (Australia, Fiji, New Zealand and Tonga). (140)

Utility values were generated from an adolescent population using choice-based TTO utility valuation technique. The utility values were econometrically modelled for the rest of 2,790 health states using a multiplicative statistical method.

The instruments' development study shows evidence of content and construct validity with exceptionally good exploratory factor analysis and structural equation modelling. Five of the six dimensions had a correlation coefficient of ≥0.73 (lambda loading gamma coefficients are equivalent of correlation coefficient) except for vision (=0.51). (166) The instrument also demonstrated reliable psychometric evidence when adapted for adolescents in four pacific countries. (140)

2.6.2.6 16-Dimensional Health-Related Measure and 17-Dimensional Health-Related Measure

The instrument name (16D) is an abbreviation of the 16 multiple choice questions, each question representing one health-related attribute (dimension). 16D was modified from 15D instrument -developed using a decomposed top-down approach from a review of Finnish health policy documents by experts in paediatrics, child psychiatry and neurology. (126, 167) 16D focuses on 12-15 age group but can be used up to 18yrs according to developers (www.15d-instrument.net/16d-and-17d/16d/). Even though neither the patient nor the public were involved in developing domains, the developers did involve the patient and public in the refinement stage and in testing the instrument. The 16D instrument has five response levels of its 16 dimensions: vitality, vision, breathing, distress, hearing, sleeping, eating, discomfort and symptoms, speech, appearance, school hobbies, mobility, friends, mental function, excretion and depression.

Utility values for the 16D instrument were elicited from a sample of 213 school going adolescents (12 - 15 years) in Finland using non-choice-based rating scale technique. The rest of 1.5×10^{11} health states had utility values generated using statistical additive method.

The measure is practically easy to complete taking between five to ten minutes. There is evidence of test re-test reliability with reliability coefficient r=0.91 (mean difference standard deviation =2) over and above the accepted standards. (167) Repeatability coefficient by individual level dimensions ranged from 86-100%, and majority (11 out of 16 dimensions) had repeatability coefficient ≥97% (only sleeping, and discomfort and symptoms <94%). (167) The measure has also demonstrated known group validity in its ability to distinguish between different HRQoL states of sick children reporting more problems on similar constructs compared to healthy children. (167) Further, content validity is based on the fact that 16D was developed by a team of children and adolescent experts and domains were mostly modified from an existing 15D instrument. Even though children and adolescents were not initially involved in the development phase, their input were included during piloting phase.

Like its 16D counterpart, the 17D instrument has 17 questions (dimensions) each representing one health-related domain. The instrument is targeted for a much younger age group (8-11years), although it can be used in children < 8 years but with assistance from a proxy. The 17 dimensions were developed using a decomposed top-down approach by a team of experts in paediatrics, child and adolescent psychiatry, child neurology and health economics mostly from literature review and by adapting the 16D. For example, 16D distress dimension was dropped whereas vision, vitality and depression were reformulated for comprehensibility for this young age group. (168) Three new domains were then added: ability to concentrate, anxiety, and learning ability and memory resulting in 17 dimensions. Children were involved in refining the instrument through a pilot study followed by an interview on relevance of domains. The final decision on domains to retain, however, was entirely that of the developers.

Each of the 17 dimensions has 5 levels of response generating 7.6 x 10^{11} health states. Utility values were elicited from 115 parents of children involved in the study using non-choice-based RS valuation technique. By employing additive statistical method, utility weights for the rest of the 7.6 x 10^{11} health states were generated.

There is evidence of practicality as children not only completed the form but also reported it to be enjoyable. (168) Test-retest reliability is evident with coefficient, r = 0.95 with cases lying within 2 SD of the mean difference and for dimensions it was between 0.91 - 1.0. 11 out of 17 domains had reliability coefficient r =0.97 with only two (sleeping and breathing) r<0.93. There is strong content validity, having been developed from validated 16D and a clear methodology of how the questionnaire was formulated is presented. The questionnaire was revised after input from children who were interviewed and the lower age (8years) was based on evidence of self-report. (168) Children with more health problems scored lower on particular dimensions reported more problems on particular related dimensions demonstrating convergent validity. However, there is no evidence of inter-rater correlation between children and proxies especially on speech being evaluated at the time.

2.6.2.7 Quality of Well-Being Self-Administered (QWB-SA)

The Quality of Well-Being self-administered version (QWB-SA) was modified from QWB, (169) using a decomposed top-down approach for adolescents (12-18years). The QWB is one of the earliest (1976) adult preference-based HRQoL instruments. (170) QWB is the first preference-based HRQoL measure to have been developed for use in economic evaluations.

(105) The QWB scale was not specifically developed for children and required a trained interviewer to complete the form. This led to the development of a much simpler self-administered QWB-SA scale in 1997. (105, 170) The difference between QWB-SA and most other preference-based HRQoL instruments is that it combines functioning, and symptoms and problems in the same scale. The QWB-SA has three dimensions (mobility, physical activity and social activity) and 58 symptoms/ problems that fall in four categories (chronic symptoms/problems, acute physical symptoms and mental health symptoms).

Utility values for the QWB-SA have been generated from a community sample of 435 (final sample N=430) English speaking adults in San Diego, California, USA. Non-choice-based RS valuation technique was used to generate utility values for a sample of 12 health states. The utility values for the rest of the 945 health states were extrapolated using additive statistical method.

The QWB-SA has been found to perform similar to the adult QWB in healthy population. (169) The QWB is a well validated instrument, it is practical to use and has content validity similar to the SF-36. (171) There is evidence for content validity since QWB-SA was developed from the validated QWB. The QWB-SA and QWB were found to be highly correlated indicating evidence of convergent validity. (169) Most of the other psychometric properties were not assessed at the development stage of the questionnaire. (169) This includes internal consistency reliability, test re-test reliability and known-group validity.

2.6.2.8 HUI Mark 2 and HUI Mark 3

The Health Utilities Index Mark 2 (HUI2) was developed in 1992 using a decomposed top-

down approach. HUI2 is a self-completed by respondents and is among the earliest child and adolescent measures. It has its origin from HUI1, developed in 1982, (105) to evaluate neonatal cancer interventions. (105) Subsequently, the HUI2 was modified from the HUI1 in 1992, for use in children and adolescent population (6-16 years). Initially, it was developed for use in cancer but has become one of the most widely used generic childhood preference-based HRQoL measure. (165) The HUI2 descriptive system was first developed from literature review and surveys by a team of experts. Thereafter, following cognitive interviews with 84 children and parent pairs (same gender and living in same household), the contents were refined. In the final stage, the team of experts selected seven domains: sensation, mobility, emotion, cognition, self-care, pain and fertility. Each of the domains has 3-5 response levels.

The HUI2 utility values were obtained from 194 parents using choice-based SG valuation technique. (86) Utility values for the rest of its 24,000 health states were derived using a multiplicative statistical method of its multi-attribute utility function.

The HUI3 was developed in 1998 using a decomposed top-down approach after concerns about conceptual overlap of HUI2 domains especially between self-care and other attributes. (84) To enhance independence, 'self-care' was thus replaced with dexterity. Further, questions regarding fertility perhaps due to age were eliminated. The sensation domain was divided up into three distinct domains: vision, hearing and speech. This resulted in the HUI3 being comprised of eight domains: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain and discomfort. Each of these domains has between 5 and 6 response levels. There was neither patient nor public engagement in generating descriptive system domains. The HUI3 has three modes of administration: self-report, face-to-face and telephone interview.

The HUI3 utility values were firstly elicited using non-choice-based VAS valuation technique from 250 adult (≥16years) general Canadian population. For the second step, the utility values were transformed using choice-based SG utility valuation technique. Finally, the utilities for the rest of 972,000 health states were then generated using multiplicative statistical method.

The reliability is mixed with high reliability (Kappa, K, = 0.61-0.80) for vision, ambulation and emotion; moderate (K=0.41-0.60) for hearing, cognition and pain; and low for speech and dexterity (K=0.137 and 0.347 respectively). (172) There is evidence of practicality, face validity, acceptability and reliability of both HUI2 and HUI3. (173) Evidence of clinically known-different groups has also been observed as well as construct validity and responsiveness.

Tables 2.6 and 2.7 shows respectively the general characteristics and utility value properties of the generic children and adolescent preference-based HRQoL measures. Table 2.8 shows the psychometric properties

Table 2.6 Generic child and adolescent preference-based HRQoL instruments arranged in chronological order of development

Measure (Acronym)	Age range (years)	Author, year (Ref)	Domain development	Domain refinement	Instrument Testing/ pilot	Completion	No. of Domains /Items	Respo-nse levels	Time to complete (min)	Recall period	Country developed (languages)
HUI3	6-16	Feeny et al 1995(84)	Experts	No	No	Self (proxy for 5-8yrs)	8	5-6	8-10	Past 4/2/ 1week	Canada, France (35)
HUI2	6-16	Feeny et al 1992(105)	Experts (surveys and literature review)	Yes, 84x Child & parent pairs rated items	No. Experts decided attributes	Self (proxy for 5-8yrs)	7	3-5	8-10	Past 4/2/ 1week	Canada (35)
QWB-SA	12-18	Kaplan et al 1996(170)	Experts (QWB)			Self	3	2-4	NA	3 days	USA (9)
17D [1]	8-11	Apajasalo et al 1996(168)	Experts (16D, literature review)	Pilot study with 79 healthy children	Final dimension decision by experts	Self	17	5	5-10	Today	Finland (4)
16D	12-15	Apajasalo et al 1996(167)	Experts (review of health policy)	Yes (Patient/ public involved)	Yes (Patient/ public involved)	Self	16	5	5-10	Today	Finland (5)
AQoL-6D	≥15	Richardson et al 2012(143)	Experts (AQoL)	No (as was developed from AQoL)	Yes, semantics and language were tested	Self	6	4-6	5-10	Past week	Australia (5)
CHU-9D	7-17	Stevens 2009(157)	Children (focus group discussion with children)	Yes,	Yes, patients and public	Self	9	5	4-5	Today/la st night	UK (7)
EQ-5D-Y	8-11	Wille et al 2010(159)	Experts (EQ-5D)	Yes (Patient/ public involved)	Yes (Patient public involved)	Self and/ or proxy	5	3	3	Today	Europe/ UK (26)
AHUM	12-18yrs	Beusterien et al 2012(158)	Experts (EQ-5D, SF-6D, and literature review)	Yes. 11x Hunter syndrome children and 27 parents	No	Self	6	4-7	N/A	N/A	UK (N/A)
CH-6D	7-12	Kang 2016(156)	N/A	N/A	N/A	Self	6	3-4	N/A	N/A	South Korea (N/A)

[1]Developed for 8-11yrs. Authors suggest it can be used for 6,7 via proxy and adolescents ≤16yrs; [2]SA self-administered; [3]IA interviewer administered
CH-6D Child Health 6D, AHUM Adolescent Health Utility Measure, AQoL-6D Assessment of Quality of Life 6D, EQ-5D EuroQol 5D, EQ-5D-Y EQ-5D Youth version, CHU-9D Child Health Utility 9D, 17D 17 Dimension, QWB-SA Quality of Well-Being self-administered, HUI2 Health Utility Index Mark 2, HUI3 Health Utility Index Mark 3

Table 2.7 Descriptive system development and utilities valuation technique of generic child and adolescent preference-based HRQoL instruments

Measure (Ref)	Health states	Valuation technique	Source of preference/ population (Age years)	HSUV statistical method	Instrument boundaries (min, max)	Year of preference made available
HUI 3 **(107, 165, 173, 174)**	972,000	SG/ RS	Adult (>=16 years)	Multiplicative	0.036, 1.00	2002
HUI2 **(107, 165, 173, 175)**	24,000	SG/ RS (VAS used then converted to a SG using power function)(105)	Adult proxy	Multiplicative	-0.03, 1.00	1996
QWB-SA (adults and children)**(107, 169-171)**	945	RS	Adults (>=18yrs)	Statistical additive except for symptom 1 problem	0.08, 1.00	1976
17D**(122, 168)**	7.6x 10^{11}	RS	Adult proxy	Additive	0, 1.00	1996
16D**(122, 167)**	1.5x 10^{11}	RS	Adolescents (12-15 years)	Additive	0, 1.00	1996
AQoL-6D**(107, 140, 166, 176)**	7.8 x10^{13}	TTO	Students (adolescents)	Multiplicative/ exponential	0.053, 1.00 min/country: Fiji 0.094; Australia 0.072; Tonga 0.068; NZ 0.053	2004
CHU-9D**(107, 142, 163, 165, 177)**	1,943,125	SG	Adults (mean age= 49yrs)	Additive	0.34, 1.00	2012
EQ-5D-Y**(107, 160, 178)**	243	Adult utilities TTO, VAS	Adults	Statistical additive	-0.59, 1.00	1993
AHUM **(158)**	16,800	TTO	Adults (>=18yrs)	Additive	0.42, 0.99	2012
CH-6D **(156)**	2,304	TTO	General population	N/A	N/A	N/A

CH-6D Child Health 6D, AHUM Adolescent Health Utility Measure, AQoL-6D Assessment of Quality of Life 6D, EQ-5D EuroQol 5D, EQ-5D-Y EQ-5D Youth version, CHU-9D Child Health Utility 9D, 17D 17 Dimension, QWB-SA Quality of Well-Being self-administered, HUI2 Health Utility Index Mark 2, HUI3 Health Utility Index Mark 3, NZ New Zealand

[1] EQ-5D-3L adult health states; AQoL-6D was valued concurrently but separately for each country; CHU-9D -UK children values was from adults with children in mind whereas in Australia it was from adolescents using BWS DCE technique

Table 2.8 Showing psychometric properties of generic children and adolescent preference-based HRQoL instruments

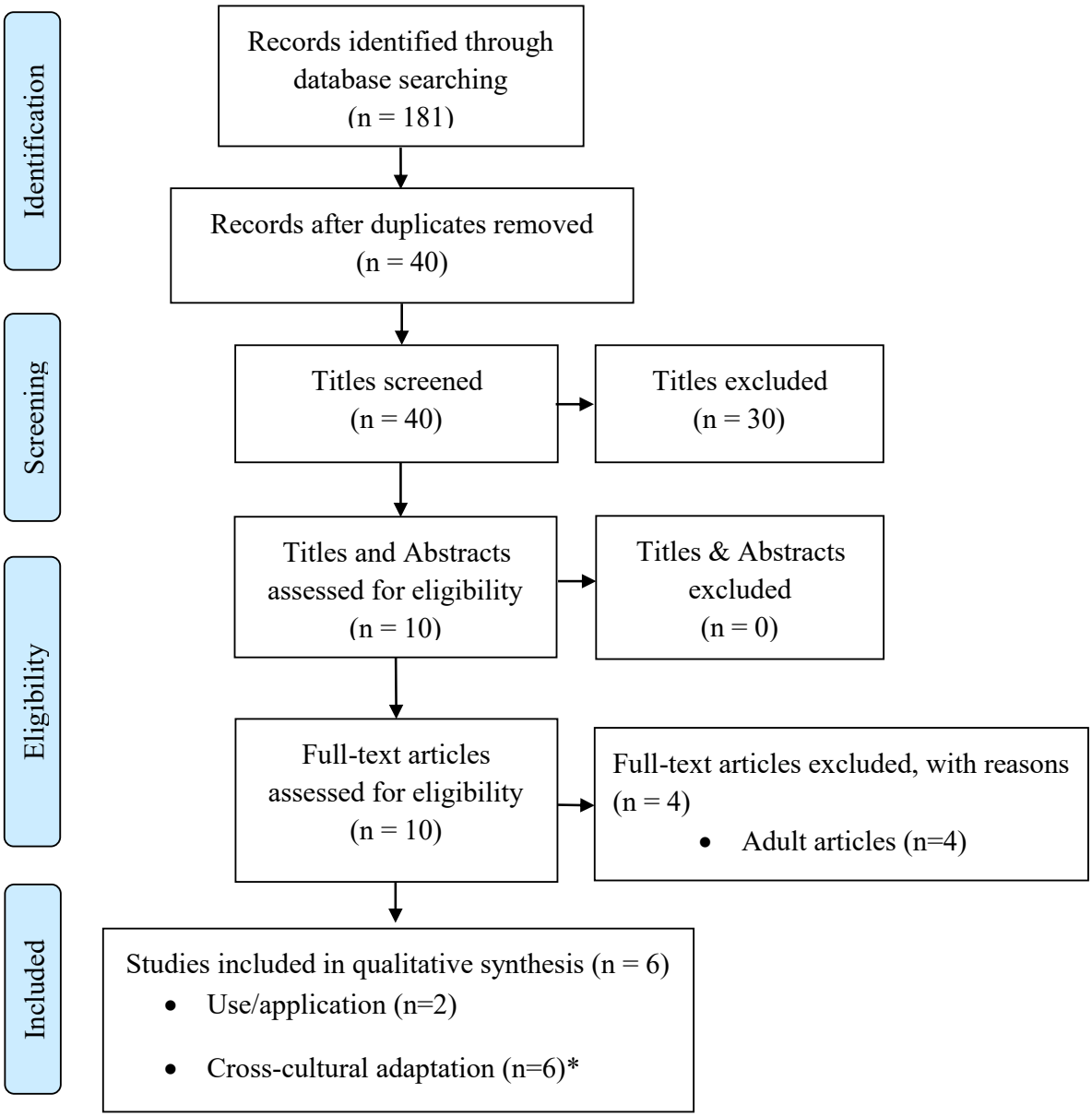
Psychometric property/ Measure	Conceptual framework[a, b,c,d]	Cross-cultural language adaptation[a,b,c,d]	Reliability (dependability/ consistency of an instrument)[a,b,c,d]			Validity (accuracy/ represent exactly what is on the ground) [a,b,c,d]							Interpret-ation of scores[a,b, c,d]	Practi-cality/ Burden [a,b,c,d]	Validated in sSA
			Internal consistency [a,b,c,d]	test re-test [a,b,c,d]	inter-rater [a,b,d]	Content [a,b,c,d]	Criterion-related[a,b,c,d]		Construct [a,b,c,d]			Responsi-veness[a,b,c d]			
							conc	pr	Conv[a]	Disc[a]	KG[a]				
HUI2 & HUI3* (165, 173)	NR	+	0	+	NR	NR	NR	NR	NR	NR	+	+	+	NR	Yes
QWB-SA(169, 171)	NR	NR	NR	NR	NR	+	NR	NR	+	NR	NR	NR	NR	+	No
17D (168)	+	NR	+	+	-	+	NR	NR	NR	NR	+	NR	NR	+	No
16D(167)	+	NR	++	+	NR	+	NR	NR	NR	NR	+	NR	NR	+	No
AQoL-6D(166, 176)	+	+	NR	NR	NR	+	NR	NR	+	NR	NR	NR	NR	NR	No
CHU-9D(157, 165)	+	NR	+	NR	NR	+	NR	NR	+	NR	+	NR	NR	+	No
EQ-5D-Y(159)	+	+	NR	+	+	-	NR	NR	-	NR	+	NR	+	+	Yes
AHUM(158, 165)	NR	NR	NR	NR	NR	+	NR	NR	NR	NR	NR	NR	NR	NR	No
CH-6D(156)	NR	NR	NR	NR	NR	NR	NR	NR	+	NR	+	NR	NR	NR	No

[a] US FDA (145); [b] MOT (146) ; [c] ISOQOL (147); [d] COSMIN (179); *Psychometric evaluation for both measures was extracted from one article provided by the developer; +: indicate meeting minimum review criteria; -: indicate not meeting minimum review criteria; NR: indicate not reported; HUI2 Health Utility Mark Index 2; HUI3 Health Utility Mark Index 3; QWB-SA Quality of Well-Being Self-Administered; 16D 16Dimension; 17D 17Dimension; AQoL-6D Assessment of Quality of Life 6Dimension adolescent; CHU-9D Child Health Utility 9 Dimension; EQ-5D-Y EuroQol 5Dimension Youth; AHUM Assessment of Health Utility Measurement; CH-6D Child Health 6Dimension; ICC- Intra-class correlation coefficient; MID-meaningfully important difference; +: indicate meeting minimum review criteria; -: indicate not meeting minimum review criteria; NR: indicate not reported.
Conc: concurrent; PR: predictive; Conv: Convergent; Disc: Discriminant; KG: Known group

2.6.3 Cross-cultural validation of generic childhood preference-based health-related quality of life measures and use in sub-Saharan Africa

Of the 181 studies that were identified in the search for this objective, only six publications had used any one of the above generic childhood preference-based HRQoL in sSA. (Fig 2.3) The HUI3 and the EQ-5D-Y were the two generic childhood preference-based HRQoL instruments that attempted cross-cultural validation in six studies in sSA settings. (159) Five of the six studies were for the validation of the EQ-5D-Y and this was in one country alone, South Africa. and the remaining one used HUI3 in Uganda. Tables 2.9 – 2.11 contain a summary of the findings.

Figure 2.3 Flow-diagram for generic childhood preference-based HRQoL instruments applications and cross-cultural validation in sSA.



* application/use of childhood PBMs (n=2) but also discussed aspects of validation

2.6.3.1 The HUI3

i) Linguistic equivalence

The HUI3 language translation involved forward translation only from the original English language into several local languages spoken in Uganda. (180) Members of the study team translated HUI3 into Luganda, Lugisu, Lugwere, Lunyole, Lusoga, Kiswahili, Ateso, Dhapadhadola, Samia and Kubsabiny as they went along interviewing families that they saw. However, the authors didn't specify whether the study members were qualified translators or that a backward translation was done prior to using the translated questionnaire.

ii) Conceptual equivalence

Cognitive interviews were not conducted prior to using the translated HUI3 among children and parents conversant in the translated language in Uganda. The study members translated the HUI3 into the local language as it was being administered to children with spina bifida and their proxies.

iii) Psychometric validation in sub-Saharan African setting

The validation in sSA setting involved comparison of the performance of the HUI3 tool and the VAS marked 0 to 10, and the scores of children and their proxies (completed from the child's perspective). There was moderate correlation (0.488) between the HUI3 and VAS scores of children in open and special schools, and even lower correlation (0.380) between those of children and their proxies.

2.6.3.2 The EQ-5D-Y

i) Linguistic equivalence

The EQ-5D-Y was not translated into any native South African language, with the EQ-5D-Y English UK version used.

ii) Conceptual equivalence

Cognitive interviews to establish conceptual equivalence were done for the EQ-5D-Y in four countries (Germany, Italy, Spain and Sweden), but not in South Africa during the development stage. (159) The adaptation involved revising wording of the adult EQ-5D dimensions into a child-friendly version followed by psychometric validation in five countries including South Africa.

iii) Psychometric validation in sub-Saharan setting

The psychometric validation study demonstrated that the EQ-5D-Y was feasible in South Africa since only 2% of the children had inappropriate or missing responses.(160) Further, there was convergent validity between the EQ-5D-Y and both the KIDSCREEN-10 and KIDSCREEN-27 on similar dimensions (correlation coefficient r= -0.17 to -0.52); the EQ-5D-Y and self-reported general health (r= 0.25-0.27); and with the EQ-5D-Y VAS (Pearson, r= -0.56 to 0.55). The negative correlation between the EQ-5D-Y and KIDSCREEN is because these are scored in opposite directions (higher score on EQ-5D-Y indicates better HRQoL whereas for the KIDSCREEN higher scores mean worse HRQoL). Known-group validity was evident in that the EQ-5D-Y was able to differentiate reported problems between with ($p<0.05$). (160) The p-value refers to the statistical significance between those reporting ‘fair’ or ‘poor’ self-reported health and more problems on EQ-5D-Y dimensions.

Scott et al.(181) in South Africa found that the EQ-5D-Y demonstrated test-retest reliability across both the descriptive component and the VAS scores (kappa, k=0.365 to 0.653 and ICC=0.77); convergent validity with the PedsQL 4.0 amongst the acutely ill children in the study ($p<0.001$); and known-group validity amongst groups with different health states both by composite score (derived from a model in order to generate a single utility score for the descriptive component) and VAS scores (all differences $p<0.001$), but not on composite scores.

In yet another South African psychometric validation study of the EQ-5D-Y, Jelsma and Ramma (182) demonstrated low inter-rater reliability between children and their adult proxies on all dimensions (r <0.33), and no known-group validity between children with and without disabilities on 'doing usual activities', 'having pain or discomfort', and 'feeling worried, sad or unhappy' dimensions and on the EQ VAS scores, mean 88.4 and 87.9, respectively. However, there was evidence of known-group validity on the 'mobility' and 'looking after myself' dimensions ($p<0.001$). The last of the EQ-5D-Y validation studies, by Jelsma, (183) demonstrated that the EQ-5D-Y and the EQ-5D-3L shouldn't be used interchangeably since the former performs better in children and adolescent populations in this setting. The EQ-5D-Y had statistically significant fewer missing responses on both dimensions ($\chi^2=9.404$; d.f.=1; $p=0.002$) and the VAS scale ($\chi^2=26.159$; d.f.=1; $p<0.001$), respectively, than the adult E-5D.

Table 2.9 Summary cross-cultural validation of generic childhood preference-based HRQoL measures in sSA (n=6)

Psychometric property/ Measure (author/ country)	Linguistic translation (equivalence) into any native sSA language		Conceptual equivalence in sSA	Psychometric validation								
				Reliability (dependability/ consistency of an instrument) in sSA			Validity (accuracy/ represent exactly what is on the ground) in sSA					Practicality / Burden in sSA
	Forward	Backward	Cognitive interviews	Internal consistency	test re-test	inter-rater	Content	Criterion	Construct		Responsive-ness	
								Concurrent predictive	Convergent/ Discriminant	Known group		
HUI-Mark3 (Sims-Williams et al/ Uganda)	+	-	-	NR	NR	- +	NR	NR	-	NR	+	+
EQ-5D-Y (Scott et al 2017/ South Africa)	-	-	-	NR	+	NR	+	-	+	+ - +	+ +	+
EQ-5D-Y (Jelsma & Ramma 2010/ South Africa)	-	-	-	NR	NR	-	NR	NR	NR	-	NR	+
EQ-5D-Y EQ5D (Jelsma 2010/ South Africa)	-	-	-	NR	NR	NR	NR	+	NR	NR	NR	+
EQ-5D-Y (Wille et al 2010/ South Africa)	-	-	-	NR	NR	NR	NR	NR	NR	NR	NR	+ + +
EQ-5D-Y (Ravens-sieberer et al 2010/ South Africa)	-	-	-	NR	NR	NR	NR	NR	+	+	NR	+

+: indicate meeting minimum review criteria; -: indicate not meeting minimum review criteria; NR: indicate not reported.

Table 2.10 Summary of children and adolescent preference-based HRQoL measures used in sSA

Measure	Country HRQoL used	Author, year	Age range (yrs.)	Study objective	Study design	study sample size	Methods of recruitment/ selection	Results	Study evaluation
HUI3	Uganda	Sims-Williams HJ, Sims-Williams HP, Mbabazi Kabachelor E, Warf BC. Arch Dis Child. 2017;102 (11):1057-61.(180)	10-14	ascertain the QoL of surviving children with spina bifida	Cross-sectional	68	Surgical cases Both children and parent (proxies) completed HUI3 and VAS. The parent (proxies) reported HRQoL from their own perspective	There was high completion rate by both children and adult proxies, 94% and 100% respectively. The HUI3 HRQoL scores for children and adult proxies was correlated. However, there was poor correlation between QoL VAS scores of children and caregivers.	The HUI3 was found to be acceptable and able to distinguish between children and adult proxy scores in this population. However, there was no agreement between QoL VAS scores of children and that of their adult proxies.
EQ-5D-Y	South Africa	Jelsma J, Ramma L. Health and Qual Life Outcomes. 2010;8(1):72.(182)	7-12	to compare the HRQoL between healthy children and those with disabilities; and between parents' proxies and children's own health	Cross sectional	628	All children in eligible grades	There was low completion rate (28.2%-28.4%). The EQ-5D-Y demonstrated practicality (few missing responses). Children with disabilities (mean VAS score=87.9) didn't perceive their HRQoL worse compared to those without disabilities (mean VAS score= 88.4). Low correlation ($r=0.33$, $p<0.001$) between children and female parent proxies' VAS scores in open schools and no correlation between children and parent proxies in special schools ($r=0.16$, $p<0.22$).	The EQ-5D-Y seems was feasible in this population judging by the few missing responses. However, the low response rate by children could not be ascertained whether it was due to lack of comprehension of the measure by the children in this setting.

HUI3 Health Utilities Index Mark 3, *EQ-5D-Y* EQ-5D Youth

Table 2.11 Quality appraisal of studies identified for application or adaptation of generic childhood preference-based HRQoL measures in sSA (n=3)

	CASP case-control study checklist*	Sims-Williams et al (180)	Jelsma & Ramma (182)	Wille et al (159)
1	Did the study address a clearly focused issue?	Y	Y	Y
2	Did the authors use an appropriate method to answer their question?	Y	Y	Y
3	Were the cases recruited in an acceptable way?	Y	Y	Y
4	Were the controls recruited in an acceptable way?	Y	Y	N/A
5	Was the exposure accurately measured to minimise bias?	N/A	N/A	N/A
6a	Aside from the experimental intervention, were the groups treated equally?	Y	Y	Y**
6b	Have the authors taken account of the potential confounding factors in the design and/or in their analysis?	Y	N#	Y
7	How large was the treatment effect?	Mean VAS score=0.61	Mean VAS score=0.5	N/A
8	How precise was the estimate of the treatment effect?	N/A	N/A	N/A
9	Do you believe the results?	Y	Y	Y
10	Can the results be applied to the local population?	Y	Y	Y
11	Do the results in this study fit with other available evidence?	Y	Y	Y

*The CASP case-control checklist was used because there is no CASP checklist for appraising cross-sectional studies which was the case with all the three studies; Y=Yes; N/A= Not Applicable; #Was not part of the objectives of the study; **Was a multi-national study across seven countries

Summary assessment of literature review results

The review found that in previous studies both the HUI3 and the EQ-5D-Y did meet some but not all the cross-cultural validation criteria. The HUI3 was translated into a native sub-Saharan language but the translation did not involve backward translation nor cognitive interviews to establish conceptual equivalence. Additionally, some but not all the psychometric properties were evaluated for the HUI3.

On the other hand, previous studies of the EQ-5D-Y have established conceptual equivalence but have not translated the measure into a native language spoken in sSA. Like the HUI3, not all the psychometric properties had been evaluated in a single study for the EQ-5D-Y. When all the findings from these individual EQ-5D-Y studies are considered together, it is evident that the psychometric properties have been evaluated in great depth for the EQ-5D-Y except for internal consistency, which was not reported in any of the reported studies. In summary, the findings show that most psychometric properties were not reported by individual studies. However, of the properties reported, the majority did meet the psychometric criteria.

2.7 Discussion of literature review findings

This review aimed to identify the available generic childhood preference-based HRQoL instruments and to determine the extent to which they have been used and/or cross-culturally validated in the sSA region. Ten measures were identified that were all originally developed in HICs and mostly adapted from adult versions of the measures. The HUI3 and the EQ-5D-Y were the only two measures used in child and adolescent populations in sSA. The review further highlighted that identified generic childhood preference-based HRQoL instruments differed greatly on several other issues.

2.7.1 Existing childhood preference-based health-related quality of life instruments

Descriptive system

There is a need for greater involvement of children and adolescents themselves in the development of conceptual frameworks underlying childhood preference-based instruments. This view is supported by previous findings. (107) The CHU-9D is the only measure identified by the review whose dimensions were generated directly from children through qualitative interviews and analysis (although there was no data for the CH-6D). (107) The other eight instruments had employed some form of top-down approach. This clearly demonstrates the lack of involvement of children in development of preference-based HRQoL instruments. Suffice to say that if an instrument is to measure HRQoL in children and adolescent, it should involve them not only in the latter developmental stages but right from the start. Involvement of children and adolescents at the initial stage is a fundamental point as this ensures instruments' capturing domains that matter to this population group. (107, 142)

Further, the generic childhood preference-based HRQoL measures have largely been adapted from adult versions of the measures. Even though it has been demonstrated that adult domains/dimensions can relate to children and adolescents, (159, 160, 184, 185) it is also well known fact that children have different perspectives and goals. (142) This implies that dimension structures should ideally be developed directly from children themselves within specific contexts. The current generic childhood preference-based HRQoL measures, except CHU-6D, have in a way 'imported' domains that risk measuring what not matters.(107) This could not only negatively impact on concepts being measured but also, when used, in an economic evaluation lead to misleading conclusions and policy.

The instruments also differ in terms of the age range of the children they can be utilised in: 6-16years for HUI2 and HUI3; 7-12 years for CH-6D; 7-17 years for CHU-9D; 8-11years for

EQ-5D-Y and 17D; 12-15years for 16D; 12-18years for QWB-SA and ≥15 years for AQoL-6D. While there is still ongoing debate on the range of age an instrument should cover, proponents for an all-inclusive age range for children and adolescent argue that it is practical to have one instrument that captures HRQoL across this whole group. (81, 126) It has similarly been asserted that children are at different levels of growth (need and development) and therefore the need to having instrument with different age ranges especially that which separates children from adolescents. (126) While this could be advantageous, it does bring in some methodological and practical challenges for example to do with pulling data from different age groups. (81) Additionally, it can be argued that if there were differences in development and need between children and adolescents, the same should apply to adults. For example, aspirations and needs of a 20year old adult are different from the one aged 40, which are further different from that of a 60-year-old. To overcome this, some psychometric instruments (such as SF-36) have population norm data to the measure published in the manual. (186) The user can then identify where these different age groups sit on the scale (i.e., where their respective means and ranges are). The skill of a well-developed generic measure is in generating content that is relevant across the whole of the adult population. This is a bit different in paediatrics as there are developmental issues at play.

Underlying utility theory

There is great diversity in not only availability of utility values but also how these were generated. Three instruments elicited utility values using TTO alone (CH-6D, AHUM and AQoL-6D), combination of TTO and VAS in the case of EQ-5D-Y, RS (16D, 17D and QWB-SA), SG for CHU-9D, and a combination of SG and RS for HUI2 and HUI3. Literature has been dominated with largely choice-based techniques (TTO and SG) as better techniques since choice is involved. (105) In this regard, seven of the measures (CH-6D, AHUM, AQoL-6D,

EQ-5D-Y, CHU-9D, HUI2 and HUI3) either entirely used these methods or a combination of methods that involved a choice-based technique.

Of these ten instruments, only the AQoL-6D has used adolescents in elicitation of utilities in four pacific countries of Australia, Fiji, New Zealand and Tonga. All the other instruments used an adult general population in utility generation either through adult elicitation or proxies, or a combination of the two. Due to the complicated nature of eliciting utilities, even among adult population, this remains tricky and the main reason why there is lack of childhood utility values. The use of adult preferences even when generated using children health states like CHU-9D is still debatable. (137) New techniques like DCEs, are under development in order to overcome this challenge although this technique has not escaped its own level of controversy. For example, DCEs are said to lack the ability to directly provide utility values and hence need to be used in conjunction with other valuation techniques. (144) A more recent approach is the use of BWS which is different from traditional DCE. In a BWS, a respondent is presented with a scenario and one at a time indicates best and worst attribute for a number health states. (139) In a feasibility study of using BWS for CHU-9D, this valuation technique was found to perform much better compared to TTO, SG as well as the traditional DCE. (139) Even so, the BWS utility valuation technique is at an exploration stage at this time.

Psychometric properties

Validity

The findings from the review show that almost all the identified generic childhood preference-based HRQoL instruments had established practicality/burden and convergent validity. However, there was variability in achieving other psychometric properties. Conceptual framework had not been established as most identified measures did not involve children, although some involved qualitative data from adults, in the initial phase of formulating

dimensions. There is lack of criterion validity which might be due to lack of agreed consensus on a 'gold standard' HRQoL instrument to compare a new instrument against. (187) Lastly, since most of these instruments have been developed in HICs, only the EQ-5D-Y and AQoL-6D adolescent were adapted cross-culturally at the initial stage. This is further discussed below.

Linguistic and conceptual equivalence

It has been long observed that children are still developing their physical and functioning capacity and therefore using directly or just doing a forward and backward translation of an instrument is not ideal. (126) This raises the need not only to use child-friendly instruments but also need for a conceptual and linguistic equivalence in adapting the instrument. (188) A number of studies have also indicated the need for not only a tool for this population but also one that should take into consideration cultural context. (151) Several studies have similarly highlighted the need to assess health from children and adolescents. (159) In a review of use of children HRQoL in resource-limited settings by Paltzer et al (189), it was found that the most common adaptation involved a forward and backward translation without subsequent need for conceptual equivalence. However, even if there was conceptual equivalence of the included items, there may well be other concepts and items that are missing. This may be due to some concepts which are important in the local setting but not for HIC or because they are important for children but not adults or both. This is explored in chapter 7 of this thesis.

In another systematic review (190) regarding cross-cultural adaption of instruments, it was established that many studies take on an 'absolutist' view conception of health a finding highlighted previously. (188) An earlier work on equivalence, identified three approaches for translating and adapting questionnaires cross-culturally: absolutist, universalist and relativist. (191) These theories are discussed in detail in chapter 3. Briefly, an absolutist approach

assumes that constructs are understood in the same way across cultures. This therefore means that culture has minimal impact and so the only thing that is said to be different are the values that are attached to different dimensions. (191) The main problem with this absolutist approach, however, is that it 'imposes' concepts onto another culture without first establishing if this were the case.

In summary, there was substantial evidence of psychometric performance on some aspects (convergent, discriminant, practicability/burden). There is lack of evidence on criterion validity, conceptual framework involving the children in development of new instruments. The content validity was rarely evaluated as this would involve qualitatively asking the respondents if they consider the measure covered all relevant dimensions of health. The AQoL-6D adolescent and EQ-5D-Y are the only two instruments that had employed some conceptual and linguistic equivalence during the developmental stage. (140, 159)

2.7.2 Preference-based health-related quality of life instruments in sub-Saharan Africa

The use of generic childhood preference-based HRQoL instruments in sSA settings is sparse. In part, this may reflect a lack of policy relevance in sSA settings; although funders have set standards for childhood outcomes as evidenced by SDG goal three, which have primarily focussed on reducing child mortality. (192) While this is important, there should be some effort to measure the HRQoL gained through the SDG interventions using the existing generic preference-based measures. It may also reflect the fact that the measurement of HRQoL in children has lagged that of their adult counterparts. Eliciting robust self-reported information across the developmental stages of childhood remains a challenge. To be able to use childhood preference-based HRQoL instruments effectively in sSA requires more research work. The

work should aim at developing methods of eliciting reliable information from children themselves via questionnaires and other modes of administration.

These literature review findings suggest that there is potential for the EQ-5D-Y to be used in sSA, but more work is needed to investigate its conceptual equivalence in this setting. There has been no cross-cultural validation in sSA of the ten identified childhood preference-based HRQoL except for the HUI3 and the EQ-5D-Y. The cross-cultural validation of the HUI3 involved translation and cognitive interviews with children and parents, and psychometric evaluation. (180) As outlined above, some adaptations of existing measures have taken an absolutist approach. This means that, without seeking to establish a conceptual equivalence, measures are exported, after crude translation, into the local cultural milieu. (189, 190) However, local culture has an impact on health and health-related concepts. (193) A qualitative study that looked at conceptualization and operationalization of 'health' from among adults in Kenya found that the definition mainly focussed on positive and negative subjective perspective aspects. (194) This included contentment, cleanliness, corporeal capacity, co-operation and completeness of an intending individual. This emphasizes the individual role to make judgements based on their own feelings but also taking into consideration that of the world around them. The same study also established that there was no clear distinction between body and mind as is the case with most existing generic health instruments. This all points to the need for local involvement of intended participants in understanding and defining health concepts elsewhere including the sSA setting.

This was further emphasized through the work of adapting the WHOQOL-BREF, a generic HRQoL instrument, in Malawi. (195) The work was taken in context of maternal, neonatal and

child health measurement and in the fact that although QoL can be measured using acceptable global instruments it is important to consider the context where these are used. For example, this study found out that although people might be sick, hungry or without money they still answered that they were fine when asked how they were doing. The application of instrument developed from HICs raises issues of transferability due to cultural and socio-economic differences. (196) While there could be similarities in domain structure cross-culturally, there will likely be differences too. It is therefore important to establish conceptual equivalence before using instruments cross-culturally. Developers of the EQ-5D-Y themselves had suggested that future studies should consider other aspects of HRQoL that are important to children and adolescents for inclusion in the measure. (197) There are other issues including age bands as well as the best responder (whether child or adult proxy) (198) which also needs addressing cross-culturally. Moreover, the childhood preference-based HRQoL instruments have only been applied in a small selection of sSA countries, raising questions about their applicability across this sSA region. Apart from Uganda, where the HUI3 was applied, the use of the generic childhood preference-based HRQoL instruments was limited to South Africa alone. However, South Africa is not representative of most economies in sSA as it is ranked as an upper middle-income country according to the World Bank. (159, 160) The use of the EQ-5D-Y and HUI3, and indeed any other childhood preference-based HRQoL instrument therefore needs further evaluation in other sSA settings.

It is important to mention some of strengths and weaknesses of the literature review. The search generated a high volume of data, and a large volume of articles were excluded during the title screening stage, which could have led to some relevant articles being excluded. Nonetheless, the findings related to the first study objective are largely consistent with the most recent review of existing generic childhood preference-based HRQoL measures. (122) Although the

measures were appraised for their psychometric properties, it is worth highlighting that this was limited to the original work. Despite the fact that psychometric evaluations for widely used measures, such as the HUI2 and HUI3, are ongoing, for practical reasons only original psychometric evaluations of each measure were included. (173) The searches were also limited to peer reviewed material, which could have overlooked important evidence in the grey literature. In relation to this, all studies other than those in the English language, were excluded which could also have excluding some local studies. Finally, the quality appraisal of the studies applied in sSA settings was performed using the CASP case-control checklist due to the lack of toolkits designed specifically for cross-sectional studies at the time of the review.

2.7.3 Conclusion of literature review findings

In conclusion, the findings of this literature review (199) show that application of generic childhood preference-based HRQoL instruments in sSA settings has hitherto been limited to the HUI3 and EQ-5D-Y. Most adaptations of existing measures take an absolutist approach, which assumes that measures can be used across cultures. There is a need to undertake cross-cultural linguistic and conceptual equivalence, testing and psychometric validation across a range of sSA cultural contexts. Unless cross-cultural validation of preference-based HRQoL measures is robust when applied in the sSA region, the application and interpretation of economic evaluations reliant on preference-based outcome measures will fail to reflect local relevance in this part of the world.

While the literature review identified and assessed identified childhood preference-based HRQoL instruments, it did not propose the use of any instrument against the other in the sSA. The next step in this doctorate research was to select an instrument that could be cross-

culturally adapted in for use in Malawi. In order to achieve this, a systematic scoring method was adapted solely for this purpose alone. The next section deals with this selection process.

2.8 Selection of generic childhood preference-based health-related quality of life instrument for use in Malawi

The scoring method developed in order to assess suitability of the generic childhood preference-based HRQoL instruments for cross-cultural validation (adaptation and validation) in Malawi, uses a four criteria list. The criteria was adapted from the Brazier and Deverill checklist for selecting a measure for use in economic evaluation. (154) The checklist was expanded by including all the 14 properties consolidated in section 2.4 in Table 2.4 above (replicated in Table 2.12 below). Each of the ten generic childhood preference-based HRQoL instruments identified in the literature review was scored against these 14 psychometric properties. Secondly, each instrument was also scored for the descriptive system development process and the degree to which there was children involvement. A third criteria for which each instrument was assessed was the underlying utility theory- available for and involvement of children in generation of these. The fourth criteria instruments were also scored against was the age range covered by the instrument: children, adolescents or both. Finally, a total score was obtained across all the four criteria to identify potential candidate for cross-cultural validation in Malawi.

The detailed scoring for each of the ten identified generic childhood preference-based HRQoL instrument is as follows:

i. a comprehensive psychometric property list (14 points i.e., one point for each of the psychometric properties) as per consolidation of the four guidelines in Table 2.12

ii. involvement of children and adolescent in the descriptive system development (three points i.e., one point each for involving children at the time of development, pilot and psychometric evaluation)

iii. underlying utility theory (five points i.e., one point for use of choice-based utility valuation technique; two points for utility values developed from among children/adolescents; one point for availability of utility values in several countries; and one point for utility values developed in sSA)

iv. age (one point i.e., if it covers across children and adolescent age groups).

2.8.1 Criteria 1- Psychometric property scoring

Table 2.12 provides the psychometric property scores for all the identified childhood instruments. Conceptual framework was established in development of all of the instruments. The EQ-5D-Y was the only one that had conceptual, cross-cultural and language equivalence its development involving a sSA country. The CHU-9D conceptual framework is in its development and has also demonstrated good practicality, face and construct validity across cultures. (200) However, the cultural and language equivalence was not reported in that cross-cultural adaptation paper. It can only be speculated that since the UK (where instrument was developed) and Australia (recipient) have cultural similarity, this might have been deemed unnecessary. The findings from this literature review show that almost all the identified generic childhood preference-based HRQoL instruments had established practicality/burden and convergent validity. Whereas there was absence of criterion validity due to lack of a 'gold standard' HRQoL instrument to compare a new instrument against, reporting on other psychometric properties was also missing.

2.8.2 Criteria 2-Involvement of children in development of descriptive system of preference-based health-related quality of life instruments

This was limited to only the CHU-9D that was developed entirely with children and thus claims good content validity.(157)

2.8.3 Criteria 3- Underlying utility theory development

2.8.3.1 Choice-based valuation technique:

All the identified generic childhood preference-based HRQoL instruments had used either one or a combination of TTO, SG and/or RS except for 16D, 17D and QWB-SA that used RS alone.

2.8.3.2 Children's values

Of all the identified instruments, only AQoL-6D and 16D have elicited preference directly from adolescents during the primary valuation. (122) The HUI2 and CHU-9D are the other two instruments that have children utility values elicited from adult proxy, adults completing in behalf of children during the primary valuation task. However, the CHU-9D and EQ-5D-Y have since developed utility values directly from adolescents in subsequent valuation in Australia and Japan respectively. (142, 201) All the other instruments have done so using adult values.

2.8.3.3 Availability of values in more than one country

Half of the instruments have more than one utility values generated either during primary or subsequent valuation other than the country the instrument was developed. (122, 144) The exception is CH-6D (no information), the CHU-9D, AHUM, 16D and HUI2 have utility values in multiple countries. The EQ-5D-Y is the only instruments with utility values developed across thirteen countries across Europe (nine countries), USA, Japan, New Zealand and Africa (Zimbabwe). (144)

2.8.3.4 Availability of utility values in sub-Saharan Africa

The EQ-5D-Y is the only childhood preference-based HRQoL instrument that has utility values developed in the sSA region (Zimbabwe), and these were elicited in adults. (202) Even though the utility values were developed from adults using EQ-5D-3L, these are used in evaluations in both adult and children population.

2.8.4 Criteria 4: Age range for which the instrument covers

The age range opted was 7-17 years. The reason for choosing seven as the minimum age is because it has been established that children of this age are able to comprehend and self-complete health status questionnaire without need for proxies. (159) It is also preferable that an instrument be applicable to different age groups so as to make it easier to compare interventions across childhood. (164) There are different schools of thoughts against a common weight for health states between children and adolescents. Evidence, however, show that various HRQoL instruments utility values correlate between those of children and adolescents as opposed to children and adults. (203)

Of the identified instruments, AQoL-6D (≥15years), 16D (12-15years), QWB-SA (12-18years) do not qualify on this criterion alone. The CHU-9D is for children 7-11years but has been validated for 6-7 years and 11-17 years. (162) The EQ-5D-Y is for children aged 8-11years, but the adult EQ-5D-3L version is recommended for use in adolescents ≥16years and a choice of the youth version or adult version for age 12 to 15 years. HUI2 and HUI3 are the only other instruments that covers across all children and adolescent age groups (6-16years).

Table 2.13 provide a summary of the ten identified instruments across all the four scoring criteria listed above. The Table also presents the total score for each instrument.

Table 2.12 Summary psychometric properties for all identified preference-based measures as reported by developers

Psychometric property/ Measure	Conceptual framework [a,b,c,d]	Cross-cultural & language adaptation[a,b,c,d]	Reliability (dependability/ consistency of an instrument) [a,b,c,d]			Validity (accuracy/ represent exactly what is on the ground) [a,b,c,d]							Interpretation of scores[a,b,c,d]	Practicality/ Burden [a,b,c,d]	Validated in sSA	Total Score*
			Internal consistency [a,b,c,d]	test re-test [a,b,c,d]	inter-rater [a,b,d]	Content [a,b,c,d]	Criterion[a,b,c,d]		Construct [a,b,c,d]			Responsiveness[a,b,c,d]				
							Conc	Pr	Conv[a]	Disc[a]	KG[a]					
HUI2 & HUI3* (165, 173)	NR	+	0	+	NR	NR	NR	NR	NR	NR	+	+	+	NR	++	5
QWB-SA(169, 171)	NR	NR	NR	NR	NR	+	NR	NR	+	NR	NR	NR	NR	+	NR	3
17D (168)	+	NR	+	+	-	+	NR	NR	NR	NR	+	NR	NR	+	NR	5
16D(167)	+	NR	++	+	NR	+	NR	NR	NR	NR	+	NR	NR	+	NR	5
AQoL-6D(166, 176)	+	+	NR	NR	NR	+	NR	NR	+	NR	NR	NR	NR	NR	NR	4
CHU-9D(157, 165)	+	NR	+	NR	NR	+	NR	NR	+	NR	+	NR	NR	+	NR	6
EQ-5D-Y(159)	+	+	NR	+	+	-	NR	NR	-	NR	+	NR	+	+	++	7
AHUM(158, 165)	NR	NR	NR	NR	NR	+	NR	NR	NR	NR	NR	NR	NR	NR	NR	1
CH-6D(156)	NR	NR	NR	NR	NR	NR	NR	NR	+	NR	+	NR	NR	NR	NR	2

*score= (total for +) – (total for -). 0= mixed evidence

Conc: concurrent; PR: predictive; Conv: Convergent; Disc: Discriminant; KG: Known group

Table 2.13 Checklist for judging preference-based measures for use in economic evaluation adapted from Brazier and Deverill (154)

Characteristic[1]		CH-6D	AHUM	EQ-5D-Y	CHU-9D	AQoL-6D	16D	17D	QWB-SA	HUI3	HUI2
Psychometric properties (across 14 properties, ref Table 2.12) =14points		2	1	7	6	4	5	5	3	5	5
Dimension development children involvement (across 3 steps, ref Table 2.6) = 3points	development	NR	0	0	1	0	0	0	0	0	0
	refinement	NR	1	0	1	0	1	1	0	0	1
	testing/pilot	NR	0	1	1	1	1	0	0	0	0
Underlying utility theory[2] (over 4 criteria, ref Table 2.7)= 5points	technique choice-based	1	1	1	1	1	0	0	0	1	1
	children values[2]	0	0	0	1	2	2	0	0	0	1
	utilities available in ≥1 country	0	0	1	1	1	1	0	0	0	1
	available tariffs developed in any country in sSA region[3]	0	0	1	0	0	0	0	0	0	0
Age range (instrument covering 7-17yr) (Y=1/ N=0) =1point		1	0	1	1	0	0	1	0	1	1
OVERALL SCORE (23)[4]		4	3	12	13	9	10	7	3	7	10

CH-6D Child Health 6D, AHUM Adolescent Health Utility Measure, AQoL-6D Assessment of Quality of Life 6D, EQ-5D EuroQol 5D, EQ-5D-Y EQ-5D Youth version, CHU-9D Child Health Utility 9D, 17D 17 Dimension, QWB-SA Quality of Well-Being self-administered, HUI2 Health Utility Index Mark 2, HUI3 Health Utility Index Mark 3, SF-6D Short Form 6D

[1]all measures scored one point each for self-reporting and acid test (stated v revealed preference- not possible in reality to measure the latter)
[2]children values: elicited from children/adolescents=2; elicited from adults with children in mind=1; elicited from adults=0
[3]tariffs available in at least in one sSA country
[4]psychometric properties (14) + dimension development (3) + utility valuation (5) + age (1)

2.8.5 Overall scores and way forward

Four instruments (CHU-9D, EQ-5D-Y, 16D and HUI2) scored ten or more out of the total score of 23 (psychometric properties=14; dimension development=3; underlying utility theory=5; and age=1). The CHU-9D and EQ-5D-Y scored highest and second highest respectively. These two (CHU-9D and EQ-5D-Y) were selected as potential candidates for cross-cultural adaptation and psychometric validation in Malawi.

Before embarking on the cross-cultural adaptation work, the first step was to contact the developers for permission to do so. Dr. Katherine Stevens at the University of Sheffield developed the CHU-9D and was contacted. Although the University of Sheffield hold the CHU-9D rights, translation and distribution is through an agency: ICON Clinical Research (UK) Ltd. The ICON translation cost quoted at US$11,716 (Appendix 3) was prohibitive within the budget of this doctorate work. The developers were contacted to consider reduction of translation cost if this researcher could get involved in the translation process. Although the involvement was granted, this was only at sub-optimal level as the agency were to lead the whole process and the reduction in translation cost was denied. The lack of full involvement (and leading this work) plus the translation cost were deemed barriers to the CHU-9D cross-cultural adaptation. For these reasons, attempt to cross-culturally adapt the CHU-9D in Malawi was dropped.

The EuroQol Group were similarly approached for the cross-cultural adaptation of the EQ-5D-Y in Malawi. The EuroQol Group not only accepted to the full involvement but granted this researcher to lead this work in Malawi. Additionally, the work would be reviewed by the EuroQol version management committee (VMC) without charge. The VMC is a team within EuroQol Group responsible for translation of all EQ-5D versions. (Appendix 4)

An important aspect of psychometric validation is to test how the instrument performs vis a vis other measures (construct validity). The Pediatric Quality of Life Inventory ™ Version 4.0 (PedsQL ™ 4.0) child self-report and the PedsQL ™ 4.0 teen self-report were identified as potential candidates against which to validate EQ-5D-Y. The choice for the PedsQL ™ 4.0 was based on several factors. Firstly, PedsQL ™ 4.0 has been used universally for psychometric evaluation against the EQ-5D-Y including the very first EQ-5D-Y psychometric validation study. (160, 204, 205). The PedsQL ™ 4.0 is one of only a few generic HRQoL instruments that had previously validated alongside the EQ-5D-Y in a sSA setting. (181) Subsequently, the PedsQL ™ 4.0 is being featured extensively for psychometric evaluation against EQ-5D-Y-5L. (206-208) Currently, there is work underway to develop utilities for the PedsQL ™ 4.0. (209) Further, the focus of this research was to understand HRQoL from the perspective of children and adolescents in this setting. There was therefore a need to have an instrument that measures health across the children and adolescent age groups. The PedsQL ™ 4.0 child self-report covers age 7 to 12 years and the PedsQL ™ 4.0 teen self-report covers age 13 to 18years. Additionally, there was no translation costs if this work would be carried out by this researcher. (Appendix 5) The PedsQL ™ 4.0 self-report replaced the CHU-9D for this purpose.

The next few chapters will provide detailed steps that were undertaken for the cross-cultural adaptation and psychometric validation of the EQ-5D-Y and EQ-5D-Y-5L in Malawi. This also includes the adaptation and validation of the PedsQL ™ 4.0 child self-report and the PedsQL ™ 4.0 teen self-report so that they could be used as validating measures for EQ-5D-Y.

2.9 Summary of chapter 2

The chapter started with description of generic childhood preference-based HRQoL

instruments. Generic preference-based HRQoL consists of a descriptive system as well as utility values for all the health states defined by the descriptive system. While there is overlap for psychometric properties for assessing quality of health instruments, 14 properties were consolidated criteria from four psychometric guidelines: FDA, ISOQOL, MOT and COSMIN.

The chapter then reviewed ten childhood preference-based HRQoL instruments identified from the literature published between 1990-2017. All ten instruments were developed in HICs. The instruments were assessed for their descriptive system, underlying utility theory and psychometric properties. Most childhood preference-based HRQoL instruments have descriptive systems developed using a decomposed top-down approach. The utility values have mostly been developed from an adult population using the choice based TTO and SG valuation techniques. The non-choice-based RS is another valuation technique that has been used to generate utility values. While reliability, construct validity, known-group validity and practicality/burden are common psychometric properties evaluated, criterion validity and responsiveness are rarely evaluated. The review further established that the use and validation of the identified generic childhood preference-based HRQoL instrument in the sSA region has been sub-optimal. Cross-cultural adaptation and use in the sSA region have been limited to the EQ-5D-Y and HUI3.

Assessment of four scoring criteria identified the EQ-5D-Y and CHU-9D as the two suitable candidates for cross-cultural adaptation and validation in Malawi. However, due to the high translation costs as well as minimum involvement of this researcher in the adaptation process, the CHU-9D was dropped and replaced with the PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report. The following chapters will present the findings from this

cross-cultural adaptation using a mixed methods approach. First, the next chapter discusses the justification for using a mixed methods approach.

The findings from this review have been presented at various conferences. In addition, a paper has been published based on the material in the chapter. (199)

3 Chapter 3- Justification and selection of a mixed-methods approach

3.1 Chapter overview

This chapter introduces the mixed methods approach in this thesis and provides its rationale. In chapter 2, the EQ-5D-Y and subsequently the EQ-5D-Y-5L were identified as candidates for cross-cultural validation in Malawi. Cross-cultural validation ensures that instruments maintain the same concepts and perform in a similar way in a target culture as in the source culture. To achieve this, a two-stage process is necessary: i) adaptation (linguistic and conceptual evaluation), and ii) psychometric evaluation (validation). Whereas adaptation may refer to modifying the descriptive system of an existing instrument (210), in this thesis it has been used consistently with reference to linguistic and conceptual equivalence. The purpose of linguistic and conceptual evaluation is firstly to establish that the translated words have the same meaning in the target setting as the original source document. The second reason is to make sure that the concepts have the same meaning in the target as in the original cultural context. Both these steps belong to stage one and are carried out qualitatively. The second stage of psychometric evaluation involves carrying out a series of quantitative psychometric tests on the linguistically and conceptually adapted version. This is done to establish whether the instrument 'behaves' or performs in the same way in the target language/culture as in the source language/culture. For these reasons, cross-cultural validation of a scale requires using both qualitative and quantitative psychometric processes.

3.2 Overview of qualitative approach

Whatever approach is employed, the aim of qualitative research methods is to ensure that the target population 'voices' are heard. Qualitative approaches do offer insight into "meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things". (211) Qualitative research is therefore about understanding a phenomenon from people's perspective or getting views of the topic under study. (212) Data is usually gathered by observation (ethnography and participant observation) or through interviews or focus groups. Ultimately, qualitative methods answer the "why" and "how" type of research questions using the different designs, some of which are briefly discussed below.

Ethnography involves living with the people or community under investigation in order to get a 'first-hand' understanding of their behaviour. (213) This kind of research could range from a few months to years depending on nature of research and objectives. One reason why this kind of research is conducted over a long period is because by nature people tend to 'conceal' their behaviour when observed. It might therefore require considerable time before people start to act normally and researchers are accepted as part of the community they are studying.

Focus group discussion (FGD) has been described as "a carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment' (214) The goal of FGD is to get views and opinions about a certain topic from the people under study using a semi structured topic guide. The groups made up of six to 12 people, depending on participants age, may last anything between 45 to 60 minutes. The discussions are guided by a facilitator with the aid of a topic guide. In this way the researcher is able to see the world from the participants' point of view. (215) The group setting provides participants a comfortable environment where they don't feel intimidated and are able to 'naturally' express themselves better as opposed to a one-to-one interview. A group setting not

only allows participants respond to questions from a facilitator but also affords them an opportunity to discuss these points among themselves. FGDs have been successfully used with both adults and children, sometimes with the use of developmentally appropriate prompts. (216) Such prompts like ice-breakers, enable participants to open up and comfortably participate.

In-depth interviews involve a one-to-one discussion with a participant. (212) The goal might be to get insight about an individual's feelings and emotions on a subject. It may also be used to gather information on sensitive matters that might not be appropriate for a group situation. Insight can also be sought on a questionnaire layout, language used if it is common to respondents, or any other aspect that might be deemed important. The in-depth interview is frequently used by the researcher to develop an understanding of a concept. Discussion is usually based on a semi-structured interview guide.

Cognitive interviews are more structured than most qualitative research methods. They are quite specific to psychometrics having grown out of the need to have a detailed and systematic way of evaluating survey questions and instruments. (217) Through cognitive interviews, a researcher can establish whether the concepts meaning are understood by respondents. Further, cognitive interviews enable a researcher to explore respondents' thought process when presented with a task such as answering a questionnaire. By observing, thinking aloud as well as verbal probing, a researcher can establish respondents' answers to questions as well as factors influencing those responses.

In summary, qualitative approaches are important in development, designing as well as in establishing equivalence, comprehension and refinement of concepts. (218) No one qualitative

method is better than the other, (219) as this depends on the objectives of the study. For example, it is claimed that FGDs generates rich volume of data (220), but it has the potential of tagging along participants to opinions of others. (219) The richness of data is said to be in two respects: elaborateness and the aggregated total data since the group discussions involve more than one individual opinion. While this may be true, some studies have found that for matters of personal or family opinion, individual interviews yield more participation. (221) The problem being that due to preference or characteristics for some, a one-to-one interview might be too daunting and not produce the desired results. The same could said for a group discussion for others although it has been argued that there is safety in numbers. (222) Additionally, with assistance from their peers' children could express themselves. (219) The facilitator can also ensure that the views of such persons are directly solicited.

3.3 Overview of quantitative approach

In general, a quantitative approach addresses the "what" and "how much" question: proportion, or number of people responding in a certain way. Quantitative approaches are concerned with empirical or experimental tradition and the epistemological stance that there is a single factual truth to be discovered. (223) This is different from the epistemological stance of much qualitative work where reality is socially constructed and is therefore determined by people's experience rather than being a single truth. Statistical significance is often calculated to establish the truth of the phenomenon from a population. A representative sample, as it is practically impossible to study the whole population, is studied and results inferred. (224) The statistical significance of 95%, a threshold broadly accepted in statistics world, is usually calculated and conclusions drawn upon it. This means findings could be said with 95% certainty that any observed differences are based on true difference and not due to chance. Table 3.2 provides a summary of qualitative and quantitative methods.

Table 3.1 Summary of qualitative and quantitative research methods

Attribute	Qualitative	Quantitative
Purpose	Understand phenomenon complexity- what, how or why	Understand magnitude of difference- how much.
Sample size	Smaller	Generally larger
Sampling method	Selective purposive (judgmental)	Probability sampling
Data inference	Context specific- no one pre-existing reality as a result gives more local understanding of a phenomena	Generalizable- one single reality that is being explored but might local understanding
	Provides understanding of concept	Provides quantitative output
Data collection	Generally, takes more time	Depends on the size of the study. Generally, takes relatively short time compared to qualitative study
Data analysis	Often time consuming	Relatively takes short time
Data interpretation	Mostly inductive- detailed description to allow broader context	Generally deductive – based on apriori hypothesis
Relationship to research process	Subjective- researcher is part of the construct being studied.	Objective- researcher and what is being measured are separate
Bias	Researcher may have an influence results interpretation	Depending on design of the study researcher may have little, if any, influence on results

Source: own synthesis based on (212, 225-227)

3.4 Advantages of the mixed methods approach.

In general, qualitative and quantitative approaches to research methods address different types of questions. Each is appropriate for different aspects of this thesis research work. As discussed above, qualitative approaches help understand meanings to "how" and "why" type of questions whereas quantitative addresses "what" or "how much". In practice often qualitative and quantitative approaches are used to complement each other. Some, 'purists', on both ends of the spectrum contend that these approaches are from two opposite ends of the continuum and shouldn't mix. (223) Others have even gone further to advocate for 'incompatibility thesis' between qualitative and quantitative methods. (228) However, research is about understanding or finding answers to questions to a certain phenomenon. This require use of observation, (227) collection, analysing and interpreting data to provide answers. Both approaches involve these processes to give an interpretation of the findings. The only difference is the underlying theory of science which is the starting point for determining which methods to use. Often, quantitative methods come from a more logical positivist position, whereas qualitative methods arise from a more social constructivist perspective. A third view suggests that utilizing both approaches together can give a more comprehensive picture and a fuller and richer answer to the research question. (229) So, instead of viewing qualitative and quantitative (mixed) approaches as being in opposite ends, they represent an interactive continuum. (223)

The mixed methods approach takes various forms depending on the involvement of one method against the other in a study. Some authors have proposed eight different combinations of the mixed-methods matrix. (223) The matrix is drawn in terms of the level of usage of one method compared to the other. In this research, the mixed-methods undertaken involved collecting both qualitative and quantitative data and analysing these to better understand "HRQoL". To the most part, these were carried out separately as sub-studies within this thesis research.

3.5 Overview of psychometrics.

Psychometrics, introduced in chapter 2, is an essential part of both the development of new instruments and the cross-cultural adaptation of existing instruments (230). Psychometrics can be traced back to over 100 years with an origin in psychology. Precisely, it originated in the mid-19th century through work of psychophysicists – an experimental tradition with interest in quantifying differences between stimulus and response. (231) Psychometrics is made up of two words: 'psych' and 'metrics' from psychology which is the study of human behaviour and measurement respectively. It follows that psychometrics is concerned with the "measurement of subjective judgements using numerical scales and the evaluation of the measurement properties of scales". (231) The two main properties often used in connection with psychometrics are reliability and validity. (232) Reliability and validity are the means by which we know whether the judgements and decisions made by respondents are meaningful. (233)

There are currently two main paradigms for investigating these main psychometric properties: traditional psychometric methods (predominantly made up of classical test theory) and modern psychometrics (which includes both item response theory and Rasch measurement theory).

3.5.1.1 Classical Test Theory

Traditional psychometric methods have been in use for a long time and the most common of these is the Classical Test Theory (CTT). The CTT is known as 'classical' because as the name suggests its use can be traced back to beginning of 20th century to the work by Spearman. (234) In those early stages, rating scales were associated with the field of education and psychology; that's why the word 'test' is used. (235) Generally, CTT assumes a linear link between the observed- or test-score level and true score. (236) The premise for CTT is that the observed

score is the result of the true score plus measurement error, represented by the formulae:

Observed (test) score= true score + error score (measurement error)

CTT says that the random error around scores for a person given an infinite number of tests taken would be normally distributed. Therefore, the expected value of the error (i.e., the mean of the distribution of errors over infinite tests taken) is 0. (237) Standard error of measurement values varies depending on individual score, but scores developed using CTT all have the same standard error irrespective of where they lie in the distribution. The CTT is sample oriented. This means that for an instrument to be suitable in a setting other than where it was developed, it must be evaluated each time before being certified fit for use in the target population.

CTT has been instrumental to development of scales for more than a century. Despite this, there are a few problems with the assumptions held under CTT. Firstly, it is impossible to determine the values of true or error score from the principal assumption for CTT. After all, the true score is not just a function of respondents but also the instrument and other factors. (237) CTT therefore fails to show the mathematical relationship of the observed, true and error scores. This makes it a weak assumption which may lead to weak conclusions. Secondly, the fact that CTT scores have the same standard error irrespective of the distribution means that these scores are only valid at the group level and not at the individual patients/respondents. This is one of the arguments for moving towards modern psychometrics. Traditional psychometric methods, for which CTT is part, also assume that scores are sample dependent. This involve testing the scale in samples with different characteristics to ascertain that it works in different groups. (237) Further, traditional psychometric methods assume that the final score is scale dependent. This means that one will yield different outcomes for the same population when assessed using different scales that measure the same thing. A problem with this assumption is that people's functionality cannot be assessed independent of the scale or items.

However, functionality or level of judgement is the same irrespective of the measurement used. Additionally, at best a score developed using CTT will give ordinal level data rather than true interval. For these reasons and other assumptions, the psychometric world has seen emergence of modern methods.

3.5.1.2 Modern psychometrics: Item response theory and Rasch measurement theory

The item response theory (IRT) and Rasch measurement theory (RMT) are called modern because they have come after and intended to improve on CTT assumptions. IRT has its origin around the 1920s out of Thurstone's law of comparative judgement, (234) and later refined by Frederick Lord in the 1960s. (238) There is an important distinction between IRT and RMT in that the latter retains the importance of the construct. If the data don't fit the model, the RMT approach is to go back and interrogate the data and not the model. The IRT approach, on the other hand, would be to add parameters to help the fit. That is, if the data don't fit the model, the IRT approach is to continue adding/reducing parameters until you find a model that best fit the data.

3.5.1.3 Advantages and disadvantages of traditional (CTT) and modern psychometrics (IRT and RMT)

Scores developed using modern psychometrics are interval rather than ordinal. Further, standard error of measurement values are also generated for individual scores meaning that the scores can be used at the individual level. This one advantage of the modern psychometrics over the CTT which only compares scores at the group level.

The advantage of the CTT is that it is generally simpler to compute scores compared with modern psychometric models. This is because CTT is based on a simple linear relationship that true score is a sum of the observed score and measurement error. However, the disadvantage

with this is that the underlying theory and mathematical model could neither be tested nor refuted making it an easy assumption to satisfy. Another disadvantage with the CTT approach is that it assumes that the scores are both scale and sample specific. Table 3.2 contains a brief summary of the different characteristics of the CTT and modern psychometric models.

Table 3.2 Summary of the CTT and modern psychometrics

Attribute	CTT	Modern psychometrics
Record	Have long track record (from mid to late19th century) – classical	More modern (from around the 1920s) psychometric method
Model	Simple mathematical linear model	Tend to involve complex non-linear models
Total sum score	Total score is sum of individual score and measurement error	Score is the function of person ability and item difficulty
Measurement error	It is assumed to be constant across people	Individual standard errors of measurement can be generated
Sample invariance	Sample based i.e. very much dependent on group in which it is estimated	Data that fit the model generate scores that are said to be "invariant" (i.e., not dependent on the particular sample or the particular scale)
Sample size	Sample size are generally small, n=200-500	Depends on purpose of analysis but generally requires large sample size

Adapted from the following sources: (234, 236, 237, 239)

3.6 The importance of constructs and methods to develop and understand them in psychometrics

Constructs are the unobservable concepts that are being measured and in the context of this thesis, HRQoL. Whichever psychometric paradigm is used, a clear understanding and

description of the construct being measured is essential in order to be able to interpret, understand and apply the scores that are derived from a numerical scale. For example, in order to fully understand what it means to have a 10 point improvement in HRQL, it is first necessary to understand what HRQL is and how both low and high (and the gradations in between) HRQL are characterised.

As discussed in chapter 2, there are mainly two approaches employed in developing content for the HRQoL construct: bottom-up and top-down. Bottom-up approaches involve the target population in formulating constructs' domains or descriptive system. Top-down approaches generally involve researchers determining dimensions that an instrument should have with limited involvement of the target population Most generic childhood preference-based HRQoL instrument have been developed using the top-down approach. (107) Ideally these two approaches should be combined. Chapter 2 also showed that all generic childhood preference-based HRQoL instruments, except for the CHU-9D, were developed using a decomposed top-down approach. This was then followed by quantitative psychometric evaluation of reliability and validity. All this shows the importance of both the qualitative and quantitative skills in developing constructs as well as evaluating performance of the resulting instruments.

3.7 Why psychometrics needs a mixed approach

Psychometric principles provide standardised criteria for developing and evaluating an instrument's performance. Having standards to harmonize and gauge products and services is not uncommon. Manufacturers of goods and services have sets of standards, by accredited agencies, for which quality of items/ products are assessed. The USA Federal Drug Agency (FDA) is one such example. According to the USA FDA website, it states the following as one of its mission "The Food and Drug Administration is responsible for protecting the public

health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices; and by ensuring the safety of our nation's food supply, cosmetics, and products that emit radiation". (240) These standards ensure that goods /products are of high quality following a set standard. This is no different in health where patient reported outcome measures (PROMs) are used to measure health status, QoL or HRQoL. In health, psychometrics ensures that same standards apply for developing and evaluating instruments to maintain good quality and generalizability. With this in mind, the third of the four USA FDA missions states: "FDA is responsible for advancing the public health by helping to speed innovations that make medical products more effective, safer, and more affordable and by helping the public get the accurate, science-based information they need to use medical products and foods to maintain and improve their health". (240) The USA FDA guidelines is only one of the four guidelines identified in chapter 2 that outline psychometric properties and their criteria. The other three guidelines being: ISOQOL, MOT and COSMIN guidelines. (146, 147, 241) Psychometric principles therefore provide a set of formulas or statistics and criteria for subsequently making these judgements.

There is a lot of agreement and overlap of psychometric properties and evaluation criteria from the above four guidelines. The approach taken by this research is a consolidation of all psychometric properties from the four guidelines. This is not only a comprehensive approach but also ensures that all properties are considered which is fundamentally important when cross-culturally validating instruments. Further, the psychometric properties are both qualitative as well as quantitative in nature. This is because without robust qualitative work to establish the conceptual framework and understand the construct then it is impossible to undertake quantitative validity testing. Quantitative validity testing requires formulation of a priori hypotheses about how the construct is expected to relate to other similar/dissimilar

constructs.

Table 3.3 contains the 14 consolidated psychometric properties which are both qualitative and quantitative in nature. The qualitative properties include: conceptual framework, content validity and cross-cultural adaptation and practicality/burden; and quantitative: reliability (internal consistency, test-retest and inter-rater), validity (concurrent, predictive, convergent, discriminant, known-group and responsiveness), and interpretation of scores. Interpretation of scores pertain to instruction for scoring and interpreting minimal important difference. For this reason, interpretation of scores has been grouped together with other quantitative psychometric properties. Practicality/burden can be assessed qualitatively, perhaps by observing how participants complete a given instrument, but this is also done quantitatively e.g. by computing completion rate, response rate and completion time.

Table 3.3 shows the 14 consolidated psychometric properties

Psychometric property	Conceptual framework[1-4]	Reliability (dependability/ consistency of an instrument) [1-4]			Validity (accuracy/ represent exactly what is on the ground) [1-4]						Interpretation of scores[1-4]	Practicality/ Burden [1-4]	Cross-cultural adaptation: Conceptual & linguistic equivalence[1-4]
		Internal consistency [1-4]	test re-test (reproducibility)[1-4]	inter-rater [1, 2, 4]	Content [1-4]	Criterion[1-4]		Construct [1-4]			Responsiveness[1-4]			
						concurrent	predictive	Convergent	Discriminant	Known group				
Evaluation criteria from psychometric guidelines	• conceptual basis for item content • target population involved • info on dimension & distinctiveness of scales • intended level of measurement • documentation defining & describing concepts included & intended population • documentation of how concept(s) are organized	**Cronbach's alpha, α** - expected to be α:≥0.70 for group level comparisons ≥0.9 for individual comparisons	**usually assessed using Inter Class Correlation (ICC). At times assessed using kappa** • expected ICC, kappa, κ≥0.70 • kappa may be used for nominal scales (κ≥0.70) • *clear statement of the time interval is required*	**usually assessed using ICC (≥0.70)**: • Inter class correlation coefficient	no single criteria for this but generally that the questionnaire is comprehensive: • items represent all concepts in the conceptual framework • covers all health-related domains • include qualitative interview to evaluate target population understanding of all included items	It generally relies on the existence of a 'gold' standard instrument. Were the 'gold' standard available, there would be no need for new measures. Where this has been applied: • provide rationale and support for using this. • correlation between the two instruments should be ≥0.70 • for screening instruments, this can be compared between screening results against diagnosis		there should be a priori hypotheses among similar constructs and that unrelated constructs should diverge. Assessed by: • **convergent**: correlation with tools measuring same construct ≥0.50 • **discriminant**: no correlation between tools measuring different constructs		**compare scores** for groups known to clinically differ on relevant variables. Assessed by **Cohen's effect size**: • <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, >0.8 large	usually assessed by: • **standardized response means** (SRM) = mean diff / SD of score change • SRM>0	• logic and methods used • responder definition • Info on how data from measure be reported and displayed • representative mean(s) and std dev(s) in the reference population • guidance on minimally important diff in scores	Several ways to assess this and could be a mixture of: • instructions for each item • completion rate • completion time • literacy level • response rate	• process used to translate & culturally adapt instrument • linguistic-≥2 forward translation; ≥1 backward translation; expert and lay review; field test • details of translators • how inconsistencies (if any) were reconciled • translators worked independently

Blue shaded areas represent qualitative methods; Green shaded areas represent quantitative approaches

All psychometric properties of an instrument (new or old) need to be evaluated to establish if they are good enough. This is especially the case if the instrument is to be used in a setting other than where it was developed. The evidence from Table 2.8 in chapter 2**,** showed that developers of generic childhood preference-based HRQoL instrument have evaluated some but not all psychometric properties. This was similarly the case when the preference-based HRQoL instruments were cross-culturally adapted in a sSA setting. For these and the following additional reasons, it was necessary to carry out a mixed methods approach for Malawi.

3.7.1 Conceptual topics

The evaluation of concepts as part of the qualitative work/methods are a necessary and important part of this doctorate work. As established up to this point, the concepts in most generic childhood preference-based HRQoL instruments have been 'borrowed' from adult measures. This has involved either adapting or simply using the same adult concepts in the childhood measures on the assumption that these are applicable in children. Societies are predominantly embedded in the approach that children are a unit and as such adults do things for them. (242) This has led some researchers to coin the 'missing child' paradox. (243) It is important in research involving children, that they (children) should be viewed as equal partners or as 'experts' for their views and opinions in matters that affect them. Except for the CHU-9D, all generic childhood preference-based HRQoL instruments were developed from the adult counterpart. The wording was accordingly adapted to fit that of children, but concepts were the same. However, changing the language does not guarantee getting to the root of the conceptual issue. Establishing that these concepts are what the children themselves consider important is vital.

On the other hand, some researchers have argued for the use of the parents/caregivers who are

guardians as better placed to not only know their children's present welfare but also their long-term health and welfare. This may be true and should probably dominate the measurement of HRQoL among children. It is an approach that has guided utility valuation for preference-based HRQoL instruments. (244) However, the proxy question has drawn substantial debate on who is more appropriate respondent when assessing children's HRQoL. (245) While it may be the case that adult domains/dimensions may relate to children and adolescents, (159) and therefore get opinions from adults, it is also a known fact that children have different perspectives and goals. (246) The Convention of the Rights of the Child declares that 'states parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child' (247) Depending on age and comprehension, children should therefore have a say in matters especially those relating to their health.

The guiding principle of children and adult proxies is that there is agreement of responses. Empirical evidence from a systematic review on this subject has, however, shown that there is generally disagreements between children and proxies (caregivers/parents/guardians). (245) After all adults might measure HRQoL based on their own subjective assessment as opposed to child's own behaviour, experiences and goals. (248) The Wilson and Cleary model, the oldest and most cited HRQoL development model, suggest that values and preferences of an individual will affect their overall HRQoL. (249) In addition to this model, the International Classification of Functioning, Disability and Health (ICF) (250) also recognizes the impact of personal factors and the environment in perception of health by individuals. As such parents may report on children's "health status" but not specifically on children's HRQoL. (81) This implies that concepts and subsequent domains or dimension structures should ideally be developed directly from children themselves within specific contexts. (142) This is consistent

with previous authors (189) who have emphasized the need for involvement of children within the relevant cultural context. (191, 193)

Lastly, as discussed in chapter 2, generic childhood preference-based HRQoL instruments have not been developed in sSA region (189, 199), let alone Malawi. No review is available to inform the transferability of concepts from adults to children or between cultures in a sSA setting. This raises the need to investigate the conceptualisation of health from a cultural perspective as well as from the child's own point of view.

3.7.2 Linguistic concerns

It is time consuming and expensive to develop health instruments simultaneously in several languages. As a result, instruments are usually developed in one or two internationally spoken languages and then translated into other languages depending on need. In doing so there is need to establish that the language used in these instruments is understandable to the target population. This is particularly the case when using the instruments in language/s other than the one they were developed in. The importance of this cannot be overemphasized in the sSA region where several Bantu languages are spoken. There are at least twelve tribes in Malawi, as discussed in chapter 1**,** and each one of these speak their own language with Chichewa being the main one. The literature review in chapter 2 showed that there has been no peer reviewed cross-cultural validation of the childhood instruments except for the HUI3 and EQ-5D-Y. But even at that, all the steps necessary for linguistic and conceptual equivalence of the HUI3 were not followed. (199) Further, the EQ-5D-Y linguistic and conceptual equivalence and use in any of the Bantu languages spoken in sSA has been limited. It is therefore important to establish that the translated language used is not only understandable to the children but also that the meaning is the same as in the source language.

3.7.3 Education challenges in childhood

Given the educational situation for most children in Malawi, introduced in chapter 1, this presents a challenge for comprehension of self-completed instruments. Malawi is a signatory to the UN to promote advancement of education in pursuit of UN SDG 4.7- universal literacy and numeracy. (251) There are two major sectors of the education system in Malawi: formal and informal, both with unique challenges. For example, despite free primary education in the formal sector, not everyone is getting even basic education and uncompletion rates are also high. This brings in issues of competency to self-complete and understand a health questionnaire. The cross-cultural validation of an existing instrument into any of Malawian languages need to be evaluated with these aspects in mind.

Part of the role of psychometrics is in resolving these cross-cultural methodological challenges. The linguistic and conceptual issues are resolved using qualitative research methods. Through quantitative methods, adapted instruments performance in a new setting or population is assessed.

3.8 How the mixed methods approach was used in this thesis

This thesis research utilized both quantitative and qualitative methods to address the research question. Qualitative methods were used in the Malawi adaptation of EQ-5D Y to ensure linguistic and conceptual equivalence and to investigate wider content validity. This included cognitive interviews and FGDs, among children and adolescents. The cognitive interviews were carried out to assess local comprehension of concepts and item wording developed in a UK setting. Chapter 4 provides the findings of assessments of linguistic and conceptual equivalence from English into Chichewa of the selected instruments: the EQ-5D-Y, EQ-5D-

Y-5L, PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report. Further, there was need to evaluate the content validity of the EQ-5D-Y which has been described as a health status measure. This aimed to establish whether the EQ-5D-Y covers all aspects of health that children consider important in a Malawian setting. Generally, content validity is about the extent to which an instrument includes all relevant concept or dimensions that define that construct. The FDA guidelines mentioned above define content validity as "the extent to which the instrument measures the concept of interest" (145) Qualitative work with the target population of a particular instrument is an important way of establishing this. In this thesis, FGDs were used to get an understanding of HRQoL from the perspective of children and adolescents, and if this aligns itself to the definition in existing instruments. Chapter 7 contain findings for this part of the study.

Quantitative methods were used to evaluate the psychometric performance of the adapted Chichewa versions against standard criteria. The adult EQ-5D-3L was originally developed from an econometric perspective i.e., as a preference-based HRQoL instrument to generate utility values that could be used in economic evaluation of health interventions. The psychometrics that were later applied were based on CTT. For this reason, the EQ-5D-3L was restricted to five items (dimensions), inevitably creating a problem for content validity. This is a well-known trade off. It is likely that scales such as EQ-5D-3L that had a specific original intention as quick and easy to administer utility instruments would not fare well under the scrutiny of modern psychometric methods. One reason for this is because such utility instruments were never intended to achieve the aim that these modern psychometric methods test. The EQ-5D-Y five items were developed from the adult EQ-5D-3L and so it would likely show problems in IRT or RMT. It would only be ideal to use the same psychometric paradigm as the original psychometric evaluation for a like-to-like comparison.

For these reasons, it was appropriate to employ the CTT approach to psychometrically evaluate the Chichewa EQ-5D-Y version for this thesis. This would ensure evaluating the Chichewa EQ-5D-Y version using the same methods as previously applied for the adult EQ-5D psychometric work. In addition, the EQ-5D-Y has never been evaluated in a Malawian sample. It is therefore appropriate to use the same psychometric methods (based on CTT) in this new sample. It was imperative to establish how the translated Chichewa version perform among children and adolescents in this setting. The CTT psychometric assessment is presented in chapter 5 (for the PedsQL ™ 4.0 self-report) and chapter 6 (for the EQ-5D-Y and EQ-5D-Y-5L).

3.9 Psychometric properties and criteria for evaluating childhood health-related quality of life instruments and how they were evaluated in this thesis

In chapter 2.4, I outlined the psychometric properties and criteria that were used for evaluating instruments that were identified in the literature review. The importance of these psychometric properties have been outlined and justified above for evaluating the EQ-5D-Y and PedsQL ™ 4.0 self-report in this thesis research. The psychometric properties and criteria for evaluating childhood HRQoL as used in this thesis will now be discussed in detail below.

3.9.1 Conceptual framework

This is the first step in the development of an instrument. The framework provides detailed description of the conceptual basis for development of the construct to be measured. It is assumed that any instrument will be robust and have an existing good conceptual framework. It is for this reason that this is less detailed and evaluated when adapting an existing instrument as is the case when developing one. (149) Evaluation of the conceptual framework is however

important especially when adapting the instrument in a setting other than the one it was developed. The conceptual framework is usually assessed by ensuring that there should be conceptual basis for item content, dimension and distinctiveness of scales, intended level of measurement and evidence of defining and describing concepts. Importantly, there should be evidence of involvement of the target population, patient input and/or the general population in the development of an instrument's dimensions. (146, 147)

3.9.2 Conceptual and linguistic adaptation

This is the extent to which a scale shows linguistic and conceptual equivalence, and adequate measurement properties of the adapted version have been demonstrated. It is usually assessed by all of the following processes: forward translation, backward translation and cognitive interviews. Forward translation is from the scale source language to the target one and it is recommended that this be done by two independent translators. The forward translators are usually native speakers of the target language but are also conversant with the original source language. The backward translation process involves translating back what was translated in the target language into the original source language. Backward translators usually are native speakers of the source language but are also conversant with the target one. This process ensures that the concepts and their meaning are the same in the target as in the source language. There should be details of translators, how inconsistencies were reconciled and if translators worked independently. (146, 147, 151) A final step involves establishing conceptual equivalence through cognitive interviews. This involves administering the translated version on a sample of the target population and getting feedback about their understanding of the concepts and ideas.

3.9.3 Item performance

Psychometric analysis is usually conducted at the scale (several items) level rather than at the level of individual items. However, it is also useful to investigate the performance of individual items within each scale as this may help to explain the performance of the scales. For a scale to perform robustly, there is a need for each item to be contributing to it. In practice this means that participants need to be able to answer the item and hence there should be low levels of missing data. High levels of missing data may indicate that participants found the item difficult to understand, or felt that the content did not apply to them or that they were not prepared to disclose the information on a questionnaire. (237) Secondly, each level of the response scale needs to be fully utilized so that the distributions of responses for each item are relatively even. "Endorsement" indicates 'the proportion of participants who tick a particular item response. The pattern of endorsement may be influenced by several factors including illiteracy, comprehension, difficulties in concentration, or reporting biases such as 'faking good"'. (237) Faking good happens when participants give a false positive response and therefore does not represent how they are really feeling. Endorsement may also be affected by poorly worded response labels. Endorsement frequency investigates how any or some of the factors might have affected respondent's choice of answers to a questionnaire. Endorsement investigation involved both as maximum endorsement frequency and as adjacent aggregate endorsement frequency. (252) Thirdly, each item must have a unique contribution to the whole scale i.e. items are not over-lapped or redundant with each other. The criteria used to investigate each of these aspects at item analysis are outlined in detail below.

Missing data: A criterion of 5% was used to evaluate missing data. (231) Items that had ≥ 5% missing data were flagged as potentially problematic.

Maximum Endorsement Frequency (MEF): This is the endorsement at the extremes of the response scale. Items are considered problematic if they have >80% endorsement at either end of the scale. (237) This indicates that responses are concentrated in the top or bottom end of the scale (floor or ceiling effect).

Aggregate adjacent endorsement frequency (AEF): AEF is the extent to which adjacent response options in the middle of the response scale sum to at least a specified minimum. (253) Items are considered problematic if any two or more adjacent response options summed to <10%. (253) This informs whether there are some response options that are not being used.

Item redundancy: This reflects the extent to which each item within the scale makes a unique contribution and is not over-lapping with other items in the scale. It is assessed by evaluating the inter-item correlation between all items in the scale. In this study, items with inter-item correlations > 0.75 were considered problematic. (237)

3.9.4 Reliability

Reliability is the degree to which an instrument is free from random error to give consistent or dependable results each and over time. Reliability takes three forms: internal consistency, test-retest and inter-rater.

Internal consistency is the extent to which all the items in a scale are homogenous and reflect the same concept. For instance, a scale measuring social functioning might ask how one relates with friends and others in school/work/home, playing games, chatting and other social aspects of social life. It is expected that these items should be correlated to each other or homogenous

since they are measuring the same social domain. This correlation or homogeneity in a scale is usually assessed using Cronbach's alpha, α, or KR-20 for dichotomous scores. (146, 148)

Cronbach's α is calculated using the following formula:

$$\alpha = \frac{k}{k-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right)$$

where *k*= number of items, σ_X^2 = total score variance, and σ_i^2 = individual score variance. Cronbach's α is usually measured on a scale of 0 to +1. Higher values of α indicate greater homogeneity of an instrument. A reliable instrument will have α ≥0.70 for group level comparisons and ≥ 0.90 for individual comparisons. (146, 147)

Test-retest reliability is the extent to which a scale is stable over time, assuming no expected change in the interim. It is expected that assuming nothing else has changed, the score from a participant would remain the same over a short period of time. A clear statement of time interval for test-retest is required- usually assessed within a period of two weeks. Test-retest reliability is assessed using intraclass correlation coefficient (ICC) and kappa coefficient, ϰ. Both ICC and ϰ are measured on a scale of 0 to +1 with a higher value indicating stronger evidence of the instrument's reproducibility. The criteria for either ICC or ϰ of ≥ 0.70 may be used for nominal scales. (146, 148)

Inter-rater reliability describes the extent to which a scale produces similar results when the scale has been rated by different people. It is only applied to instruments that are rated by two or more raters, neither of whom are providing a self-report and the results compared. Inter-rater reliability is the degree to which the responses from the different raters agree. The criteria

for inter-rater reliability is ICC ≥ 0.70, where a high value indicates high levels of agreement between two assessors. (146, 149)

3.9.5 Validity

Validity refers to an instrument's ability to measure what it claims to measure. As an example, a scale that measures weight might be said to be reliable if it gives consistent readings each time. If that same measuring scale is poorly calibrated, the readings although reliable would be inaccurate (not valid). Validity is thus a way of ensuring that the instrument is not only reliable but that it measures what it purports to measure. This shows the importance of the psychometric property of validity for any good instrument. There are several types of validity, including content validity, construct validity (convergent and discriminant), item convergent and discriminant validity, criterion related validity, known-groups validity, empirical validity and responsiveness.

3.9.5.1 Content Validity

Content validity has been said to be probably the most important of all different forms of validity. (125) It is the extent to which the items in the scale represent the conceptual breadth of the construct. All the relevant concepts that make up the construct should be included in a scale. There is no statistical criterion for assessing content validity of an instrument. Generally, an instrument is expected to be comprehensive and that it should cover all relevant domains. This can be achieved by ensuring that at development stage, there is target population involvement in identifying and selecting items (and domains) of a given construct. (147, 149)

3.9.5.2 Construct Validity

Construct validity refers to the degree to which an instrument confirms an priori hypothesis

i.e., whether truthful to measuring what it claims to measure. It is usually assessed by measuring the instrument's convergent validity, discriminant validity and known-group validity.

3.9.5.2.1 Convergent Validity

Convergent validity is the extent to which similar dimensions of two or more instruments are related. Convergent validity is assessed by an a priori hypothesis usually using correlation coefficients. For convergent validity, a coefficient of ≥ 0.50 would indicate moderate to strong correlation. (148) It is expected that similar dimensions on different instruments would have moderate to strong correlation as evidence of convergent validity, the higher the better.

3.9.5.2.2 Discriminant Validity

Discriminant validity on the other hand measures the extent to which unrelated dimensions between scales should not be similar. As with convergent validity, discriminant validity is also assessed by an a priori hypothesis usually using correlation coefficients. Discriminant validity is assessed for lack of correlation between dimensions measuring different concepts which is usually <0.2. The smaller the coefficient the better the discriminant validity.

3.9.5.2.3 Item convergent and discriminant validity

Item convergent and discriminant validity is concerned with whether items are more associated with items of their hypothesized scale than with items of other scales. This was investigated using a method based on the Multitrait-multimethod (MTMM) matrix developed by Campbell and Fiske. (254) The MTMM is a matrix of correlations arranged by multiple traits (constructs) by multiple methods. However, correlations matrix in this study were generated using one (mono) method only (i.e., all data were reported by children), so the MTMM was modified to

a Multitrait-*mono*method. The ratio of correlation coefficient matrix for the same construct was divided by the correlation coefficient matrix of the two constructs. A ratio below 0.85 (conservative approach) the scale is said to have met the criteria for discriminant validity.(255)

3.9.5.3 Known-group Validity

Known groups validity is the capacity for the scale to detect differences between groups known to be clinically different (for example cases and controls). It is usually assessed by an a priori hypothesis that scores for clinically different groups would be different. The statistical significance of the difference between clinically different groups is usually assessed using p-value of <0.5 although this does not indicate the magnitude of the difference. The magnitude is better assessed using Cohen's effect size statistic calculated using the mean difference divided by standard deviation of the reference (healthy) group as follows:

$$\frac{mean\ of\ healthy\ children - mean\ of\ sick\ children}{standard\ deviation\ of\ healthy\ children}$$

where <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, >0.8 large. (150) The higher the effect size, the better the scale is at distinguishing differences between two or more clinically different groups.

3.9.6 Criterion Validity

Criterion-related validity is the extent to which scores on the new instrument are correlated with scores of the 'gold standard' that measures the same construct. This can be administered either concurrently (concurrent validity) or by an a priori model of prediction (predictive validity). Criterion-related validity may be evaluated by comparing a 'new' scale and an

existing validated gold standard tool measuring the same construct. A correlation ≥ 0.70 between the 'new' scale and an existing validated tool is used as a criterion. (151) Due to lack of 'gold standard' instruments, assessing criterion related validity for most scales is not realistically possible. Also, were there a 'gold standard' there would be no need for developing new measures.

3.9.7 Empirical validity

The EQ-5D is a preference-based health-related quality of life measures used for measuring HRQoL but also for use in cost-effectiveness analysis. As such, other than undergoing rigorous psychometric evaluations outlined above, the EQ-5D needs to demonstrate that it is able to detect changes in health status. (91, 154) The point health economists put forward is that preference-based measures do not really measure HRQL but that they measure the value (preference) placed on specific health states. (91) Therefore, if validity is about the extent to which the instrument measures what it purports to measure then it is important to evaluate how and to what extent the instrument measures preferences not HRQL. It was hypothesized that the utility scores generated by the EQ-5D-Y and EQ-5D-Y-5L would detect differences in external indicators of health status. Additionally, it was expected that the EQ-5D-Y-5L would possess greater empirical validity compared to the EQ-5D-Y. This hypothesis was based on the fact that the EQ-5D-Y-5L, like the adult EQ-5D-5L counterpart, was developed on the premise to address ceiling effect as well as sensitivity. (256) As such the EQ-5D-Y-5L is expected to have better characteristics compared to the EQ-5D-Y. The EQ-5D-Y and EQ-5D-Y-5L ability to assess differences in external indicators of health status was investigated in two ways.

3.9.8 Responsiveness

Responsiveness is another part of validity which has been referred to as part of longitudinal construct validity in some quarters. (105) Responsiveness is usually defined as the ability of the scale to show change in relation to a treatment or intervention of known efficacy. For example, a patient receiving treatment is expected to experience improvement of his or her health condition because of the treatment with a known efficacy. A good scale should have the ability to detect these health improvements when the person is administered a health instrument. Responsiveness is usually assessed by standardised response means (SRM) or effect size. There is no standardized criterion to measure responsiveness since this is very much dependent on the instrument, construct and the intervention/treatment that is causing the change. Generally, responsiveness statistics are evaluated against an a priori hypothesis, indicating the size of change that is expected. (147)

3.9.9 Interpretability of scores (used for minimal important difference)

This is the degree to which one can easily assign meanings to an instrument's quantitative scores. The interpretability of scores therefore refers to a set of instructions of how to use a scale and assign meaning to the values, and as such there is no single assessment for it. A scale is expected to at least have some, if not all, of the following: responder definition, information on how data from the instrument is reported and displayed, representative mean(s) and standard deviation(s) in the reference population, and guidance on minimally important differences in scores. (146, 147, 149, 151)

3.9.10 Practicalities versus work burden to administering a scale

Practicalities and work burden refer to the amount of work involved on those completing and/or how to administer the scale. There are several ways for assessing practicality of a scale

including questionnaire length, formatting, size of font (as it affects readability) and completion rate. Completion time, response rate, any indication of coercing for a particular response and resources used to administer the questionnaire are also used to assess the practicality aspect.

All quantitative data analysis were performed using IBM SPSS 26.0.0. for Mac (IBM Corp. Armonk, New York, USA). (257) Qualitative data were analysed using Microsoft office word.

In summary, this thesis purposely takes a mixed methods approach and uses the strengths of both qualitative and quantitative psychometric methods to address whether and to what extent the EQ-5D-Y is reliable and valid in the Malawi context. The psychometric properties and criteria evaluated used in this thesis are summarized in Table 3.4 below.

Table 3.4 EQ-5D-Y and EQ-5D-Y-5L psychometric properties investigated in this thesis

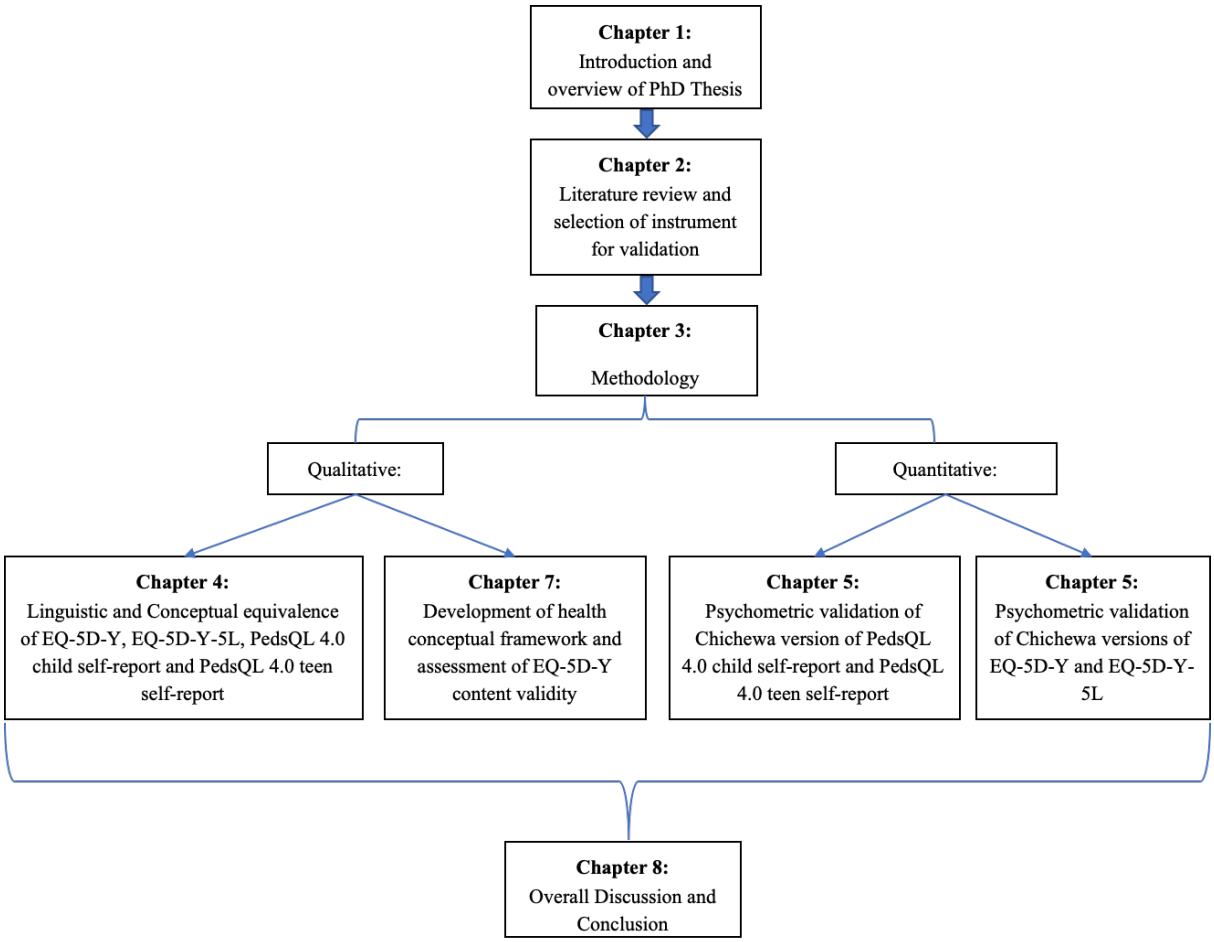
Source for psychometric properties: FDA, MOT, ISOQOL, COSMIN, Brazier et al (145-148, 154)

Main Psychometric property	Criteria	Justification for evaluation in this thesis	Where it should be investigated:
1. Conceptual framework			*Chapter 7*
2. Cross-cultural and language adaptation			
Linguistic equivalence			*Chapter 4*
Conceptual equivalence			*Chapter 4*
3. Item analysis			
Missing data			*Chapters 5 & 6*
Maximum endorsement frequency			*Chapter 5 & 6*
Aggregate adjustment endorsement frequency			*Chapter 5 & 6*
Item redundancy			*Chapter 5 & 6*
4. Reliability			
Internal consistency			*Chapter 5 & 6*
Test re-test			*Chapter 5 & 6*
Inter-rater			*Chapter 5 & 6*
5. Validity			
Convergent			*Chapter 5 & 6*
Discriminant			*Chapter 5 & 6*
Known-group			*Chapter 5 & 6*
Criterion			*Chapter 5 & 6*
Empirical			*Chapter 6*
Content			*Chapter 7*
6. Responsiveness			*Chapter 5 & 6*
7. Practicality/burden			
Missing data			*Chapter 5 & 6*
comprehension			*Chapter 4*
Completion time and rate			*Chapter 4*

Figure 3.1 below provides a flow of how the various psychometric properties within the mixed method approach as outlined in Table 3.4 above were evaluated in this thesis. The qualitative findings for investigating linguistic and conceptual equivalence are in chapter 4 and those that explored children's concepts of health in Malawi are in chapter 7. The quantitative psychometric methods enabled the use of well established, gold standard criteria to determine measurement properties of the EQ-5D-Y in Malawi (chapters 5 and 6).

3.10 Flow diagram of studies

Figure 3.1 Flow diagram of studies within this thesis.



3.11 Summary of Chapter 3

This chapter has provided justification for using the mixed methods approach to address whether and to what extent the EQ-5D-Y is appropriate for use in economic evaluation in Malawi. Cross-cultural validation of an existing instrument in a setting other than where it was developed requires a qualitative understanding of the construct being measured in the target language. It also requires a quantitative understanding of the psychometric performance of the instrument. This in turn is also dependent on the qualitative conceptual understanding in order to formulate appropriate hypotheses about validity.

Having provided this justification, the next chapter presents findings of the first part of the qualitative work. The chapter presents findings to establish linguistic and conceptual equivalence of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report.

4 Chapter 4- Linguistic and conceptual equivalence in the Chichewa language (Malawi) of selected scales.

With:
Ngwira, L.G.; Jelsma, J.; Maheswaran, H.; Derrett, S.; Kapakasa, F.; Petrou, S.; Niessen, L.; Smith, S
VIHRI 2021
1st EuroQol African meeting (Oral presentation)
1st EuroQol Early career researchers meeting (Oral presentation)

Derrett, S.; Herdman, M.; Ngwira, L.G.; Yohe Moore, E.; Jelsma, J.
The Patient 2021

4.1 Overview of chapter

In chapter 2, the EQ-5D-Y was selected for cross-cultural adaptation for Malawi. The PedsQL ™ 4.0 was selected as an appropriate validating measure and also needed to be cross culturally adapted. The EQ-5D-Y was developed in 2010 (159) and has demonstrated psychometric robustness in both healthy and sick populations around the globe. (160) The EQ-5D-Y has neither been cross-culturally adapted nor validated in Malawi. Additionally, the recently developed beta (trial) EQ-5D-Y-5L has only been cross-culturally adapted or psychometrically evaluated in Europe. As such, during cross-cultural adaptation of the EQ-5D-Y in Malawi, it was considered appropriate to do the same for the EQ-5D-Y-5L.

Cross-cultural adaptation requires application of the mixed methods approach as discussed in chapter 3. As part of psychometric evaluation, construct validity entails correlation between similar constructs measured by different instruments. The PedsQL ™ 4.0 child self-report (ages 8-12) and teen self-report (ages 13-18) were used for this purpose.

This chapter will summarise the cross-cultural adaptation – establishing linguistic and conceptual equivalence - of the EQ-5D-Y, EQ-5D-Y-5L and PedsQL ™ 4.0 from English into Chichewa.

4.2 Introduction

As introduced in chapter 2, the development and use of childhood preference-based HRQoL instruments has been on the increase since mid-1990s. (88) The popularity of preference-based HRQoL instruments has emerged from their use for resource allocation decisions. There are several childhood preference-based HRQoL measures, identified in chapter 2, that have been developed over the years. (122, 258) However, the use of these preference-based HRQoL instruments in LMICs (189) and sSA in particular has been very limited. (199) The EQ-5D-Y and PedsQL ™ 4.0 were identified as potential candidates for cross-cultural validation into Chichewa for Malawi. As discussed in chapter 3, cross-cultural validation involves adaptation and validation which require a mixed methods approach.

The EQ-5D-Y, developed by adapting the adult EQ-5D into a child friendly language version, has five dimensions (mobility, looking after oneself, usual activities, pain or discomfort, and worried, sad or unhappy) and the same three severity levels (no/not a problem, some problems, a lot of problems). (159) The major changes made were regarding the use of child friendly language to describe dimension of Mobility (walking about was added to the descriptor, as well as confined to bed option was changed to a lot of problems); self-care (included washing and dressing in the response options); usual activities (included child activities including playing and going to school); and anxiety/depression changed to worried, sad or unhappy.

The EQ-5D-Y was then psychometrically validated simultaneously in four European countries (Germany, Italy, Spain and Sweden) and South Africa. (160) In that study, the EQ-5D-Y was found to be feasible with little missing responses, was well understandable by respondents, with good test re-test reliability, convergent and known-group validity. However, these settings

are not in LMICs and although South Africa was included it is not representative of most countries in the sSA region. Also, the EQ-5D-Y was not translated into any of the local South African languages to ascertain if there were any issues with the conceptual and linguistic equivalence. This needs investigating and this chapter aimed to address this.

The review of literature also show that the EQ-5D-Y has been criticised for being insensitive to small changes in health status and for exhibiting ceiling effects. (259) In order to overcome this, the EuroQol Group recently developed simultaneously in four languages (German, Spanish, Swedish and English) a beta (trial or pre-official) version, the EQ-5D-Y-5L. The EQ-5D-Y-5L has the same five dimensions but with five severity levels per dimension: no/not; a little; some; a lot (or quite); and extreme problems (or extremely or cannot). (259) While the official EQ-5D-Y-5L version has not been released, the beta EQ-5D-Y-5L version has been tested in Cantonese amongst populations in Hong Kong using a standard EQ-5D-Y translation protocol. (260) The EQ-5D-Y has demonstrated psychometric robustness including cross-cultural adaptation across settings. (261) This is not the case with the EQ-5D-Y-5L owing to it being relatively new. The cross-cultural adaptation of the EQ-5D-Y for Malawi was therefore carried out alongside that of the beta EQ-5D-Y-5L version. Throughout this thesis the EQ-5D-Y-5L referred to is the beta version.

For the English (UK) EQ-5D-Y-5L to be appropriate for use in other settings, the translated version must be linguistically and conceptually equivalent, (193) and there should be no issues of severity level inversion. (262) Severity (or response) level inversion happens when a worse response option is ranked higher (or better) than a better one, e.g. ‘a lot of problems’ is ranked better than ‘having a little bit of problems’. A key requirement for HRQoL measures is that the response options should be discrete, and also respondents are able to rank them hierarchically in order of severity. (262) It is highly likely that severity level inversion would translate to

preference inversion in a utility evaluation exercise, and thereby impacting on the economic evaluation and policy derived from such preferences.

The relevance of childhood preference-based HRQoL instruments for Malawi were outlined in chapter 1. Briefly, the HSSP Ⅱ (2017-2022) medium-term health strategy contain EHP in pursuit of UHC. Future design of EHP will require appropriate instruments for carrying out cost-effectiveness, particularly CUA, for determining services to be included within the package. The adult EQ-5D is one of the mostly used measures and it is hoped that both the EQ-5D-Y and EQ-5D-Y-5L will be appropriate cross-culturally for economic evaluation of children and adolescents interventions. It is therefore imperative to establish whether the newly developed EQ-5D-Y-5L achieves conceptual equivalence in other settings and languages including those of sSA. Additionally, that it ensures any severity level inversion is resolved since this could have a direct impact on utility values derived from such a measure. Up to the time of doing this research, there was no Chichewa (Malawi) language version of the EQ-5D-Y and EQ-5D-Y-5L.

The Peds ™ 4.0 is not a preference-based measure but a generic HRQoL measure that has been used alongside the EQ-5D-Y to assess construct validity in a sSA setting. (181) The appropriateness of the PedsQL ™ 4.0 as a validating measure was discussed in chapter 2.8.5.

4.3 Aims and objectives

The aim of the study described in this chapter was to cross-culturally adapt the EQ-5D-Y and the EQ-5D-Y-5L into the Chichewa language. The general performance of the EQ-5D-Y-5L in relation to the EQ-5D-Y in terms of other psychometric properties including ceiling effect is in chapter 6. The work in this chapter which looks at the adaptation only was conducted in

formal collaboration with the Euroqol Group, using their EQ-5D-Y-5L protocol for the methods. In addition, the VMC reviewer reviewed each step of the adaptation process. The secondary objective was to cross-culturally adapt the PedsQL ™ 4.0 into Chichewa language. All the work presented in this chapter was either carried out or led by this researcher with supervision for the EQ-5D-Y and EQ-5D-Y-5L as per the translation protocol requirements.

4.4 Methods

4.4.1 Participants and recruitment

The study recruited both healthy and ill, children and adolescents and was undertaken in Blantyre, Malawi's commercial capital. In this study, children were defined as those aged ≤12 years, and those aged >12 years as adolescents. The healthy participants consisted of a convenience sample of children and adolescents attending primary and secondary schools in Blantyre. The sick children consisted of a convenience sample of children visiting the out-patient department at the Queen Elizabeth Central Hospital (QECH) in Blantyre. QECH is the largest referral and teaching hospital in Malawi. Written assent and consent were obtained from both children and parents/guardian, respectively, for each interview. (see Appendix 6-9) For sick participants, the invitation came at the end of clinical care. For healthy participants, invitations were through the schoolteacher. For both sets of participants, questionnaires were distributed by the research team at the end of clinical care or on arranged school day.

The study was approved by Ethic Committees of the Malawi College of Medicine (P.10/18/2509) and Liverpool School of Tropical Medicine (19-045). (see Appendix 10 and 11 respectively). Approval was also granted by the Blantyre district health office. (Appendix 12)

4.4.2 EQ-5D-Y and EQ-5D-Y-5L

4.4.2.1 The Instrument: EQ-5D-Y

The EQ-5D-Y (Figure 4.1) consists of a descriptive health system with five dimensions: 'mobility', 'looking after oneself', 'doing usual activities', 'having pain or discomfort', and 'feeling worried, sad or unhappy', (and visual analogue scale (EQ VAS)), each with three severity levels (or responses): 1 "no problems", 2 "some problems /a bit", and 3 "a lot of problems /very". Self-rated health is also assessed with an EQ VAS on a thermometer like vertical scale with scores ranging between 0 (representing worst imaginable health) and 100 (representing best imaginable health).

The five dimensions and three responses generate a total of 243 (3^5) unique health states. A single digit expresses the response level selected for that dimension such that the five-digit number for five dimensions describes specific health state. For example, the health state '11111' indicates no problems in all the five dimensions, '22312' indicates 'some problems walking about, some problems to wash or dress, a lot of problems doing usual activities, no pain or discomfort, and a bit worried, sad or unhappy', and '55555' indicates 'a lot/very' in each dimension. The EQ-5D-Y has a recall period of "today". (159)

Figure 4.1 The EQ-5D-Y English (UK) self-complete paper version

Health Questionnaire

English version for the UK

Under each heading, please tick the ONE box that best describes your health TODAY.

MOBILITY *(walking about)*

I have no problems walking about ❑
I have some problems walking about ❑
I have a lot of problems walking about ❑

LOOKING AFTER MYSELF

I have no problems washing or dressing myself ❑
I have some problems washing or dressing myself ❑
I have a lot of problems washing or dressing myself ❑

DOING USUAL ACTIVITIES *(for example, going to school, hobbies, sports, playing, doing things with family or friends)*

I have no problems doing my usual activities ❑
I have some problems doing my usual activities ❑
I have a lot of problems doing my usual activities ❑

HAVING PAIN OR DISCOMFORT

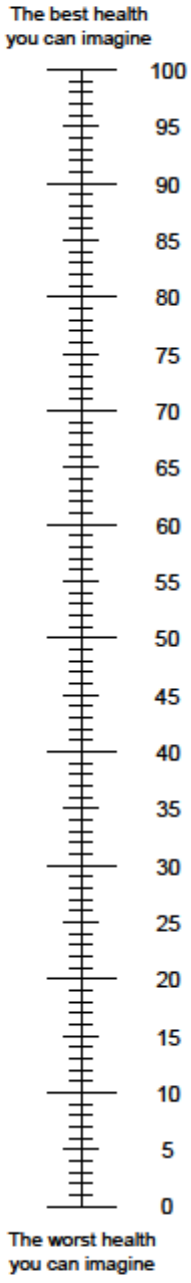
I have no pain or discomfort ❑
I have some pain or discomfort ❑
I have a lot of pain or discomfort ❑

FEELING WORRIED, SAD OR UNHAPPY

I am not worried, sad or unhappy ❑
I am a bit worried, sad or unhappy ❑
I am very worried, sad or unhappy ❑

- We would like to know how good or bad your health is TODAY.
- This line is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Please mark an X on the line that shows how your health is TODAY.
- Now, write the number you marked on the line in the box below.

YOUR HEALTH TODAY =



1

4.4.2.2 The Instrument: EQ-5D-Y-5L

The EQ-5D-Y-5L instrument consists of the same five dimensions as the EQ-5D-Y version: 'mobility', 'looking after oneself', 'doing usual activities', 'having pain or discomfort', and 'feeling worried, sad or unhappy' (and EQ VAS), each with five severity levels: 1 'no problems/not', 2 'a little bit of a problem', 3 'some problems /quite', 4 'a lot of problems/really' and 5 'extreme problems/extremely/cannot'. (Figure 4.2)

The five dimensions and five responses generate a total of 3,125 (5^5) unique health states. A single digit expresses the level selected for that dimension such that the five-digit number across the five dimensions describes the specific health state. For example, the health state '11111' indicates no problems in all the five dimensions, '52314' indicates 'cannot walk, a little bit of a problem to wash or dress, a lot of problems doing usual activities, no pain or discomfort, and really worried, sad or unhappy', and '55555' indicates extreme problems in each dimension.

Figure 4.2 The EQ-5D-Y-5L English (UK) self-complete paper version

Health Questionnaire

English version for the UK

Experimental version

you can imagine

Under each heading, please tick the ONE box that best describes your health TODAY.

MOBILITY *(walking about)*

I have no problems walking about	❑
I have a little bit of a problem walking about	❑
I have some problems walking about	❑
I have a lot of problems walking about	❑
I cannot walk about	❑

LOOKING AFTER MYSELF

I have no problems washing or dressing myself	❑
I have a little bit of a problem washing or dressing myself	❑
I have some problems washing or dressing myself	❑
I have a lot of problems washing or dressing myself	❑
I cannot wash or dress myself	❑

DOING USUAL ACTIVITIES *(for example, going to school, hobbies, sports, playing, doing things with family or friends)*

I have no problems doing my usual activities	❑
I have a little bit of a problem doing my usual activities	❑
I have some problems doing my usual activities	❑
I have a lot of problems doing my usual activities	❑
I cannot do my usual activities	❑

HAVING PAIN OR DISCOMFORT

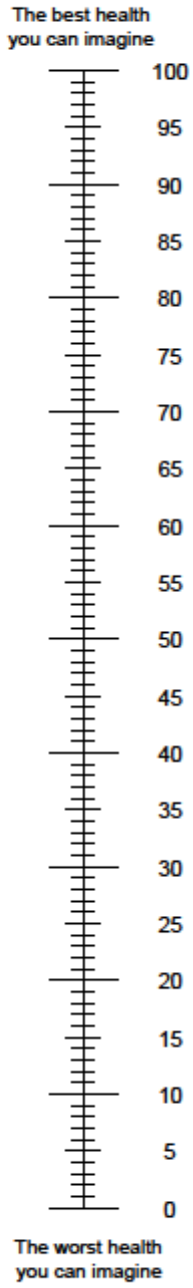
I have no pain or discomfort	❑
I have a little bit of pain or discomfort	❑
I have some pain or discomfort	❑
I have a lot of pain or discomfort	❑
I have extreme pain or discomfort	❑

FEELING WORRIED, SAD OR UNHAPPY

I am not worried, sad or unhappy	❑
I am a little bit worried, sad or unhappy	❑
I am quite worried, sad or unhappy	❑
I am really worried, sad or unhappy	❑
I am extremely worried, sad or unhappy	❑

- We would like to know how good or bad your health is TODAY.
- This line is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Please mark an X on the line that shows how your health is TODAY.
- Now, write the number you marked on the line in the box below.

YOUR HEALTH TODAY = [ ]



4.4.2.3 The EQ-5D-Y and EQ-5D-Y-5L cross-cultural adaptation procedure

The procedure consisted of two phases; translation and cognitive interviews for both the EQ-5D-Y and EQ-5D-Y-5L. The EQ-5D-Y-5L adaptation also included a third phase; a card ranking exercise. In addition, there is a slight text difference of the EQ-5D-Y and EQ-5D-Y-5L. For these reasons, the adaptation process is discussed differently for the two instruments.

The EuroQol Group provided the following for the cross-cultural adaptation process of the EQ-5D-Y from English into Chichewa: the EQ-5D-Y English (UK) version, the adult EQ-5D-3L Chichewa version and the EuroQol translation protocol (Figure 4.3). (263) The adult EQ-5D-3L Chichewa version has been validated for orthopaedic patients in Malawi. (264) The only difference between the EQ-5D-Y and EQ-5D-3L is the age-appropriate languages.

The EQ-5D-Y-5L cross-cultural adaptation process followed that of the EQ-5D-Y. So, other than the EQ-5D-Y English (UK) source document, the EuroQol Group provided both the approved Chichewa EQ-5D-Y and adult Chichewa EQ-5D-3L as reference material for the EQ-5D-Y-5L adaptation. The EuroQol Group also provided the draft EQ-5D-Y-5L translation protocol (Figure 4.4). This is referred to as a draft translation protocol since it had not officially been approved at the time of this translation work.

4.4.2.4 Capabilities of the translation team

The forward translation team was made up of two experienced translators within the Malawi-Liverpool-Wellcome Trust Clinical research (MLW) programme. The MLW programme has been involved in clinical research since mid 1990s and usually has a strong team of experienced translators. The adult EQ-5D-3L Chichewa version was provided to them as a guide only, as per EQ-5D translation guidelines, in composing the EQ-5D-Y version. The backward translation team were original English speakers but also experienced linguist in Chichewa

language. The English version was not provided to backward translators as per EQ-5D translation guidelines.

Figure 4.3 Translation process for the EQ-5D-Y adapted from the EuroQol translation process

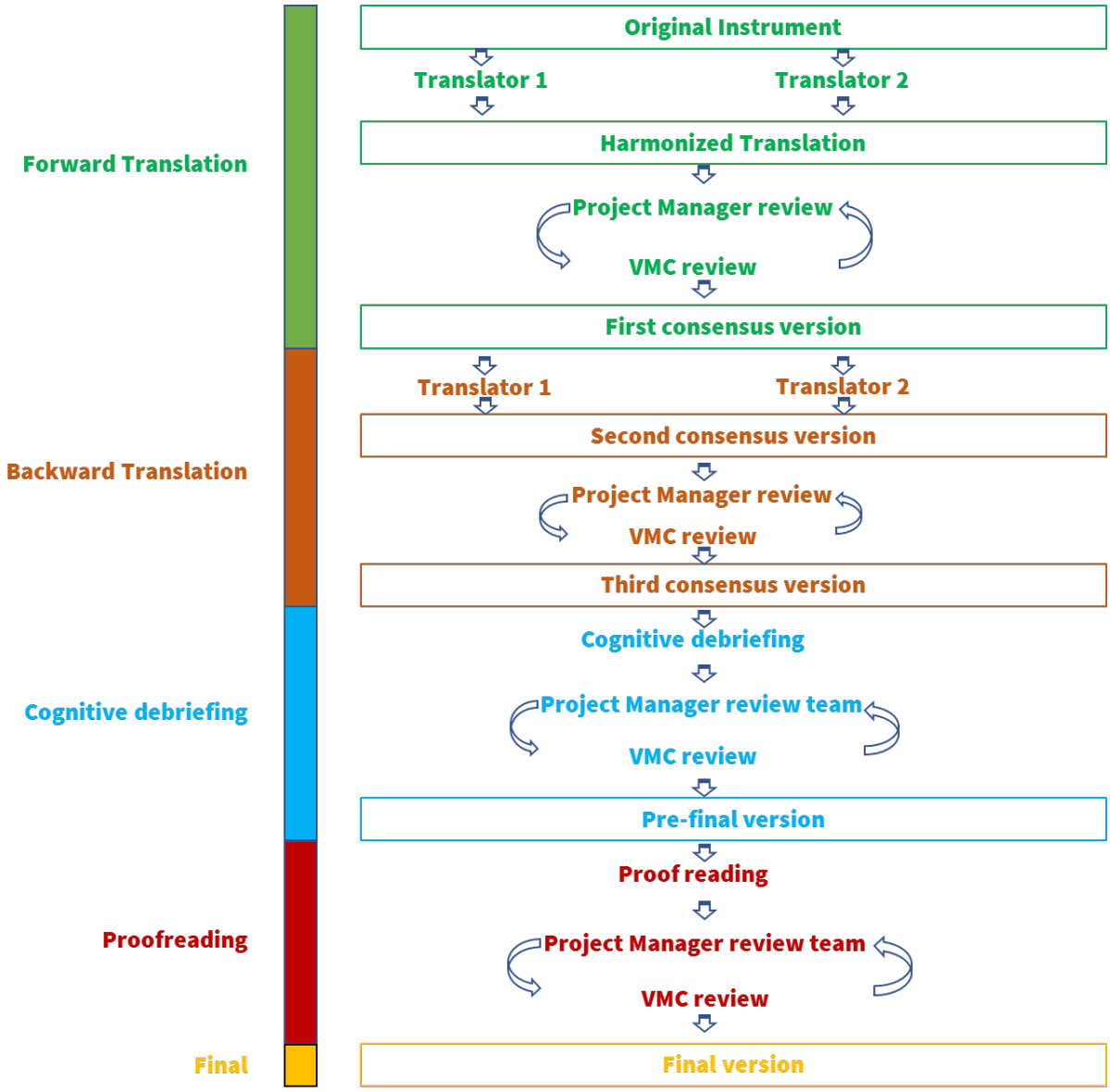
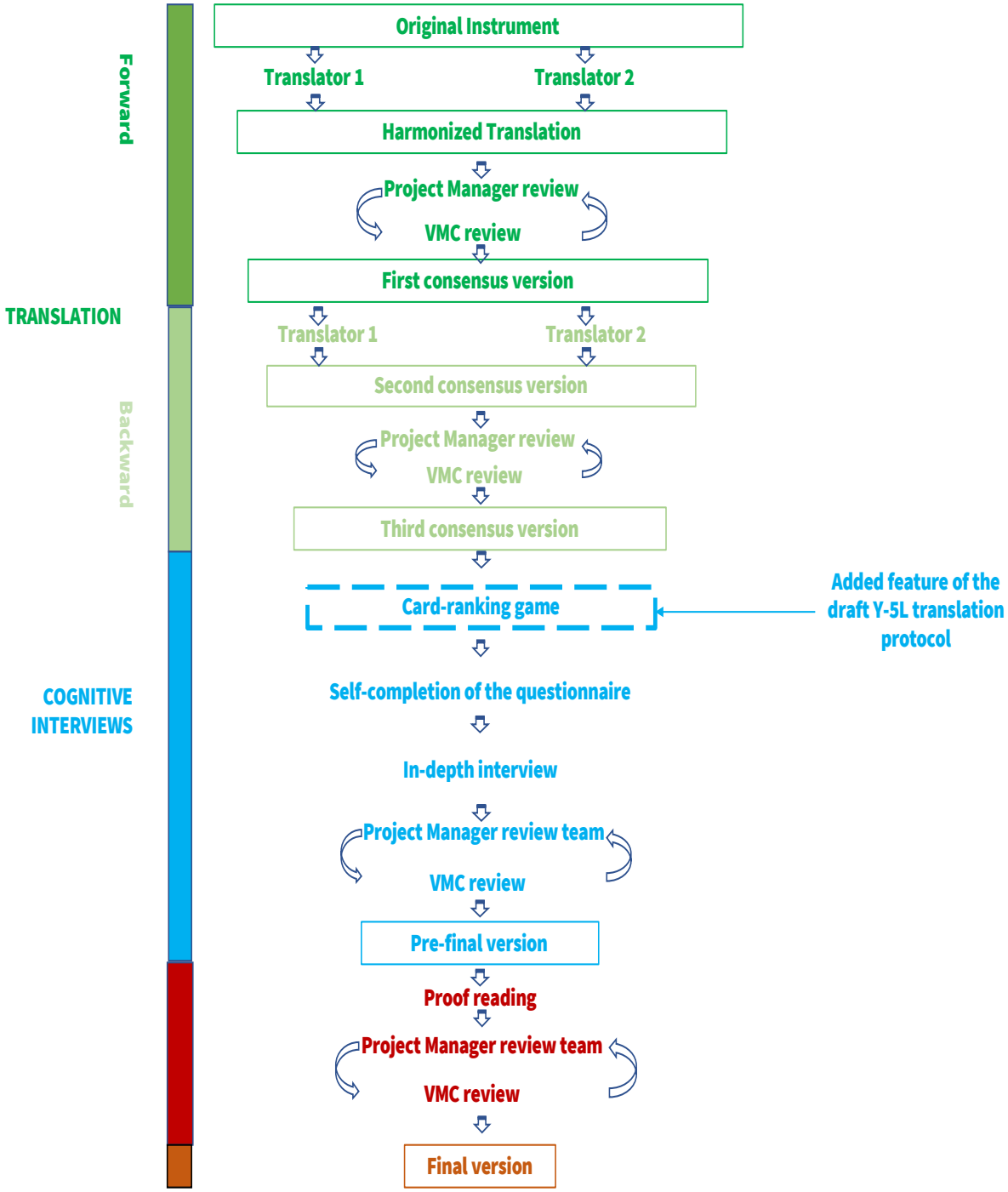


Figure 4.4 Translation process for the EQ-5D-Y-5L adapted from EuroQol translation process



4.4.2.5 Translation process

i) EQ-5D-Y

The translation for the EQ-5D-Y was done between October 2018 and July 2019. First, two experienced local translators undertook forward translation of the source English document into Chichewa. The translators sought for consistency between the EQ-5D-Y and EQ-5D-3L Chichewa version as far as this was appropriate. The translators worked independently of each other guided by the developers recommended approaches. The first consensus version was established after a meeting between the two direct translators and this researcher who is also fluent in Chichewa. Any differences in terminology used and inconsistencies were resolved at this stage to come up with the first consensus version. The first consensus version was then forwarded to the VMC reviewer including a full report of the translations by the two forward translators. Any inconsistencies and changes made to some text was discussed in an iterative way, justified, and agreed upon with the VMC reviewer.

After approval by the VMC reviewer, the first consensus version was then back translated from Chichewa into English by two native English speakers, different to the previous two translators and conversant with Chichewa, who did so independently of the source document. The backward translations were compared between the two backward translators and this researcher. Any inconsistencies and differences in choice of words used by the two translators was discussed and resolved. Changes that were deemed necessary from the backward translation were incorporated into the first consensus version to come up with the second consensus version. The full report of the backward translation and second consensus version were together sent to VMC reviewer. The VMC reviewer and this researcher discussed in an iterative manner, any inconsistencies and discrepancies in the second consensus. For any changes made to the second consensus version, justification was provided to produce the third

consensus version which was then piloted through cognitive interviews.

ii) EQ-5D-Y-5L

The EQ-5D-Y-5L translation was done between March and December 2019 and is the first such study to have used the draft translation protocol which includes the ranking exercise. The same forward translators for the EQ-5D-Y independently translated the EQ-5D-Y-5L from English to Chichewa. The backward translation to English was done by another set of two translators different to those who did the EQ-5D-Y. The translation involved the additional severity levels and any new text for the EQ-5D-Y-5L that are not in the EQ-5D-Y to produce a consensus version. The EQ-5D-Y-5L severity level descriptions were discussed by translators and this researcher to reach a consensus before forwarding to the VMC for review and ultimate approval. The translated questionnaire was then piloted through cognitive interviews.

4.4.2.6 Cognitive interview process

i) EQ-5D-Y

The participants took part in the following two cognitive interview steps: i) *self-completion of the EQ-5D-Y questionnaire*, and *ii) cognitive interviews*.

ii) EQ-5D-Y -5L

The EQ-5D-Y-5L cognitive interviews was more involving compared to the EQ-5D-Y as it included an additional card ranking exercise which was conducted first. The participants therefore took part in the following three cognitive interview steps: a) *card ranking exercise (a new translation protocol feature) to evaluate comprehension of the EQ-5D-Y-5L severity level;* b) *self-completion of the EQ-5D-Y-5L questionnaire*, and c) *cognitive interviews*. The card ranking exercise were conducted in four rounds made up of different participants each time. The first round of the interviews was conducted with four healthy and three sick children aged

8-15 years according to the VMC translation protocol. However, in order to resolve any apparent severity level inversion, a further three subsequent cognitive interviews were conducted with healthy children alone; it was not considered likely that severity inversion would relate to the health status of participants. The first ten participants took part in all the three steps of cognitive interviews. The rest (n=8) of the participants took part in the card ranking exercise only because it was seen that no new information was emerging from the last two completed cognitive interviews.

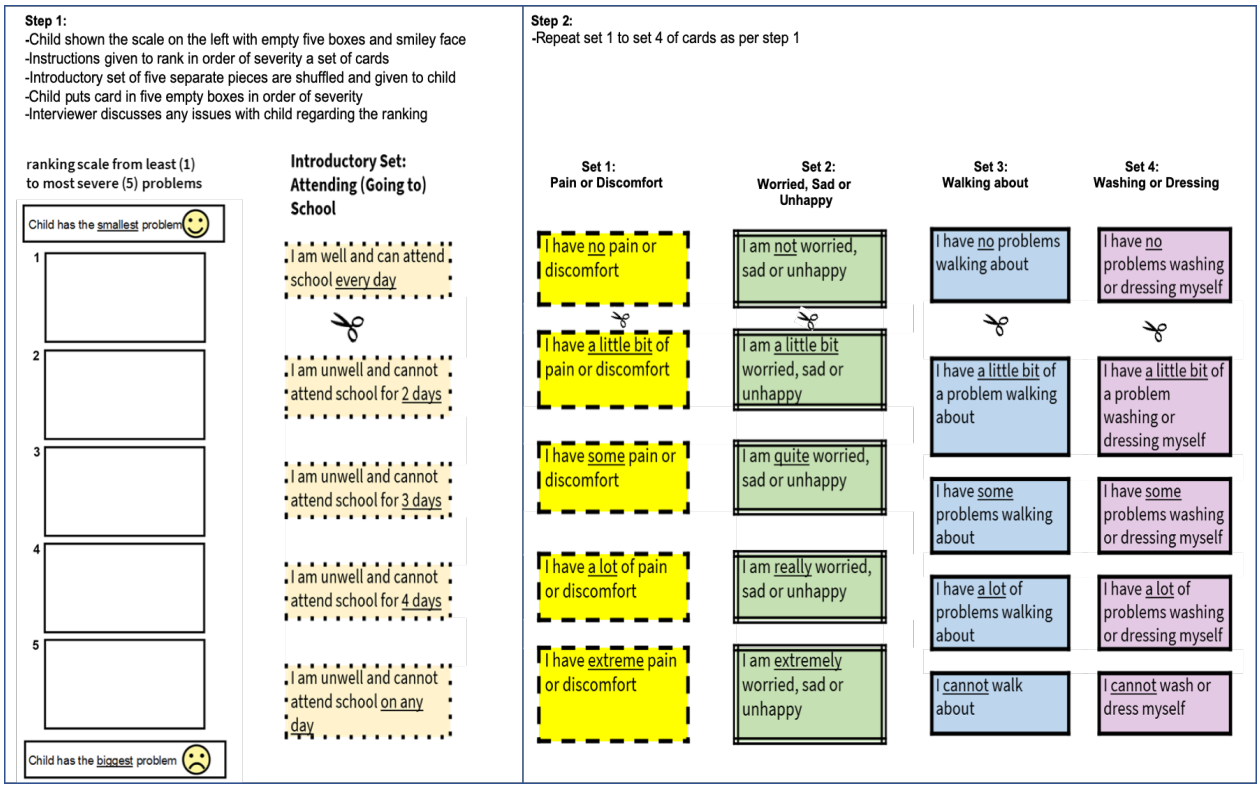
A data collection form was developed to record a few demographics including age, gender, grade, venue of interview and time taken to complete each of the questionnaires. The form did not contain any subject names. Instead, a number was allocated to each respondent. Additionally, probe questions to assess how the interviewee understood the questionnaire were included in the case report form as per attachment in the Appendix 13.

4.4.2.6.1 Card ranking exercise (EQ-5D-Y-5L only)

The card ranking exercise (Figure 4.5) was specifically developed by the VMC to detect severity level inversion in the beta EQ-5D-Y-5L so that this could be investigated and corrected during the adaptation process. Being the first such study to have employed the card ranking exercise, this doctorate research contributed to the evaluation of this added feature which has since been published. (265) The ranking exercise allows assessment of severity level inversion during the translation process by asking children to rank EQ-5D-Y-5L severity levels from least to more severe problems. All the five severity levels have the same qualifiers for ‘mobility’, ‘looking after oneself’, and ‘doing usual activities’. On the other hand, the ‘having pain or discomfort’ dimension has the same qualifiers as the first three dimensions except for the fifth severity level. The ‘feeling worried, sad or unhappy’ dimension also has unique qualifiers for the third to fifth severity levels. Due to similarity of qualifiers for all the severity

levels for the first three dimensions (mobility, doing usual activities, and looking after oneself), the VMC did not keep the 'doing usual activities' dimension as part of the card ranking exercise. (265) As such, the five ranked card sets consists of an introductory (warm-up) set, and four of the five EQ-5D-Y-5L dimensions.

Figure 4.5 The card ranking exercise in four of the five EQ-5D-Y-5L dimensions



Source: EuroQol Group

At the start of the interviews, participants first ranked the *introductory set* and any issues with understanding the card ranking process were clarified before moving on to the other four sets. Interviewers, with prior interview guide, were interested in observing whether the intended card order of the translated Chichewa cards were in the order anticipated (from no problems to extreme problems). The individual severity level ordering by each child and any observed difficulties to do so were recorded on a data sheet by the interviewer, which was later

summarized into a chart for all participants. The correct rank orderings were given an orange colour whereas a blue colour represented incorrect rankings. The blue squares gave an indication of where the problems were that needed addressing for the next iteration. This was done until the severity level inversion was resolved.

4.4.2.6.2 Self-completion of the questionnaire

i) EQ-5D-Y

Individually, the participants completed the Chichewa EQ-5D-Y questionnaire. Instructions were read out loud by the interviewer on how to fill the questionnaire. The children then self-completed the questionnaire, without assistance from either the interviewer or guardian/parent.

ii) EQ-5D-Y-5L

Following the card ranking exercise, participants different from those for the EQ-5D-Y, completed the Chichewa EQ-5D-Y-5L questionnaire. By design, the Chichewa EQ-5D-Y-5L was administered to the children *after* the card ranking exercise. This was done to avoid any possible biases that may have emerged from children seeing the intended ordering of the severity levels when self-completing the EQ-5D-Y-5L. If necessary, children who struggled to understand the instructions for self-completion (n=7) of the questionnaire were given help to understand the task. However, the children were not given help to respond to the questionnaire. Any difficulties that the children experienced when self-completing the questionnaire were noted by the interviewer. Additionally, any assistance given to help them understand the instructions was also recorded on a separate sheet of paper.

4.4.2.6.3 cognitive interviews about the questionnaire

i) EQ-5D-Y

After completing the questionnaire, each participant took part in a one to one interview about his/her understanding of the questionnaire. The interview questions were open ended and were adapted from the EQ-5D-Y translation protocol. Participants were first asked to comment on a) the comprehensibility of the whole questionnaire, b) words they found difficult to understand, and c) what they could change/add to make the instruction easy to understand. Participants were also asked to comment on their understanding of health, and best/worst health they could imagine among others. Further, the participants were also asked to reflect on how they understood each question and severity level. Specifically, the participants were asked their understanding of key qualifiers: 'some', 'a lot of ', 'very', and concepts like 'pain', 'discomfort' and 'looking after oneself'. The participants were asked their comprehension by giving specific examples of someone they could imagine to be in a particular health condition.

All the responses recorded on the EQ-5D-Y questionnaire were compared in terms of health state generated by such responses against the EQ VAS. At the end of the interviews the children were thanked for their participation and were compensated with an exercise book and pen/pencil for their time.

ii) EQ-5D-Y-5L

Each child participating in the first two rounds of interviews then took part in a one-to-one interview about his/her understanding of the questionnaire. The interview questions were open ended and were adapted from the draft EQ-5D-Y-5L translation protocol. (263) Participants were first asked to comment on: a) the comprehensibility of the whole questionnaire, b) words they found difficult to understand (in the dimensions, severity levels and instructions), and c) what they would change/add to make the instructions easy to understand. Additionally, the

participants were asked to reflect on how they understood: i) each question and ii) severity levels and, subsequently, iii) provide real or hypothetical examples of people experiencing specific severity levels. Other than comprehension of concepts, instructions and wording of the questionnaire, participants were not asked regarding the survey in general. At the end of the interviews the same procedure as for EQ-5D-Y explained above for compensating participants was followed.

4.4.3 PedsQL 4.0

4.4.3.1 The instrument:

The PedsQL ™ 4.0 child self-report and the PedsQL ™ 4.0 teen self-report were provided by the Mapi Research Trust (Mapi Trust) who hold the distribution rights. Both the PedsQL ™ 4.0 child self-report and the PedsQL ™ 4.0 teen self-report have 23 items in five dimensions: **'**physical functioning', 'emotional functioning', 'social functioning', and 'school functioning'; each with five severity level options: 'never', 'almost never', 'sometimes', 'often' and 'almost always.' The response completion time is in the past one month. The only difference between these two versions is with reference to use of the terms 'kids' and 'teens' respectively on some items.

For each of the 23 items, participants indicate how much of a problem it has been during the past one month on a 5-point severity levels. The severity level for each item is coded as: 0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem. Item responses are then reverse scored and linearly transformed on to 0-100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0). The PedsQL™ 4.0 total scale score is obtained by scoring across all 23 items (higher = better) and the Physical Functioning sub-scale (sum of eight Physical Health items), whereas the last three sub-scales (15 items) are

combined to form Psychosocial Health summary score. The sub-scale scores are obtained through the summation of the all the items that are answered. By summing across all the items over the number of items that are answered, it takes into account the missing items if present. (266, 267) Figures 4.6 and 4.7 has the PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report respectively.

Figure 4.6 Pediatric Quality of Life Inventory ™ Version 4.0 child self-report

ID#_____

Date:_____

PedsQL™

Pediatric Quality of Life Inventory

Version 4.0

CHILD REPORT (ages **8-12**)

DIRECTIONS

On the following page is a list of things that might be a problem for you. Please tell us **how much of a problem** each one has been for you during the **past ONE month** by circling:

0 if it is **never** a problem
1 if it is **almost never** a problem
2 if it is **sometimes** a problem
3 if it is **often** a problem
4 if it is **almost always** a problem

There are no right or wrong answers.
If you do not understand a question, please ask for help.

*In the past **ONE month,** how much of a **problem** has this been for you …*

ABOUT MY HEALTH AND ACTIVITIES ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. It is hard for me to walk more than one block	0	1	2	3	4
2. It is hard for me to run	0	1	2	3	4
3. It is hard for me to do sports activity or exercise	0	1	2	3	4
4. It is hard for me to lift something heavy	0	1	2	3	4
5. It is hard for me to take a bath or shower by myself	0	1	2	3	4
6. It is hard for me to do chores around the house	0	1	2	3	4
7. I hurt or ache	0	1	2	3	4
8. I have low energy	0	1	2	3	4

ABOUT MY FEELINGS ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. I feel afraid or scared	0	1	2	3	4
2. I feel sad or blue	0	1	2	3	4
3. I feel angry	0	1	2	3	4
4. I have trouble sleeping	0	1	2	3	4
5. I worry about what will happen to me	0	1	2	3	4

HOW I GET ALONG WITH OTHERS ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. I have trouble getting along with other kids	0	1	2	3	4
2. Other kids do not want to be my friend	0	1	2	3	4
3. Other kids tease me	0	1	2	3	4
4. I cannot do things that other kids of my age can do	0	1	2	3	4
5. It is hard to keep up when I play with other kids	0	1	2	3	4

ABOUT SCHOOL ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. It is hard to pay attention in class	0	1	2	3	4
2. I forget things	0	1	2	3	4
3. I have trouble keeping up with my schoolwork	0	1	2	3	4
4. I miss school because of not feeling well	0	1	2	3	4
5. I miss school to go to the doctor or hospital	0	1	2	3	4

Figure 4.7 Pediatric Quality of Life Inventory ™ Version 4.0 teen report

ID#_____

Date:_____

PedsQL™

Pediatric Quality of Life Inventory

Version 4.0

TEEN REPORT (ages **13-18**)

DIRECTIONS

On the following page is a list of things that might be a problem for you. Please tell us how much of a problem each one has been for you during the **past ONE month** by circling:

0 if it is **never** a problem
1 if it is **almost never** a problem
2 if it is **sometimes** a problem
3 if it is **often** a problem
4 if it is **almost always** a problem

There are no right or wrong answers.
If you do not understand a question, please ask for help.

*In the past **ONE month,** how much of a **problem** has this been for you …*

ABOUT MY HEALTH AND ACTIVITIES ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. It is hard for me to walk more than one block	0	1	2	3	4
2. It is hard for me to run	0	1	2	3	4
3. It is hard for me to do sports activity or exercise	0	1	2	3	4
4. It is hard for me to lift something heavy	0	1	2	3	4
5. It is hard for me to take a bath or shower by myself	0	1	2	3	4
6. It is hard for me to do chores around the house	0	1	2	3	4
7. I hurt or ache	0	1	2	3	4
8. I have low energy	0	1	2	3	4

ABOUT MY FEELINGS ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. I feel afraid or scared	0	1	2	3	4
2. I feel sad or blue	0	1	2	3	4
3. I feel angry	0	1	2	3	4
4. I have trouble sleeping	0	1	2	3	4
5. I worry about what will happen to me	0	1	2	3	4

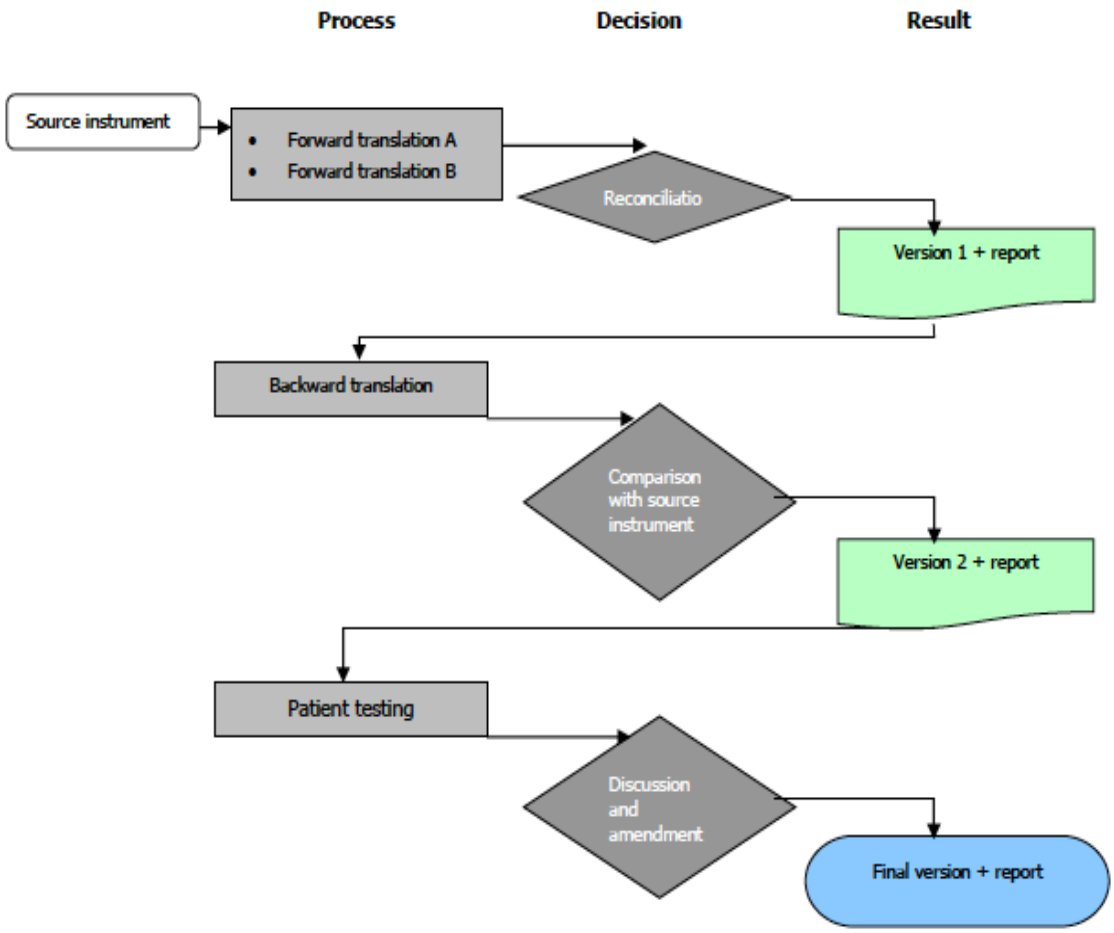
HOW I GET ALONG WITH OTHERS ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. I have trouble getting along with other teens	0	1	2	3	4
2. Other teens do not want to be my friends	0	1	2	3	4
3. Other teens tease me	0	1	2	3	4
4. I cannot do things that other teens of my age can do	0	1	2	3	4
5. It is hard to keep up with my peers	0	1	2	3	4

ABOUT SCHOOL ***(problems with…)***	**Never**	**Almost Never**	**Some-times**	**Often**	**Almost Always**
1. It is hard to pay attention in class	0	1	2	3	4
2. I forget things	0	1	2	3	4
3. I have trouble keeping up with my schoolwork	0	1	2	3	4
4. I miss school because of not feeling well	0	1	2	3	4
5. I miss school to go to the doctor or hospital	0	1	2	3	4

4.4.3.2 Cross-cultural adaptation procedure:

Translation procedure for PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report followed the translation protocol supplied by the Mapi trust (translation algorithm shown in Figure 4.8**)**. The PedsQL ™ 4.0 translation protocol is similar to the one for the EQ-5D-Y explained above. The only slight difference is that, unlike the EQ-5D-Y (and EQ-5D-Y-5L), the translation process was not overseen by developer's reviewer since they did not assign one. However, each step of the adaptation process was first reported to the Mapi trust office before proceeding to the next step. The adaptation process took place between March 2019 and December 2019.

Figure 4.8. PedsQL 4.0 Linguistic validation process algorithm (copied from PedsQL translation protocol)



4.4.3.3 Translation process

The same procedure for forward and backward translation as reported for both the EQ-5D-Y was followed for the PedsQL ™ 4.0 from English (US) into Chichewa and back from Chichewa into English. Through a phone meeting between the two backward translators and this researcher, a backward translation version was reconciled and approved for pilot testing through cognitive interviews.

4.4.3.4 Cognitive interview process

Five children and adolescents aged 8-12 years and 13-18 years self-completed and later interviewed for PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report respectively. The recommended number of cognitive interviews as per protocol is five sick children aged 8 to 12 years for PedsQL ™ 4.0 Child self-report, and five adolescents aged 13 to 18years for the PedsQL ™ 4.0 Teen self-report. However, due to logistical challenges to enrol sick participants at the time, Mapi trust agreed piloting of the measure among the healthy population only. Each of the ten participants (five each) took part in the two cognitive interview steps: i) self-completion of the PedsQL ™ 4.0 questionnaires, and ii) in-depth interviews for PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report.

4.4.3.4.1 Self-completion of the PedsQL 4.0 questionnaire

The translated questionnaires were administered to the participants for self-completion. As with the other instruments above, help to understand the instructions were given but not how to respond to the questions. Any support given was similarly recorded by the interviewer on a separate sheet of paper.

4.4.3.4.2 Cognitive interviews about the PedsQL 4.0 questionnaire

Each participant then took part in a one-to-one interview about their understanding and other aspects of the questionnaire as for the EQ-5D-Y and EQ-5D-Y-5L. Two aspects were achieved by the testing: conceptual equivalence of items and the different response ratings, and participants' understanding of the instructions and items. Specifically, participants were asked to comment on: i) any difficulty in their understanding of the items in the questionnaire, and ii) how they interpreted all items. Additionally, participants were also given an opportunity to propose alternatives where there were some ambiguities in the translated version.

The same procedure for compensating participants as explained above was followed.

4.4.4 Preparation of the final approval

The translated wordings for EQ-5D-Y and EQ-5D-Y-5L were revised following each iteration of cognitive interviews until all inconsistencies were resolved. These versions were proofread for typographical and layout errors. The pre-final versions were forwarded to the VMC and EuroQol Office for endorsement as the final translated Chichewa (Malawi) EQ-5D-Y and beta EQ-5D-Y-5L version.

A similar step was undertaken for the PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report. The pre-final versions were forwarded to Mapi trust for endorsement as the approved versions of the Chichewa (Malawi) PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report.

4.5 Results

4.5.1 EQ-5D-Y

4.5.1.1 Participants and recruitment:

The sample for the EQ-5D-Y cognitive interviews were made up of eight children (female=2, male=6). However, since only one of the eight interviewed participants were sick, the developers recommended recruiting an additional sick participant. In total, therefore, there were nine participants (2=female, 7=male) aged between 8-15years (mean =9.6yrs, median =8yrs) that were interviewed as opposed to the eight (the developers recommended sample size). Additionally, developers recommended that the cognitive interviews focus mostly on the younger group and therefore the interviews were skewed more to the younger age bracket. All the children were able to read and write (7=primary school, 2=high school), and able to self-complete the questionnaire (mean 4.9min, median 4.5min). (Table 4.1)

Table 4.1 Summary of general characteristics of the EQ-5D-Y cognitive interview participants

Participant Characteristics (n=9)	Summary
Age 8 to 15yrs	Mean 9.6yrs Median 8yrs
• Aged 8-9yrs • Aged 10-11yrs • Aged 12-13yrs • Aged 14-15yrs	5 2 1 1
Gender	Female=2 Male =7
Time taken to complete questionnaire	Mean 4.9 minutes Median 4.5 minutes
Medical condition	Healthy=7 Sick=2
Education	Primary school=7 High school=2 All able to read, write and self-complete questionnaire.
Familiarity with instrument	None

4.5.1.2 EQ-5D-Y translation

There were several conceptual and linguistic problems identified from the translation process. Most of these were identified and resolved by translators during the first consensus process. However, some translation issues, particularly pertaining to concepts were resolved during the cognitive interviews. The results of the cognitive interviews are discussed below. Regarding translation, the first amendment was from the Chichewa adult EQ-5D-3L version. The bracketed (Nyanja) was removed from the heading Chichewa (Nyanja) for Malawi' as Nyanja is no longer spoken in Malawi but in some neighbouring parts of Zambia and Mozambique. This may not have altered the concept but was done to be consistent with the language being translated into.

Conceptual issues:

Health

From the outset, it became apparent that translation of the term "health" was a problem. "Health" was translated differently by translators as 'umoyo' which is general health; 'moyo' which is life; and 'thanzi' which is about being healthy. This was discussed at length in the consensus discussion with the translators, and the agreement was that 'moyo' is about being alive and not necessarily about being healthy (good life) or health per se (umoyo). '*Umoyo'* was agreed upon as the correct translation for the term "health" through group consensus among the translators.

Pain or Discomfort

The adult Chichewa EQ-5D-3L version had translated 'pain' as '*ululu'*. This word does mean pain but implies extreme pain and is not frequently used by children. Thus 'pain' was translated as '*kuphwanya m'thupi*' (body pain) in the EQ-5D-Y-5L. The aspect of physical pain was added because pain can also be interpreted as mental anguish and the VMC's definition of

terms indicates that pain is intended to refer to physical pain. Additionally, the conjunction 'kapena' ("or") seems not to have the same effect as in English where it separates 'pain' from 'discomfort'. So "*m'thupi*" (body) was added to the descriptor even though it appears repetitively in the same sentence. This is not problematic as this is often the case in the Chichewa language.

'Discomfort' was translated '*kusamva bwino*' in contrast with '*kuphwanya m'thupi kosowetsa mtendere*' in the adult EQ-5D-3L version. The former was opted because '*kuphwanya m'thupi kosowetsa mtendere*' conveys pain of greater intensity, which differs from the VMC's intended definition of discomfort as "uncomfortable physical sensation, of a *lower* grade of intensity than pain". (263)

Doing usual activities

This was translated 'kuchita zinthu za tsiku ndi tsiku' instead of 'zochitika za tsiku ndi tsiku' as in the adult Chichewa EQ-5D-3L version. This was because 'zochitika' implies 'things or activities that happen' whereas 'kuchita' is the ability to perform certain tasks. Additionally, the response options for the adult Chichewa EQ-5D-3L version uses 'ntchito' (work) for activities, and so 'zinthu' (activities) was used for the EQ-5D-Y version.

Feeling worried, sad or unhappy

The first response option '*I am not worried, sad or unhappy'* was initially translated '*Sindikudandaula, sindikumva chisoni kapena osasangalala*'. However, *unhappy is osasangalala* and in this case there is a negation of unhappy i.e. *not unhappy meaning ndikusangala* (happy). The 'not' worried and sad had this captured already with prefix *'si'* but

the '*not unhappy'* was being captured as *'not happy'.* During cognitive interviews this was changed to '*ndikusangalala'* (happy).

Semantic issues:

Other deviations from the existing Chichewa adult EQ-5D-3L version included changing the expression 'I have (no problems) ...' that was translated '*Ndimakhala ndi*...' in the adult version to '*Ndili ndi*…'. This is because '*Ndimakhala ndi*… 'carries past/present continuous action, i.e. how the person is (always has been) up to this very moment. This does not reflect the thought in the instrument which is about how one feels on a day, 'today'. Other phrases explored and rejected included '*Ndili ndi vuto'*. The phrase does imply a problem with performing a task due to (in)ability to do so but could not because of one's attitude (child being disobedient for example) or other issues.

The Chichewa adult EQ-5D-3L version '*No <u>problems</u>*…' translated '*Ndilibe <u>vuto</u>*…' was changed in the EQ-5D-Y version to '*Ndilibe <u>mavuto</u>...'*. This is because—*problems*—as it appears in the source document belong to the 'li-ma' class in Chichewa i.e. *vuto* is singular as opposed to plural *mavuto*.

Additionally, the severity qualifier 'some problems' is translated '*mavuto ena*' in the adult EQ-5D-3L version. However, '*ena'* relates to 'other' and so '*mavuto ena*' could imply other problems on top of, or in addition to, the current problem relating to walking. '*Ena*' was therefore removed in the severity level. Further, while the qualifier '*a lot'* was translated '*kwambiri'* for 'having pain or discomfort' dimension, this was translated as '*aakulu'* for 'mobility', 'looking after oneself' and 'doing usual activities' dimensions. Even though both imply intensity of a problem, '*aakulu'* is an active verb and therefore appropriate for 'mobility',

'looking after oneself' and 'doing usual activities' dimensions. On the other hand, '*kwambiri'* is more about feeling and therefore suitable for 'having pain or discomfort' dimension.

Another feature introduced in the Chichewa EQ-5D-Y version was inclusion of a qualifier for each of the discrete concepts in the 'having pain or discomfort', and 'feeling worried, sad or unhappy' dimensions. For example, the English version reads 'I am a bit worried, sad or unhappy' and this is understood to mean a bit relating to 'feeling worried, sad or unhappy'. However, this could be quite confusing in the Chichewa version as it could be read that 'a bit' applies to 'worry' only. In order to overcome this potential problem, the qualifier was included for each of these concepts so that the responder is clear that this applies to 'feeling worried', 'sad', and 'unhappy'.

4.5.1.3 Cognitive interviews

There were a few conceptual and linguistic issues that were identified from the interviews. The concept of health was similarly observed to be difficult to explain as there was no single way that emerged to describe it in Chichewa even though the appropriate translation for this was '*umoyo'*. Most participants mentioned that 'umoyo' was about how one's daily life was/ how the body was feeling compared to 'moyo', which was the state of being alive, i.e., opposite of death. Some of the participants were not able to draw this distinction between these two terms, and a few others expressed no opinion regarding this but agreed that this was about health.

The various words that translated pain into Chichewa were also tested in cognitive interviews to ascertain the correct rendering. '*Kupweteka'* was tested alongside '*ululu'* as used in the adult Chichewa EQ-5D version. Most children apart from one participant who expressed the idea of changing '*kupweteka'* to '*ululu'*, expressed ignorance regarding what '*ululu'* meant so

‘*kupweteka’* was retained. However, some examples that were given for this concept related to emotional as opposed to bodily pain. After further testing, therefore, ‘*kupweteka’* (pain) was changed to ‘*kuphwanya m'thupi*’ (bodily pain) since the former can mean any kind of pain. Additionally, discomfort had earlier been translated *'kuphwanya m'thupi kosowetsa mtendere*' but it became apparent during the cognitive interviews that this was considered as extreme pain. This meant that only pain was being captured in this dimension. A further 15 adults within MLW were randomly asked about what they thought about the right translation for pain or discomfort. Almost all adults agreed that ‘*kuphwanya m'thupi’* is body pain but even among them there was seeming confusion in the correct rendering for discomfort. Most agreed that ‘*discomfort’* could be translated *'kusamva/kusapeza bwino’*, ‘*kusowa mtendere'*, or ‘*kunyong'onyeka*’. However, the latter is an unfamiliar vocabulary for this population. Additionally, *'kusowa mtendere'* could literally be translated ‘to be troubled’ which carries a psychological concept. One out of the three sick participants interviewed differentiated *'kusapeza bwino'* and *'kuphwanya m'thupi'* in that the latter is about getting sick whereas the former is a general feeling of being unpleasant e.g. due to fatigue. However, when asked to explain, the participant was also alluding to the fact that ‘*kusapeza or kusamva bwino’* could be equated to being sick- which could or not be the case with discomfort. Conclusively, in as much as *'discomfort'* might not have an accurate translation to capture this concept in Chichewa language, *'kusamva bwino'* was deemed the closest and most appropriate.

Regarding the use of the qualifier ‘*aakulu’* for some ‘mobility’, ‘looking after oneself’, and ‘doing usual activities’ dimensions, and ‘*kwambiri*’ for ‘having pain or discomfort’, and ‘feeling worried, sad or unhappy’ dimensions, all the participants during interviews found that this was appropriate.

Lastly, ‘health questionnaire’ was translated *'Chikalata ya mafunso a za Umoyo' (Document about health questionnaire).* During cognitive interviews the respondent’s preference was for prefix “*Chi-*“ to be dropped so that it should be *'Kalata ya mafunso a za Umoyo'.* This is because in Chichewa grammar the prefix ‘Chi-‘ implies something big which is not the case with the EQ-5D-Y. Further reflection on this translation, however, led to removal of ‘*Kalata’* which can also translate ‘letter’, and would not be an appropriate translation for ‘questionnaire’. Children further suggested making changes to some set of instructions. For example, ‘describing your health today’ the Chichewa word *‘fotokoza’* (explain/describe) was preferred as opposed to ‘*kufotokoza’* (explaining/describing). It appears children in school were familiar with using *fotokoza* (explain) from the lessons taught in class and ‘*kufotokoza’* appeared to have been asking for a bit more information. The EuroQol Group, however, decided to remove this instruction on the final Chichewa EQ-5D-Y version. Additionally, *'LEROLO'* was opted during cognitive interviews with children compared to the translated word *'pa tsiku lalero'* because even though both meant TODAY, the former was much shorter.

As regards the EQ VAS, suggestion was made during cognitive interviews that ‘*Nambala ya…’* (the number…) should be added to the scale in order to make more meaning to the text that refers to the numbers 0 and 100 on the scale. The children had earlier suggested to change ‘**a**kuntanthauza’ to ‘**i**kutanthauza’ both translates means but because the prefix ‘a-‘ relates to a person than an object, it was therefore grammatical correct to use the latter. Also, the participants suggested adding *this (uwu)* to the instruction about putting a mark X on the EQ VAS. Most participants had the mark next to the line and not on the line. So perhaps adding *this* could make it clearer the need to put the mark X on the actual EQ VAS.

When it came to assessing comprehension of concepts, these were generally understood by participants. Appropriate examples were given for different dimensions and respective response options. For example, someone with broken arm was an example given of someone with problems looking after themselves. ‘Doing usual activities’ included such examples as going to school, market, looking after siblings, and washing clothes, whereas problems to do with ‘doing usual activities’ they mentioned being sick. For the ‘having pain or discomfort’ dimension, being sick, having diarrhoea, having headache, upset tummy or tonsillitis were examples that were given for this domain. Appropriate examples were also given for the ‘Feeling worried, sad or unhappy’ dimension.

Suggested changes were discussed with translators and the VMC reviewer, and the version was amended accordingly. After no further changes were suggested during the proof-reading stage, the EuroQol office received and approved the instrument. (268) (Figure 4.9)

Figure 4.9 EuroQol approved Chichewa (Malawi) EQ-5D-Y self-complete paper version

Mafunso a za Umoyo

Chichewa cha ku Malawi

(Chichewa version for Malawi)

85

80

75

Pansi pa mutu ulionse, chonga bokosi LIMODZI limene likufotokoza bwino za umoyo wako LEROLO.

Mayendedwe *(kutha kuyenda)*

Ndilibe mavuto alionse poyenda ❑
Ndili ndi mavuto pang'ono poyenda ❑
Ndili ndi mavuto aakulu poyenda ❑

Kudzisamalira ndekha

Ndilibe mavuto posamba kapena kudziveka ndekha ❑
Ndili ndi mavuto pang'ono posamba kapena kudziveka ndekha ❑
Ndili ndi mavuto aakulu posamba kapena kudziveka ndekha ❑

Kuchita zinthu za tsiku ndi tsiku *(mwachitsanzo, kupita ku sukulu, kuchita zimene ndimakonda, masewero olimbitsa thupi, kusewera, kuchita zinthu ndi banja langa kapena anzanga)*

Ndilibe mavuto alionse pochita zinthu zanga za tsiku ndi tsiku ❑
Ndili ndi mavuto pang'ono pochita zinthu zanga za tsiku ndi tsiku ❑
Ndili ndi mavuto aakulu pochita zinthu zanga za tsiku ndi tsiku ❑

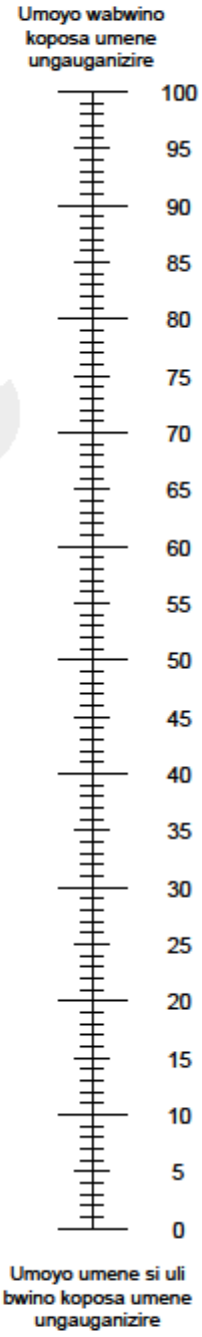
Kumva kuphwanya m'thupi kapena kusamva bwino

Sindikumva kuphwanya m'thupi kapena ndikumva bwino ❑
Ndikumva kuphwanya m'thupi pang'ono kapena sindikumva bwino pang'ono ❑
Ndikumva kuphwanya m'thupi kwambiri kapena sindikumva bwino kwambiri ❑

Kudandaula, kumva chisoni kapena kusasangalala

Sindikudandaula, sindikumva chisoni kapena ndikusangalala ❑
Ndikudandaula pang'ono, ndikumva chisoni pang'ono kapena ndikusangalala pang'ono ❑
Ndikundaula kwambiri, ndikumva chisoni kwambiri kapena sindikusangalala ❑

- Tikufuna tidziwe m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino LEROLO.
- Mzere uwu walembedwa manambala kuyambira pa 0 mpaka 100.
- Nambala ya 100 ikutanthauza umoyo wabwino koposa umene ungauganizire.
  Nambala ya 0 ikutanthauza umoyo umene suli bwino koposa umene ungauganizire.
- Chonde lemba chizindikiro cha X pa mzere uwu pamene pakuonetsa m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino LEROLO.



4.5.2 EQ-5D-Y-5L

4.5.2.1 Participants and recruitment:

The cognitive interviews were carried out in four separate rounds. The first round involved seven participants (four healthy and three sick) as per the EQ-5D-Y-5L translation protocol. Each of the next three rounds comprised of four (all healthy) participants. All participants were attending primary school, and none were already familiar with the EQ-5D-Y-5L. In total, 19 participants (10 boys, nine girls) aged 8-14 years (mean 9.6 years), were interviewed over the four rounds. After consultation with the VMC reviewer, data from first respondent was removed from the results due to some confusion the child had in understanding the task. This confusion could have arisen from either that this was the first interviewee and it is possible that the interviewer did not clearly articulate instructions, or the child may have been confused by the task, or a combination of both. Results from the remaining 18 participants (3 sick -all acutely ill- and 15 healthy) are presented. (Table 4.2)

Table 4.2 Summary of general characteristics of the EQ-5D-Y-5L cognitive interview participants

Participant Characteristics (n=18)	Summary
Age 8 to 15yrs	Mean 9.7yrs Median 9yrs
• Aged 8-9yrs • Aged 10-11yrs • Aged 12-13yrs • Aged 14-15yrs	9 6 2 1
Gender	Female=8 Male =10
Medical condition	Healthy=15 Sick=3
Education	Primary school=18 All able to read, write and self-complete questionnaire.
Familiarity with instrument	None

4.5.2.2 EQ-5D-Y-5L translation

Since the EQ-5D-Y-5L translation followed the EQ-5D-Y translation, this translation concentrated only on new text that appear in the former but are not in the latter. Still, one translation in particular was re-examined during the translation of the EQ-5D-Y-5L. '*Ndili ndi vuto* ' (I have problem) imply a problem to perform a certain task due to in/ability to do so. This could similarly imply being unable to perform a task because of one's attitude (child being disobedient for example) or issues towards that task. Before retaining this translation in the EQ-5D-Y-5L, a third person was asked and suggested translating this as 'Ndikuvutikirapo' (I have problem). However, '*Ndikuvutikirapo*' could mean a struggle which loosely imply some difficulty/problems in performing a task. This would have created difficulties in adding qualifiers for all other response levels 2-4 for dimensions 1-3. For this reason, the EQ-5D-Y translation was retained in the EQ-5D-Y-5L.

There were a few problems when translating two new response options that are in the EQ-5D-Y-5L but not in the EQ-5D-Y. The 'mobility" extreme response 'I cannot walk' was translated '*Sindingathe kuyenda*' or '*Sindingayende*'. Whereas both translations are accurate, this was changed to '*Ndikulephera kuyenda'* because '*sindinga* -' carries the thought of impossibility to walk. '*Ndikulephera'* on the other hand means 'failing' to walk at a particular time and possibility of doing so in future which is the thought for this item. Similarly, the fifth severity level in the 'looking after oneself' dimension 'I cannot', which is not in the EQ-5D-Y version, was changed to '*Ndikulephera'* instead of '*Ndimalephera'* as it appears in the adult EQ-5D-3L version. The latter implies a continuous action as opposed to how one feels or are at a point in time - today.

Some qualifiers were also generally not easy to translate. The 'extreme' qualifier in 'feeling worried, sad or unhappy' dimension was translated '*moonjeza'* and '*mopitirira'* which are used interchangeably because they got the same meaning. The former was however assessed as most appropriate and retained in this context.

In the end, the five qualifiers were translated as follows: '*Ndilibe'* (No), '*pang'ono kwambiri'* (a little bit), '*pang'ono'* (some), '*aakulu /ambiri'* (a lot/ quite), and '*ndikulephera /moonjeza'* (cannot / extreme/extremely). At this point the questionnaire was then piloted through cognitive interviews.

4.5.2.3 Cognitive interviews

Cognitive interviews *(Table 4.3)*

i) Card ranking exercise (Table 4.4)

In the first round, each of the six participants ranked five severity levels for the four dimensions to give a total of 120 rankings (Table 4.4). Of these, 43% (52/120) across all dimensions were incorrect rankings from the intended order. Of the incorrect rankings, 62% (32/52) were attributed to severity level inversion between the Chichewa wording for '*a little bit*' and '*some*'. It was evident that '*a little bit*', initially translated '*pang'ono kwambiri*', was the source of the confusion. The Chichewa words '*pang'ono*' and '*kwambiri'* mean '*less*' and '*a lot*', respectively. While the combination of these two words was supposed to imply '*a lot less*', it was interpreted as '*a lot more',* and led participants to interpret severity level two as worse than that of level three. A further 14% of incorrect rankings were due to severity level inversion between the Chichewa words for '*a lot*' (6%) and '*extreme*' (8%). The rest of the incorrect rankings were made up of a mixture of incorrect ordering among all the severity levels. Since

there were several cases of severity level inversion between ‘a little bit’ and ‘some’, this was revised and then re-tested.

The re-testing in round two indicated that the change in translation generally resolved the severity level inversion for levels two (*a little bit*) and three (*some*). The severity level inversion was only evident in one respondent for both ‘having pain or discomfort’ and ‘feeling worried, sad or unhappy’, and since this was equal to only two responses, it was judged that this change had satisfactorily resolved the problem. However, half of the incorrect rankings in round 2 were then due to severity level inversion between the qualifiers for levels four (*a lot*) and five (*extreme*), which re-emerged in all dimensions except for ‘looking after oneself’. It was discovered that the translated qualifier for severity level four (*kwambiri)* was generally understood to mean “*extreme*” by participants because in their everyday usage *kwambiri* is understood to mean *extreme*. The severity level five qualifier was therefore changed to ‘*moonjeza kwambiri’* to emphasize the intensity and this was then re-tested during the third round of iterative interviews.

In round three of testing, 12 of all 28 incorrect rankings were still due to translated severity level inversions between qualifiers for levels four (*a lot*) and five (*extreme*). The latter was therefore further changed to ‘*kwambiri zedi’* for the ‘having pain or discomfort’ and ‘feeling worried, sad or unhappy’ dimensions and then re-tested. The qualifiers for the other three dimensions were not changed because they have a different qualifier for level five (cannot), which was not problematic for the translated version.

The final round of testing showed that the changes made to the qualifier for extreme severity level five had resolved the severity inversion for levels four and five. At this point, the EQ-5D-

Y-5L translated version was deemed to have established appropriate hierarchical ordering. The translated questionnaire was then ready for proofreading ahead of self-completion and in-depth interviews.

ii) Self completion of the questionnaire

Appropriate self-completion of the questionnaire without interviewer assistance was problematic among some participants in this setting. Seven of ten participants who were aged 8-10 years tended to tick every box instead of choosing one option for each of the five dimensions, which could indicate a lack of understanding of the instructions. However, once the interviewer read and explained the instructions clearly to the participants, they were able to appropriately choose one severity level per dimension. (Table 4.4)

Table 4.5 also shows that the participants with health conditions reported health states that were sub-optimal, and reported lower scores on the EQ VAS scale, compared with healthy participants. On the other hand, all healthy participants reported perfect health on the descriptive system and corresponding high EQ VAS scores (where one child reported 95 and six participants reported 100). One sick participant, with a poor health state of 32132, had a high EQ VAS score of 95, whereas another sick participant with a better health state of 11222 provided a lower (worse) EQ VAS score of 60.

5) *cognitive interviews about comprehension of the questionnaire*

Similar to the translation work, it emerged during cognitive interviews that some participants struggled to draw a distinction between health, (umoyo), healthy (thanzi) and life (moyo). Most participants mentioned that ‘umoyo’ was about how one’s daily life was/how the body was feeling compared to ‘moyo’, which was the state of being alive, i.e. opposite of death. Some

of the participants were not able to draw this distinction between these three terms, and a few others expressed no opinion regarding, this but agreed that this was about health. Therefore, the word ‘umoyo’ was kept as decided during the translation process.

Despite this, most participants were able to give appropriate examples of the different severity levels. For example, when it came to examples to illustrate ‘a little bit’, ‘some’ and ‘a lot’ of problems in the ‘mobility’ dimension, the participants described such examples as someone with a sprained knee, an injured leg and having mobility issues due to some health problems, respectively. For ‘a little bit’ in the ‘doing usual activities’ dimension, participants gave examples of someone who was unwell but able to go to school; and similarly, appropriate examples were given for ‘a little bit’, ‘some’ and ‘a lot’ of problems in the ‘having pain or discomfort’ dimension. Examples such as someone with an upset tummy or with tonsillitis indicated appropriate comprehension for ‘having pain or discomfort’. It was also very clear that most participants interviewed understood the ‘feeling worried, sad or unhappy’ concept as evidenced by examples given for this dimension. Participants cited that losing a friend or losing a loved one to death, lack of basic needs or money, lack of food at home, or home-based violence, or lack of school materials would result in different levels of worry, sadness or unhappiness.

There were also some inaccurate responses. In the ‘looking after oneself’ dimension, two of the ten participants gave examples of severity levels relating to lack of provisions such as soap or water as examples of ‘some problems’. In the ‘having pain or discomfort’ dimension, two different participants gave examples that related to loss of a loved one in death (emotional pain) or a friend that is sick and could not attend school as things that would make someone ‘having pain or discomfort’. Another participant gave an example for ‘a lot’ of ‘having pain or

discomfort' of an adult who is diagnosed with a certain disease (e.g. epilepsy) and therefore unable to do their usual activities. From a child's perspective, an epileptic person would be affected in physical functioning, although this example may pertain more to emotional problem as opposed to bodily pain.

At the end of the interviews, when asked to make suggestions to change the wording of dimensions or severity levels at no point during the in-depth interviews did the participants make additional meaningful suggestions to further improve the Chichewa wording of dimensions and severity levels.

Table 4.3 EQ-5D-Y-5L cognitive interview summary

Characteristics	**Value**
No of interviews conducted (N)	18
No of interviews that included all three activities (i. ranking game; ii. Self-completion of questionnaire; and iii. in-depth interviews)	10
Average time taken to complete all three activities (mins)	31.25
Needed help to self-complete questionnaire	7
No. of interviews involving card ranking exercise only (*n*)	8
*Average time taken to complete the ranking exercise (mins)**	22

* based on the last 8 only, since the first 10 involved all three activities.

Table 4.4 EQ-5D-Y-5L participant rankings of the card ranking exercise progressing over four rounds

	Box number card is placed into*																	
	Round 1						Round 2				Round 3				Round 4			
	R 1	R 2	R 3	R 4	R 5	R 6	R 7	R 8	R 9	R 10	R 11	R 12	R 13	R 14	R 15	R 16	R 17	R 18
Set 1 (Pain or discomfort)																		
No pain or discomfort	1	1	1	1	1	1	1	1	1	1	1	3	2	1	1	1	1	1
A little bit of pain or discomfort	3	3	3	3	4	2	2	2	2	3	2	2	1	2	2	2	2	**2**
Some pain or discomfort	4	2	2	2	2	3	3	3	3	2	3	1	3	3	3	3	3	**3**
A lot of pain or discomfort	5	4	4	4	5	4	5	5	4	4	5	5	5	4	4	4	4	**4**
Extreme pain or discomfort	2	5	5	5	3	5	4	4	5	5	4	4	4	5	5	5	5	**5**
Set 2 (Worried, sad or unhappy)																		
Not worried, sad or unhappy	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1
A little bit worried, sad or unhappy	3	3	3	3	4	3	2	3	2	2	2	2	3	2	2	2	2	**2**
Quite worried, sad or unhappy	2	2	2	2	2	2	3	2	3	3	3	1	2	3	3	3	3	**3**
Really worried, sad or unhappy	4	5	4	4	5	4	4	4	5	4	4	5	4	4	4	4	4	**4**
Extremely worried, sad or unhappy	5	4	5	5	3	5	5	5	4	5	5	4	5	5	5	5	5	**5**
Set 3 (Walking about)																		
No problems walking about	1	1	1	1	1	1	1	1	1	1	1	5	3	1	1	1	1	1
A little bit of a problem walking about	5	3	3	3	3	2	3	2	2	2	2	2	2	2	2	2	2	**2**
Some problems walking about	2	2	2	2	2	3	4	3	3	3	3	1	1	3	3	3	3	**3**
A lot of problems walking about	3	5	4	4	4	4	5	5	4	4	4	4	4	5	4	4	4	**4**
Cannot walk about	4	4	5	5	5	5	2	4	5	5	5	3	5	4	5	5	5	**5**
Set 4 (Washing or dressing)																		
No problems washing or dressing him/herself	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A little bit of a problem washing or dressing him/herself	2	3	3	3	2	3	2	2	2	2	2	5	2	2	3	2	2	**2**
Some problems washing or dressing him/herself	3	2	2	2	3	2	3	3	3	3	3	3	3	3	2	3	3	**3**
A lot of problems washing or dressing him/herself	4	4	4	4	4	4	4	4	4	4	5	2	4	4	4	4	4	**4**
Cannot wash or dress himself/herself	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	**5**

**The box number represent respondent placing of the sorting card. No two cards were placed in the same box -otherwise these would have been given the same box number.*

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

The Chichewa translated cards were administered

In terms of the questionnaire instructions, these were found to have generally been understood by all participants except for in younger participants (≤12 years). Additionally, the participants with health conditions reported health states that were sub-optimal, and reported lower scores on the EQ VAS scale, compared with healthy participants. On the other hand, all healthy participants reported perfect health on the descriptive system and corresponding high scores (one child reported 95 and six participants reported 100) on the EQ VAS scale. (Table 4.5) However, inconsistency, as perhaps may be expected (181), between the EQ-5D-Y-5L descriptive health system and the EQ VAS score was observed. For example, one sick child with an extreme health state of 32132 had a high EQ VAS score of 95, whereas another sick child with a better health state of 11222 provided a much worse EQ VAS score of 60.

Table 4.5 Problem severity levels from the beta EQ-5D-Y-5L according to each of the dimensions

MOBILITY (walking about)										
No problems	1	1	1		1		1	1	1	1
A little bit						2				
Some problems				3						
A lot of problems										
Cannot do										
LOOKING AFTER MYSELF										
No problems	1	1	1		1	1	1	1	1	1
A little bit of a problem				2						
Some problems										
A lot of problems										
Cannot do										
DOING USUAL ACTIVITIES										
No problems	1	1	1	1		1	1	1	1	1
A little bit of a problem					2					
Some problems										
A lot of problems										
Cannot do										
HAVING PAIN OR DISCOMFORT										
No	1	1	1			1	1	1	1	1
A little bit					2					
Some				3						
A lot										
Extreme										
FEELING WORRIED, SAD OR UNHAPPY										
Not	1	1	1			1	1	1	1	1
A little bit				2	2					
Quite										
Really										
Extremely										
Descriptive health profile summary*	**11111**	**11111**	**11111**	**32132**	**11222**	**21111**	**11111**	**11111**	**11111**	**11111**
VAS SCORE	100	100	100	95	60	100	95	100	100	100

*Where '11111' represents 'No problems' reported for all five dimensions, and where '55555' would represent 'Extreme problems' on all five dimensions
The Chichewa translated version was administered

Taking into account suggested changes that emerged from the ranking exercise and in-depth interviews, the Chichewa EQ-5D-Y-5L was revised accordingly as a post-cognitive interview version. This version was forwarded to an independent linguist for a final review. No further changes were suggested, and the version was forwarded to the EuroQol office who approved the instrument. (Appendix 14) (Figure 4.10)

Figure 4.10 EuroQol approved Chichewa (Malawi) beta EQ-5D-Y-5L self-complete paper version

Mafunso a za umoyo

Chichewa cha ku Malawi

(Chichewa version for Malawi)

Pansi pa mutu ulionse, chonga bokosi LIMODZI limene likufotokoza bwino za umoyo wako LEROLO

Mayendedwe (kutha kuyenda)

Ndilibe mavuto alionse poyenda ❑

Ndili ndi vuto pang'ono penipeni poyenda ❑

Ndili ndi mavuto pang'ono poyenda ❑

Ndili ndi mavuto aakulu poyenda ❑

Ndikulephera kuyenda ❑

Kudzisamalira ndekha

Ndilibe mavuto posamba kapena kudziveka ndekha

Ndili ndi vuto pang'ono penipeni posamba kapena kudziveka ndekha ❑

Ndili ndi mavuto pang'ono posamba kapena kudziveka ndekha ❑

Ndili ndi mavuto aakulu posamba kapena podziveka ndekha ❑

Ndikulephera kusamba kapena kudziveka ndekha ❑

Kuchita zinthu za tsiku ndi tsiku (*mwachitsan[illegible]mita ku[illegible]ulu, kuchita[illegible]mene ndimakonda, masewero olimbitsa thupi, kusewera, kuchita zi[illegible]nia langa[illegible]ga*)

Ndilibe mavuto alionse pochita zinthu zanga za[illegible]siku ndi ts[illegible] ❑

Ndili ndi vuto pang'ono penipeni pochit[illegible]ga za tsiku n[illegible]iku ❑

Ndili ndi mavuto pang'ono pochita [illegible]zanga za[illegible] ndi tsiku ❑

Ndili ndi mavuto aakulu pochita zinth[illegible]ga za tsiku n[illegible]iku ❑

Ndikulephera kuchita zint[illegible]na za ts[illegible]di tsiku ❑

Kumva kuphwanya m'[illegible]pi kape[illegible]samva [illegible]

Sindikumva kuphwan[illegible]m'thupi [illegible]ikumva b[illegible]no ❑

[illegible]phwanya [illegible]thup[illegible]ng'ono p[illegible] kapena sindikumva bwino pang'ono penipeni ❑

Ndikumva kuphwanya m'thu[illegible]ang'ono [illegible]pena sindikumva bwino pang'ono ❑

Ndikumva k[illegible]'thupi[illegible] kapena sindikumva bwino kwambiri ❑

[illegible]phwanya m[illegible]pi kwambiri zedi kapena sindikumva bwino kwambiri zedi ❑

Kudandaula, kumva chi[illegible] kapena kusasangalala

Sindikudandaula, sin[illegible]va chisoni kapena ndikusangalala ❑

Ndikudan[illegible]o penipeni, ndikumva chisoni pang'ono penipeni kapena ndikusangalala pang'ono penipeni. ❑

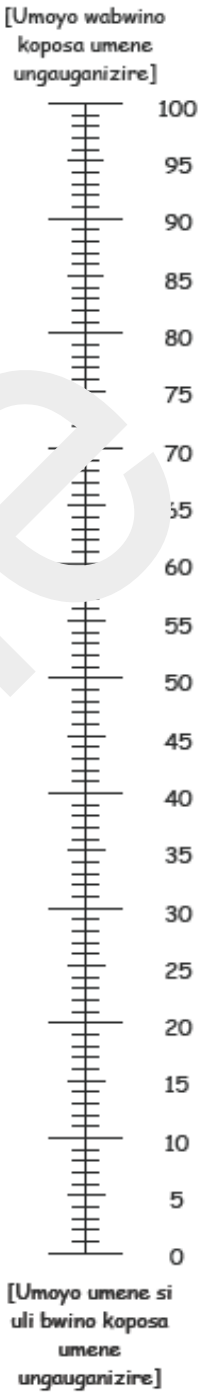
Ndikudandaula pang'ono, ndikumva chisoni pang'ono kapena ndikusangalala pang'ono ❑

Ndikudandaula kwambiri, ndikumva chisoni kwambiri kapena sindikusangalala ❑

Ndikudandaula kwambiri zedi, ndikumva chisoni kwambiri zedi kapena sindikusangalala mpang'ono pomwe. ❑

- Tikufuna tidziwe m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino LEROLO
- Mzere uwu walembedwa manambala kuyambira pa 0 mpaka 100
- Namabala ya 100 ikutanthauza umoyo wabwino koposa umene ungauganizire
  Nambala ya 0 ikutanthauza umoyo umene si uli bwino koposa umene ungauganizire
- Chonde lemba chizindikiro cha X pa mzere uwu pamene pakuonetsa m'mene umoyo wako ulili LEROLO
- Tsopano, lemba nambala imene walembapo chizindikiro cha[illegible] mzere uwu mu bokosi lili m'musimu.

M'MENE UMOYO WAKO ULILI[illegible]O =



4.5.3 PedsQL ™ 4.0 self-report

4.5.3.1 Participants and recruitment:

Five healthy primary school going participants for the child report (mean 11yrs; F=2, M=3) (Table 4.6) and five adolescents were interviewed for the teen report (mean 13yrs; F=3, M=2) (Table 4.7).

Table 4.6 Summary of general characteristics of cognitive interview participants for PedsQL ™ 4.0 Child self-report

Participant Characteristics (n=5)	Summary
Targeted age: 8 to 12yrs	Mean 11yrs Median 11yrs
Aged 10-11yrs Aged 12-13yrs	4 1
Gender	Female=2 Male =3
Medical condition	Healthy=5 Sick=0
Education	Primary school=5 All able to read, write and self-complete questionnaire.
Familiarity with instrument	None

Table 4.7 Summary of general characteristics of cognitive interview participants for PedsQL ™ 4.0 Teen self-report:

Participant Characteristics (n=5)	Summary
Targeted age: 13 to 18yrs	Mean: 13yrs*
Gender	Female=3 Male =2
Medical condition	Healthy=5 Sick=0
Education	Primary school=5 All able to read, write and self-complete questionnaire.
Familiarity with instrument	None

*The sample was mostly from the lower age group

4.5.3.2 PedsQL ™ 4.0 self-report translation process

Linguistic issues:

A few linguistic issues emerged on the translation into the Chichewa language due to limited vocabulary for some words or excess words expressing the same thing for others. For example, the qualifier '*never'* was translated '*silinakhalepo'* and '*silikhala'*, but the former was opted since the latter imply not a recurring problem as opposed to never. The qualifier '*almost always'* was translated '*ngati limakhala vuto nthawi zonse'* and '*ngati lili vuto pafupifupi nthawi zonse'*. The former was adopted as opposed to the latter because the verb 'limakhala'– implies that the problem comes and goes, but not continuous.

Conceptual issues:

There were several conceptual issues that emerged especially due to some of the concepts not being available in the local context. For example, the concept of 'It is hard for me to walk more than one ***block'*** was difficult to conceptualize since distance is not measured in blocks. Additionally, the length of blocks differs even in some states within the USA where this is used as standard practice. Distance was therefore conceptualized in terms of classrooms for the children to easily relate to where one block would be equivalent to three double classrooms. 'It is hard for me to play sports activity or exercise' was also problematic since sports and exercise have one word in Chichewa. 'It is hard for me to take a bath or shower by myself', take shower or bath is same word in Chichewa. The item 'I hurt or it aches'-was translated '*kupweteka'* which is a correct rendering for 'hurt' as well as 'aching' —but in this age group, *kuwawa* was deemed more appropriate. Additionally, aches in Chichewa is a noun and needs a supporting verb which isn't there. 'Not feeling well' in the item 'I miss school because of not feeling well', was translated as 'kusapeza bwino' and 'kudwala'. However, 'kusapeza bwino' was not adopted as it can also be used to imply lack of things i.e., poverty (although this was retained for 'discomfort' dimension for the EQ-5D-Y and EQ-5D-Y-5L).

4.5.3.3 Cognitive interviews

The cognitive interviews also revealed a major conceptual issue that needed revision. The item "It is hard for me to take a bath or shower by myself" was translated "Zikumandivuta ***kusamba***". However, it was discovered that much as 'kusamba' means to 'bathe', it has a cultural connotation to 'menstrual cycle' for the older girls. Consequently, "m'thupi" (body) was added to this item to become "Zikumandivuta ***kusamba m'thupi***". This removed any misconception of the meaning of this item. No other changes were suggested by the participants during this piloting phase.

Generally, the cognitive interviews showed that participants understood the questionnaires as they could give specific correct examples for severity levels. For example, one participant circled sometimes for 'I feel angry' and the participant said they did so because other children did not want to associate with her. There was also a general match between the responses given and the general health condition of the participants considering that the sample was that of a general population. (Table 4.8 and Table 4.9)

Table 4.8 Responses of the participants completing the PedsQL ™ 4.0 child report

	Participants options				
SUB-SCALE (item)*	R1	R2	R3	R4	R5
ABOUT MY HEALTH AND ACTIVITIES (problems with…)					
It is hard for me to walk more than one block	0	1	0	0	1
It is hard for me to run	0	0	0	1	2
It is hard for me to play sports activity or exercise	0	0	0	2	1
It is hard for me to lift something heavy	0	1	0	2	2
It is hard for me to take a bath or shower by myself	0	0	0	1	1
It is hard for me to do chores around the house	0	0	0	2	1
I hurt or it aches	0	0	0	1	1
I have low energy	0	0	0	1	1
ABOUT MY FEELINGS (problems with…)					
I feel afraid or scared	2	0	0	2	1
I feel sad or blue	2	0	0	2	1
I feel angry	2	0	0	2	1
I have trouble sleeping	0	0	0	1	2
I worry about what will happen to me	0	0	0	2	1
HOW I GET ALONG WITH OTHERS (PROBLEMS WITH…)					
I have trouble getting along with other kids	0	0	0	1	1
Other kids do not want to be my friend	2	0	0	1	2
Other kids tease me	3	0	0	1	1
I cannot do things that other kids of my age can do	3	0	0	2	2
It is hard to keep up when I play with other kids	0	0	0	1	1
ABOUT SCHOOL (PROBLEMS WITH…)					
It is hard to pay attention in class	0	0	0	1	1
I forget things	2	0	0	2	1
I have trouble keeping up with my schoolwork	0	0	0	1	1
I miss school because of not feeling well	2	0	0	2	2
I miss school to go to the doctor or hospital	2	0	0	1	1

*The Chichewa translated version was administered

Table 4.9 Responses of the participants completing the PedsQL ™ 4.0 teen report

SUB-SCALE (item)*	R1	R2	R3	R4	R5
ABOUT MY HEALTH AND ACTIVITIES (problems with…)					
It is hard for me to walk more than one block	2	0	0	0	1
It is hard for me to run	1	0	1	2	0
It is hard for me to play sports activity or exercise	1	0	1	0	0
It is hard for me to lift something heavy	2	0	1	1	0
It is hard for me to take a bath or shower by myself	1	0	1	0	
It is hard for me to do chores around the house	2	0	1	0	0
I hurt or it aches	1	0	1	0	0
I have low energy	2	0	1	0	0
ABOUT MY FEELINGS (problems with…)					
I feel afraid or scared	1	0	1	2	0
I feel sad or blue	1	2	2	1	0
I feel angry	1	0	0	1	1
I have trouble sleeping	1	0	0	0	0
I worry about what will happen to me	2	0	1	0	0
HOW I GET ALONG WITH OTHERS (PROBLEMS WITH…)					
I have trouble getting along with other teens	2	0	1	4	0
Other teens do not want to be my friends	1	0	1	1	0
Other teens tease me	1	0	1	0	0
I cannot do things that other teens of my age can do	3	0	1	2	0
It is hard to keep up with my peers	1	2	1	1	0
ABOUT SCHOOL (PROBLEMS WITH…)					
It is hard to pay attention in class	2	0	2	1	0
I forget things	2	0	3	1	1
I have trouble keeping up with my schoolwork	1	0	1	0	0
I miss school because of not feeling well	2	0	0	2	0
I miss school to go to the doctor or hospital	2	0	1	2	0

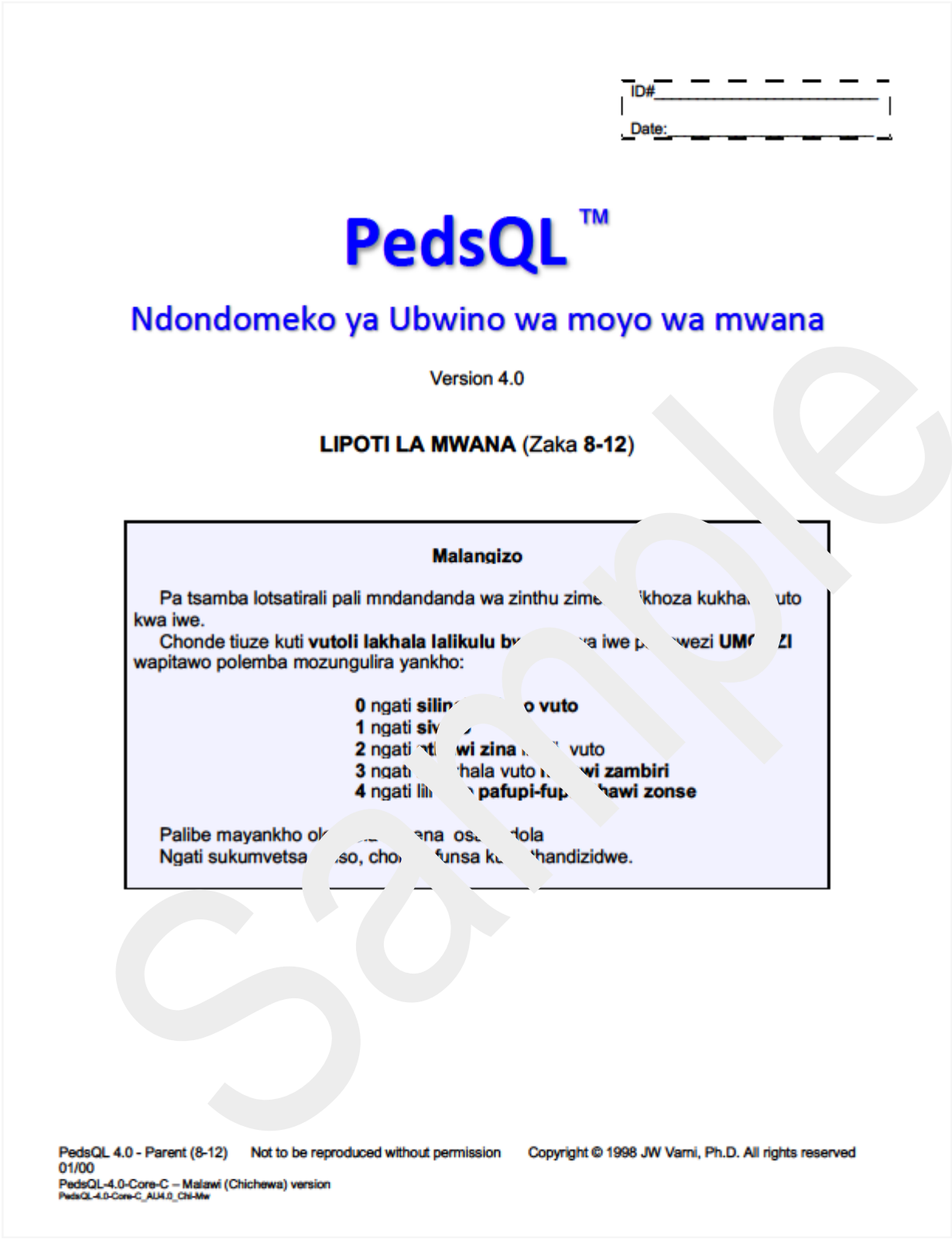
*The Chichewa translated version was administered

Although, the participants generally understood the meaning of the concepts and items, self-completion of the questionnaire without interviewer assistance was also problematic for these instruments. Similar to what was observed with the EQ-5D-Y above, some participants would circle every box instead of choosing one severity level per domain. Once the interviewer read and explained the instructions clearly to the participants, this seemed to address the problem.

After no further changes were suggested, the versions were proofread and subsequently forwarded to Mapi Trust for approval. (Appendix 15- MAPI Trust approval) The approved

versions are in Figures 4.11 and 4.12.

Figure 4.11 Mapi Trust approved Chichewa (Malawi) PedsQL ™ 4.0 Child report version



ID#_____

Date:_____

PedsQL™

Ndondomeko ya Ubwino wa moyo wa mwana

Version 4.0

LIPOTI LA MWANA (Zaka **8-12**)

Malangizo

Pa tsamba lotsatirali pali mndandanda wa zinthu zime… …khoza kukha… …uto kwa iwe.

Chonde tiuze kuti **vutoli lakhala lalikulu b**… …a iwe p… …wezi **UM**… …**ZI** wapitawo polemba mozungulira yankho:

0 ngati **silin**… …**o vuto**
1 ngati **siv**…
2 ngati …**wi zina**… …vuto
3 ngati… …hala vuto… …**wi zambiri**
4 ngati li… …**pafupi-fup**… …**hawi zonse**

Palibe mayankho ol… …ena os… …dola
Ngati sukumvetsa… …so, cho… …unsa k… …handizidwe.

PedsQL 2

*Mu **mwezi UMODZI** wapitawu, kodi **vuto** ili lakhala lalikulu bwanji kwa iwe …*

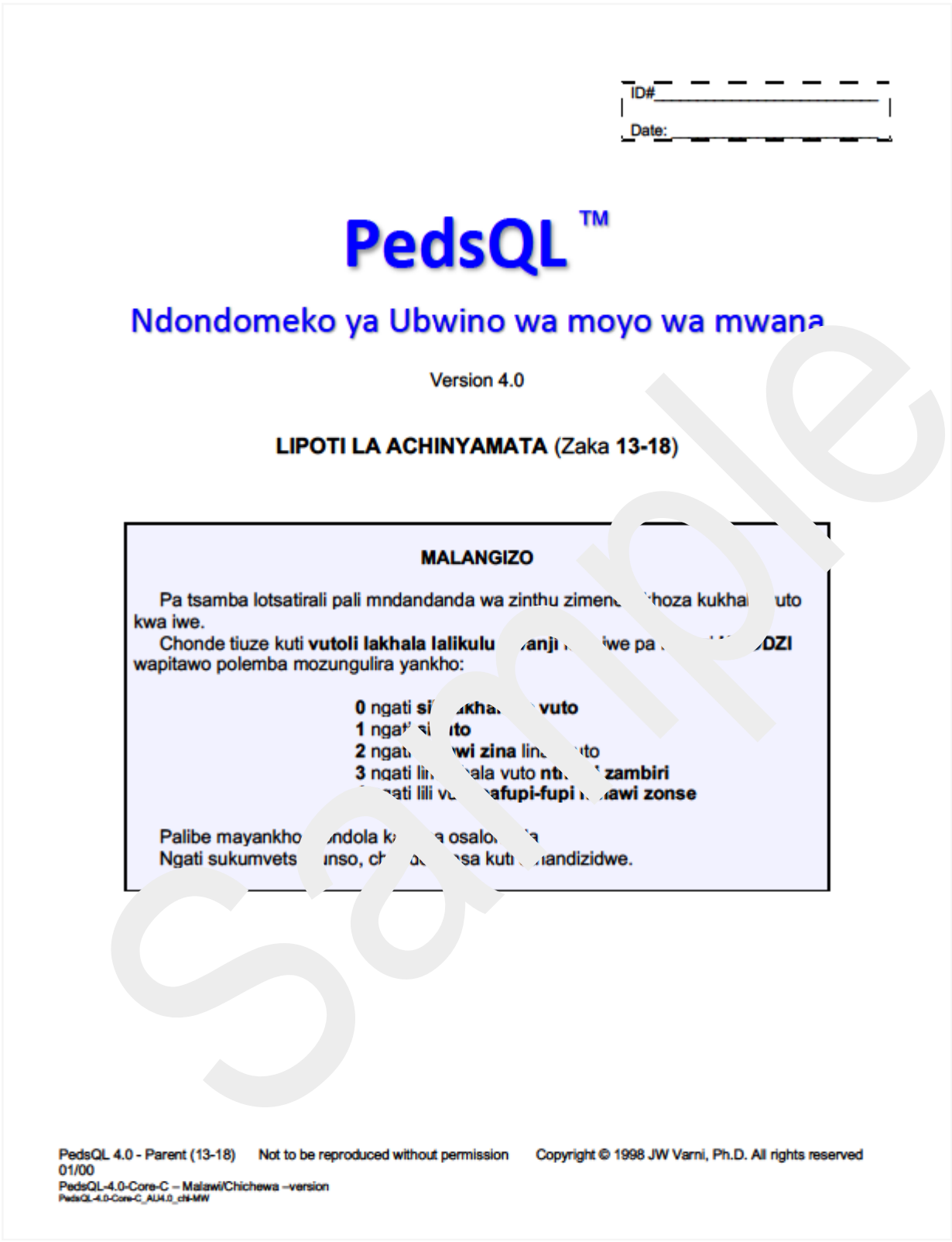
ZOKHUDZANA NDI UMOYO WANGA KOMANSO ZINTHU ZIMENE NDIMACHITA*(mavuto ndi …)*	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Nthawi zambiri**	**Pafupi-fupi nthawi zonse**
1. Zikumandivuta kuyenda mopitilira mabuloku atatu a sukulu	0	1	2	3	4
2. Zikumandivuta kuthamanga	0	1	2	3	4
3. Zikumandivuta kuchita masewero olimbitsa thupi	0	1			4
4. Zikumandivuta kunyamula chinthu cholemera	0	1			4
5. Zikumandivuta kusamba m'thupi	0	1			4
6. Zikumandivuta kugwira ntchito za pakhomo	0			3	4
7. Ndikumamva kuwawa	0	1	2		4
8. Ndikumafooka		1		3	4

ZOKHUDZANA NDI M'MENE NDIKUMAMVERA *(mavuto ndi …)*	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Kawiri-kawiri**	**Pafupi-fupi nthawi zonse**
1. Ndikumaopa kapena ndikumakhala ndi manth		1	2	3	4
2. Ndikumamva chisoni kapena ndikumadand	0	1	2	3	4
3. Ndikumakwiya	0		2	3	4
4. Ndikumavutika kugona			2	3	4
5. Ndikumadandaula ndi chimene	0	1	2	3	4

M'MENE NDIMAKHALALIRA NDI ANA E *(mavuto ndi …)*	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Kawiri-kawiri**	**Pafupi-fupi nthawi zonse**
1. Zikumandivuta kuc za ndi a	0	1	2	3	4
2. mafuna chala zanga	0	1	2	3	4
na ena amandiseka	0	1	2	3	4
Sindingathe kupanga zinthu ene a zaka zanga amachita	0	1	2	3	4
Zikuma ndi a na posewera	0	1	2	3	4

ZOKHUDZANA NDI KULU *(mavuto ndi …)*	**Silinak halepo vuto**	**Sivuto**	**Nthawi zina ndi vuto**	**Ndi vuto Kawiri-kawiri**	**Pafupi-fupi nthawi zonse**
1. Zikumano mvetsera mu kalasi	0	1	2	3	4
2. Ndikumaiwala zinthu	0	1	2	3	4
3. Zikumandivuta kuchita za ku sukulu	0	1	2	3	4
4. Ndimajomba ku sukuku chifukwa chodwala	0	1	2	3	4
5. Ndimajomba ku sukulu chifukwa ndimapita kuchipatala	0	1	2	3	4

Figure 4.12 Mapi Trust approved Chichewa (Malawi) PedsQL ™ 4.0 Teen report



ID#_____

Date:_____

PedsQL™

Ndondomeko ya Ubwino wa moyo wa mwana

Version 4.0

LIPOTI LA ACHINYAMATA (Zaka **13-18**)

MALANGIZO

Pa tsamba lotsatirali pali mndandanda wa zinthu zimen[illegible]hoza kukha[illegible]uto kwa iwe.

Chonde tiuze kuti **vutoli lakhala lalikulu [illegible]anji [illegible]**iwe pa [illegible]**OZI** wapitawo polemba mozungulira yankho:

0 ngati **si [illegible]ukha[illegible] vuto**
1 nga[illegible]**si [illegible]ito**
2 ngat[illegible]**wi zina** lin[illegible]ito
3 ngati li[illegible]ala vuto **nt[illegible] zambiri**
[illegible]ati lili v[illegible]**afupi-fupi [illegible]awi zonse**

Palibe mayankho [illegible]ndola k[illegible]a osalo[illegible]a
Ngati sukumvets[illegible]nso, ch[illegible]a[illegible]sa kuti[illegible]andizidwe.

*Mu **mwezi UMODZI** wapitawu, kodi **vuto** ili lakhala lalikulu bwanji kwa iwe …*

ZOKHUDZANA NDI UMOYO WANGA KOMANSO ZINTHU ZIMENE NDIMACHITA ***(mavuto ndi …)***	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Nthawi zambiri**	**Pafupi-fupi nthawi zonse**
1. Zikumandivuta kuyenda mopitilira mabuloku atatu a sukulu	0	1	2	3	4
2. Zikumandivuta kuthamanga	0	1	2	3	4
3. Zikumandivuta kuchita masewero olimbitsa thupi	0	1	2	3	4
4. Zikumandivuta kunyamula chinthu cholemera	0	1	2	3	4
5. Zikumandivuta kusamba m'thupi	0	1	2	3	4
6. Zikumandivuta kugwira ntchito za pakhomo	0	1	2	3	4
7. Ndikumamva kuwawa	0	1	2	3	4
8. Ndikumafooka	0	1	2	3	4

ZOKHUDZANA NDI M'MENE NDIKUMAMVERA ***(mavuto ndi …)***	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Nthawi zambiri**	**Pafupi-fupi nthawi zonse**
1. Ndikumaopa kapena ndikumakhala ndi mantha	0	1	2	3	4
2. Ndikumamva chisoni kapena ndikumadandaula	0	1	2	3	4
3. Ndikumakwiya	0	1	2	3	4
4. Ndikumavutika kugona	0	1	2	3	4
5. Ndikumadandaula ndi chimene chingandichitikire	0	1	2	3	4

M'MENE NDIMAKHALALIRA NDI ENA ***(mavuto ndi …)***	**Silinakhalepo**	**Sivuto**	**Nthawi zina**	**Kawiri-kawiri**	**Pafupi-fupi nthawi zonse**
1. Zikumandivuta kucheza ndi achinyamata ena	0	1	2	3	4
2. Achinyamata ena sam[illegible]khala a[illegible]nga	0	1	2	3	4
3. Achinyamata ena a[illegible]diseka	0	1	2	3	4
4. Sindingathe kupan[illegible]zinthu [illegible]chinyam[illegible]a amach[illegible]	0	1	2	3	4
5. [illegible]umandivuta kukhala nd[illegible]chinyam[illegible]a [illegible]na posewera	0	1	2	3	4

[illegible]KHUDZA[illegible]U ***(m[illegible]ndi …)***	**Silinak halepo**	**Sivuto**	**Nthawi zina**	**Ndi vuto**	**Pafupi-fupi nthawi zonse**
1. Zikumandivuta kumvet[illegible]a mu kalasi	0	1	2	3	4
2. Ndikumaiwala zinthu[illegible]	0	1	2	3	4
3. Zikumar[illegible]a za ku sukulu	0	1	2	3	4
4. Ndimajomba ku sukuku chifukwa chodwala	0	1	2	3	4
5. Ndimajomba ku sukulu chifukwa ndimapita kuchipatala	0	1	2	3	4

4.6 Discussion

Ultimately, acceptable versions of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report were developed in the Chichewa language. The versions established acceptable linguistic and conceptual equivalence with the original UK and US English versions. However, there were a few issues that needed to be resolved during the adaptation process.

The concept of health was very difficult to translate as three terms are used interchangeably in everyday usage in Chichewa. This was problematic as these terms relate to somewhat different concepts. This confluence of health and life concepts is also common to other Bantu languages, possibly due to different views or models of health existing in these settings. (269) In Zulu, "health" and "life" both translate to "Mpilo", and in Kinyarwanda "ubuzima" can be used for both. One study in Kenya, also found that this was the case when they tried to disentangle health. (194) The conclusion is that while the children generally did seem to understand health, the concepts of health and life in these Bantu cultures need further exploration through focus group discussions to look for better alternative local language words.

The validity of including combined concepts within two of the EQ-5D dimensions has been questioned previously. (270, 271) However, as the items are entrenched in the measure which has been universally applied (272), no changes were made to the constructs included in the 'having pain or discomfort' and 'feeling worried, sad or unhappy' dimensions. To emphasize that all two or three components of these dimensions were highlighted and given equal prominence, it was helpful to add the qualifier at each component. For example, the severity level 'I have a little bit of pain or discomfort' could imply that 'a little bit' applies only to pain and not to both pain and discomfort. This was altered to read, "I have a little bit of pain or a

little bit of discomfort". This solution is not unique and has been incorporated into other translations, such as Krio self-complete EQ-5D-3L and the Arabic for Jordan EQ-5D-Y versions. (Personal communication, VMC, 2021) Future cross-cultural adaptation in other contexts should explore the need to emphasize all components in combined dimensions to maintain the full meaning.

The Chichewa EQ-5D-Y-5L was found to be difficult to self-complete. Younger children, in particular, needed some interviewer assistance to accurately complete the questionnaire. For example, younger children tended to tick every severity level response option instead of choosing one unless the interviewer assisted. Consequently, users of the Chichewa EQ-5D-Y-5L be advised to consider using the interviewer administered version for children <12 years.

Although the Chichewa versions established both linguistic and conceptual equivalence with the English version, caution is advised when administering these versions. While most participants gave specific examples that demonstrated their understanding of the severity levels and dimensions, important lessons were learnt from the in-depth interviews. A few participants struggled to understand some of the dimensions. This also highlighted contextual differences in terms of how people understand translated dimensions of HRQoL instruments. For the EQ-5D-Y and EQ-5D-Y-5L this was particularly the case regarding 'looking after oneself' dimension. In HICs, washing is simply the action of washing the body. In LMICs, participants may interpret this as whether there is the availability of water and soap, rather than whether they have the *ability* to wash and dress themselves. This type of misinterpretation was observed during the exercise in Sierra Leone to translate the EQ-5D-Y into Krio. (Personal communication, VMC, 2021) The inclusion of needing to getting water, i.e., from a well or river in the activity of self-care was highlighted in that translated version. It was recommended

that the EuroQol Group should provide further guidance on environmental factors in the translation protocol. The omission of such clarification as water availability and 'mobility' issues might affect item equivalence particularly in some low income countries. (273) These results endorse that recommendation.

The term 'Having pain or discomfort' has proven somewhat problematic in the past. These terms are meant to refer to physical pain or discomfort, but participants often interpret them in relation to emotional or mental pain or discomfort. In this study, some of the Malawian children had interpreted these terms in relation to emotional or mental pain instead of physical pain. This finding was curious because '*m'thupi'*(body) was added in the descriptor to establish conceptual equivalence with the concept about bodily pain. It is possible that the translated word for discomfort (*kusamva bwino*) could have led others to an emotional as opposed to uncomfortable physical sensation. However, this was most appropriate word that could be used for discomfort. Generally, the 'having pain or discomfort' dimension has proved difficult to translate in the past. In the Shona translation of the adult EQ-5D-3L, it was reported that such expression as "I cannot even stay in one place" was considered for translation. In addition, discomfort was not regarded as a sensation separate from pain, (274) as such the phrase "kurwadziwa" for pain was used. Future users of the youth Chichewa EQ-5D-Y versions might need to explain to participants that the discomfort mentioned here refers to physical sensation. Another option would be to include 'physical' in the descriptor, so that the dimension reads 'physical pain or physical discomfort'.

Overall, the findings show the importance of including the cognitive interviews as opposed to only a translation to establish linguistic equivalence. The issue of the '*kusamba'* meaning menstrual cycle was very fundamental and this would not have been identified with a

translation alone. Had it not been for the cognitive interviews a wrong meaning would have been captured leading to wrong results and conclusions.

The ranking exercise was instrumental in understanding the qualifiers for the five different severity levels which the cognitive interviews alone (self-completion and in-depth) were unable to identify. It is very likely that had it not been for the card ranking exercise, coming up with accurate qualifiers to establish linguistic equivalence and correct hierarchical ranking would have been adequately established. As noted elsewhere, severity level inversion in the questionnaire is a threat to the validity of valuation sets. (275) If the severity level is either not properly understood, or does not accurately distinguish between levels, it could lead to preference inversion in utility valuation exercises. These findings highly recommend the use of the ranking exercise when adapting the EQ-5D-Y-5L cross-culturally.

It can be said that the translation process for the EQ-5D-Y-5L was found to be more thorough in comparison to the EQ-5D-Y. While potential for severity inversion might not be highly likely in a three compared to a five-response version, there was inversion that existed between levels two and three for the EQ-5D-Y-5L. This was only resolved with the card ranking exercise. It is possible that inversion might exist within the three response EQ-5D-Y version but remain to be resolved due the current standard translation protocol without ranking exercise. The same could be said of the PedsQL ™ 4.0 which does not include the ranking exercise for its five-severity level. In addition, the EQ-5D-Y and EQ-5D-Y-5L have a more rigorous translation process with the inclusion of the VMC reviewer involved each step of the way. The inclusion of the developer's reviewer is not part of the PedsQL ™ 4.0 translation process. While due diligence was taken to come up with an accurate PedsQL ™ 4.0 translation, the process could have benefitted from the developer's reviewer of the process.

Finally, as regards possible inconsistency between the EQ VAS score and that the utility score associated with the descriptive system health state, this should not be overly-interpreted. The EQ VAS is designed to provide information that complements the EQ-5D profile. (276) However, although EQ VAS scores are systematically related to health profiles they have been found to partially explain this relationship in regression analyses. (277) Further, VASs, in general, are considered to be unreliable (single item) and this is an established fact in psychometrics. There is established psychological evidence of response options being best at 5+/-2 options. (278) Additionally, people may be able to use the EQ VAS to essentially make a ranking (i.e., in the right order) but not able to use the minute gradings on the scale. So, while not wanting to read much into the EQ VAS because of these reasons, there are still lessons that can be drawn from this. There were two out of three participants who had EQ VAS scores that seemed inconsistent with their reported descriptive system health state which might indicate that there may be aspects of health lying beyond this. Future studies need to explore which other aspects of 'health' may be important to children, beyond the five EQ-5D-Y-5L dimensions. (277) Within the local context, this was evaluated in chapter 7.

A few things need to be highlighted that show the additional relevance of this research. Firstly, the stipulated samples were adequate to establish the comprehensibility of the PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report. The sample size for establishing comprehensibility of the EQ-5D-Y-5L depended on establishing severity level inversion. The EuroQol Group need to give a general guide that the sample size for the EQ-5D-Y-5L translation is dependent on establishing severity level inversion.

The official name for the different EQ-5D youth versions should be quickly addressed by the EuroQol Group. Currently, the beta EQ-5D-Y-5L version has been labelled differently by different users. (259, 260, 279, 280) One suggestion would be to change label of the three level from EQ-5D-Y to EQ-5D-Y-3L. Changing the label to EQ-5D-Y-3L would help in making a clear distinction between the three level EQ-5D-Y (Y-3L) and five level EQ-5D-Y-5L (Y-5L), which is proving difficult at the moment. Changing the label for the latter might not be straight forward since it is a copyrighted document. This might not be a problem for the EQ-5D-Y-5L since this it is still undergoing piloting before finalized as an official version.

The main limitation of the adaptation of the Chichewa versions of the measures is in the sample of participants. (263) While this was attainable during the first round of testing for the EQ-5D-Y-5L but was not the case in subsequent rounds of re-testing. The cognitive interviews for PedsQL ™ 4.0, was also only among healthy, instead of sick, children due to challenges to recruit sick children from the referral hospital at the time. That said, it is doubtful that sick children would understand the language or concepts any differently. The second limitation is that this sample was also made up of school going children who could read and write, and therefore generally able to understand the concepts. It is possible that this might not be the case in a general population comprising of school and non-school going children. It should also be highlighted that the forward translators were provided with the adult EQ-5D-3L Chichewa version which might have influenced translation of words common in both versions. However, this was at best a guide only and the translators made independent decisions and discussed on the translated words to retain in the EQ-5D-Y. As discussed above while trying to maintain consistency between the adult EQ-5D-3L Chichewa version and the EQ-5D-Y there were differences in the translated words used. Lastly, while it may be true that the geographic variation in Chichewa (e.g., by districts) could confuse meanings, it is practically impossible

to do translation in every dialect spoken. To a large extent, the local translators ensured that the translated Chichewa has linguistic and conceptual equivalent to be understood by most Malawians. There was diversity in local translators including the parts of Malawi they come from as well as where they had lived to address heterogeneity concerns.

There are strengths of this adaptation study. Firstly, this was the first study to pilot cross-cultural adaptation of the beta EQ-5D-Y-5L using the card ranking exercise which adequately addresses severity inversion. This part of the work has been recommended to the EuroQol Group for EQ-5D-Y-5L translation protocol. (281, 282) Secondly, the age of the participants in this study was deliberately skewed to younger children (8-10 years) in order to establish comprehensibility of the measure in this population. This was because the Chichewa EQ-5D-Y version, as well as the first two rounds of the EQ-5D-Y-5L interviews, had established comprehensibility of the measure in older children and adolescents (six participants aged 11-14 years [three each for the EQ-5D-Y and the EQ-5D-Y-5L]). The focus for the subsequent EQ-5D-Y-5L was therefore mostly in the younger population.

It should be noted that the intention of this study was not to resolve all the issues that occur with EQ-5D but only to resolve those that occur due to the local context. Therefore, even at the end of the adaptation process there will still be some issues that are general to EQ-5D in any language e.g., the double barrelled nature of "pain and discomfort" and "washing or dressing", "feeling worried, sad or unhappy" etc. This also include the limitation of five dimensions of the EQ-5D instruments. Addressing these issues is beyond the scope of this thesis although some of these outstanding issues within the local context are also evaluated in chapter 7.

In general, the adapted Chichewa versions were relatively comprehensible although the younger children needed some interviewer assistance for self-completion. This finding highlights the need for interviewer administered rather than the self-complete versions of the EQ-5D-Y, EQ-5D-Y-5L and PedsQL ™ 4.0 especially among children ≤12 years.

4.7 Conclusion

As has been noted elsewhere,(283) the importance of rigorous translation and cross-cultural adaptation of HRQoL cannot be overemphasised. The use of a poorly translated instrument may well invalidate otherwise carefully planned research. The EQ-5D-Y, EQ-5D-Y-5L, and PedsQL ™ 4.0 Chichewa versions were generally comprehensible to children in Malawi. The findings highlight the need for interviewer administered version rather than the self-complete version in children aged 8-10 years. Regarding concepts, the dimensions of 'mobility', 'doing usual activities' and 'feeling worried, sad or unhappy' clearly established conceptual equivalence and were understood by participants. There were mixed results with the 'looking after oneself' and 'having pain or discomfort' dimensions of the EQ-5D-Y and EQ-5D-Y-5L. The EuroQol Group should consider including specific examples in the translation protocol in order to make this concept easier to understand in LMICs or similar settings. Lastly, the ranking exercise was instrumental in identifying translation problems with the severity qualifiers for the different levels. The ranking exercise are fundamental part of the EQ-5D-Y-5L cross-cultural adaptation.

In conclusion, the EQ-5D-Y, EQ-5D-Y-5L, PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report were comprehensible for use in Malawi although with limitation to some outstanding issues highlighted above. The issues concerning "looking after oneself" and "having pain or discomfort" may not just relate to Malawi but it is not possible to make

unilateral change to these versions. These have, however, been communicated to the EuroQol Group.

The cross-cultural adaptation process is the first step in the cross-cultural validation process. The next steps are to test additional quantitative psychometric properties of the instruments in the Malawian setting. The following chapter will psychometrically evaluate the PedsQL ™ 4.0 Child self-report and the PedsQL ™ 4.0 Teen self-report. Thereafter, the EQ-5D-Y and EQ-5D-Y-5L are evaluated.

The findings based on the material in the chapter have been disseminated at several international conferences. A paper has been accepted for publication based on the material in the chapter. (284) A second paper was published based on some of the material from this chapter. (265).

5 Chapter 5- Psychometric evaluation of PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report. Chichewa versions

With:

Ngwira, L.G.; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S
(manuscript drafted)

5.1 Overview of chapter

In chapter 4, the EQ-5D-Y, EQ-5D-Y-5L, PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report into Chichewa, were cross-culturally adapted into Chichewa. In this chapter, the PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report will be psychometrically evaluated (validated) to establish whether they are valid instruments in Malawi. This is to inform whether they can be used as validating measures in the psychometric evaluation of EQ-5D-Y in subsequent analyses.

The PedsQL ™ 4.0 child self-report and PedsQL ™ 4.0 teen self-report have demonstrated psychometric robustness among healthy and paediatric cancer patients in California, USA. (267, 285) The PedsQL ™ 4.0 has subsequently been translated into more than 60 languages and demonstrated reliability, validity, sensitivity and responsiveness among healthy and chronic conditions. (286) However, although the PedsQL™ 4.0 is one of the most used generic instruments to assess HRQoL, it's use has not been established in sSA let alone Malawi.

5.2 Introduction

Over the last three decades there has been an increase in development of childhood HRQoL due to the use of HRQoL in clinical trials, practice and resource allocation decisions. (168, 287) However, the development, use and validation of such in sSA as discussed in chapter 2 has been limited. (199) At the same time, chapter 2 highlighted the potential for cross-cultural validation of the childhood health instruments in the sSA region. This was evident in chapter 4 where four HRQoL instruments including the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report were linguistically and conceptually adapted into Chichewa (Malawi). The PedsQL™ 4.0 is a generic non-preference based HRQoL measure that assesses HRQoL across four core dimensions (physical, emotional, social and school) in healthy and ill children. (285) Measuring HRQoL in childhood populations can assist in exploring the burden of disease, as well as in the evaluation of health care interventions. (288) The PedsQL™ 4.0 self-report was developed as "a generic pediatric HRQoL instrument to be utilized across diverse pediatric populations, including healthy children and children with acute and chronic health conditions". (267) The PedsQL™ 4.0 has demonstrated feasibility, reliability, validity, sensitivity, and responsiveness in healthy and ill children and adolescents. (267, 285, 286) Although the PedsQL™ 4.0 has been translated in over 60 languages, (286) its use in sSA region has been limited. (199)

As outlined in chapter 3, psychometric assessment ensures that an instrument's ability to quantify or measure how people self-describes health. The expectation is that any good instrument should demonstrate robust psychometric properties. (83, 144)

5.3 Aims and objectives

The aim of the study described in this chapter was to evaluate the psychometric performance of the Chichewa versions of the PedsQL™ 4.0 child self-report (for ages 8-12 years) and PedsQL™ 4.0 teen report (for ages 13-18 years). Specifically, to investigate: i) item level performance (missing data, response distribution, item redundancy); ii) scale reliability (items in the scale are homogenous); and iii) scale validity (scale measuring what it claims to measure).

5.4 Methods

5.4.1 Participants and recruitment

The participants and recruitment procedure are the same as outlined in chapter 4 section 4.1. All the data here were collected in same way and at the same time as the EQ-5D-Y and EQ-5D-Y-5L data reported in the next chapter 6. (Appendix 16-18 has the information leaflets and consent/assent form)

A convenience sample of healthy children was conveniently sampled from Primary and Secondary schools. Participants self-completed the Chichewa versions of PedsQL™ 4.0 child (ages 8-12 years) and teen (ages 13-18 years) versions between January 2020 to March 2020. Due to the COVID-19 pandemic data could not be collected beyond this point. Data on participants' age, school grade and gender were also recorded.

Acutely ill participants were recruited from the QECH, and the chronically ill participants were recruited from primary health clinics at the end of their clinical appointment. The children completed either the PedsQL™ 4.0 child self-report and teen self-report. In addition, data collected included sociodemographic (age, grade, gender) and clinical condition (acute or

chronic). Parents socio-demographic data were however not collected in either the healthy or sick participants enrolled in the study.

5.4.2 The PedsQL™ instrument

These were presented in the previous chapter 4. The PedsQL™ 4.0 scores were calculated according to the PedsQL™ 4.0 procedure outlined in chapter 4.4.3.1. This scoring procedure is according to the published instructions (266) to generate PedsQL™ 4.0 total scale scores, two health summary scores (psychosocial and physical) and four sub-scores (physical, emotional, social and school).

5.4.3 Psychometric analyses

Psychometric analyses were conducted using gold standard techniques and based on the guidelines outlined in detail in chapter 3. Item performance, reliability (internal consistency) and validity (convergent, discriminant and known-groups) were evaluated. The analyses are briefly outlined below.

5.4.3.1 Item performance

As discussed in detail in chapter 3.9, there are four aspects of item performance that were evaluated in this thesis: missing data, MEF, AEF and item redundancy. The criteria used to investigate each of these aspects of item performance are outlined briefly below.

Missing data: A criterion of 5% was used to evaluate missing data. (231) Items that had ≥ 5% missing data were flagged as potentially problematic.

Maximum Endorsement Frequency (MEF): This is the endorsement at the extremes of the response scale. Items are considered problematic if they have >80% endorsement at either end of the scale. (237) This indicates that responses are concentrated in the top or bottom end of the scale (floor or ceiling effect).

Aggregate adjacent endorsement frequency (AEF): AEF is the extent to which adjacent response options in the middle of the response scale sum to at least a specified minimum. (253) Items are considered problematic if any two or more adjacent response options summed to <10%. (253) This informs whether there are some response options that are not being used.

Item redundancy: This reflects the extent to which each item within the scale makes a unique contribution and is not over-lapping with other items in the scale. It is assessed by evaluating the inter-item correlation between all items in the scale. In this study, items with inter-item correlations > 0.75 were considered problematic. (237)

5.4.3.2 Reliability

Reliability indicates the extent to which a scale is free from random error and therefore able to produce consistent and reproducible results. (237) There are several types of reliability which can be assessed psychometrically but as discussed in chapter 3, it was only possible to evaluate internal consistency in this study.

Several statistics were examined to evaluate internal consistency including Cronbach's alpha, corrected item-total correlation, and Cronbach alpha when item is deleted. A criterion of Cronbach's alpha ≥0.70 (146) indicates acceptable internal consistency and the same was used in this study. The corrected item-total correlation was expected to be between 0.25 – 0.80,

(237) to indicate that it was not contributing too much or too less to the scale. The Cronbach's alpha if item is deleted was assessed by looking at the deviation between the standardized alpha for the overall scale and the Cronbach's alpha if an item is deleted. It was hypothesized that this difference would be small.

5.4.3.3 Validity

5.4.3.3.1 Convergent validity

This was evaluated by comparing the intra correlation coefficients for the four PedsQL™ 4.0 sub-scales: Physical Functioning, Emotional Functioning, Social Functioning and School Functioning. The literature has classified the strength of correlation as: negligible (0.00—0.09), weak (0.10—0.39), moderate (0.40—0.69), strong (0.70—0.89) and very strong (0.90—1.00) (289). In was hypothesized that the intra correlation coefficient would be ≥0.4 (moderate to strong) as evidence of convergent validity.

5.4.3.3.2 Discriminant validity

It was hypothesized that there was going to be no association between PedsQL™ 4.0 scores with: gender (t-test), age (correlations), or grade (ANOVA). In detail, it was hypothesized that there would be no association at 5% significance level between PedsQL™ 4.0 self-report scores and gender. There should be no difference between gender in self-reported health because these are generic instruments developed to assess health for either gender. Additionally, these instruments measure perception of health and it is expected that this is viewed in the same way irrespective of gender differences. It was anticipated that there would be no association (correlation <0.2) between PedsQL™ 4.0 self-report scores and age. A one-way ANOVA test was performed which compared the mean PedsQL™ 4.0 self-report scores per specified grades.

Any evidence of significance at the 5% significance level would indicate that grade influenced PedsQL™ 4.0 self-report score outcome.

5.4.3.3.3 Item convergent and discriminant validity

To quantify the difference between the within (intra) scale and between (inter) scale correlations, the approach developed by Henseler et al 2014 was adapted. (255) The Henseler approach is a modification of the MTMM and calculated using Microsoft excel as follows: 'the average of the correlations of indicators across constructs measuring different phenomena, relative to the average of the correlations of indicators within the same construct'. (255) This is expressed as follows (from Henseler et al):

$$\underbrace{\mathrm{HTMT}_{ij} = \frac{1}{K_i K_j}\sum_{g=1}^{K_i}\sum_{h=1}^{K_j} r_{ig,jh}}_{\substack{\text{average}\\ \text{heterotrait–}\\ \text{heteromethod}}} \div \underbrace{\left(\frac{2}{K_i(K_i-1)}\cdot\sum_{g=1}^{K_i-1}\sum_{h=g+1}^{K_i} r_{ig,ih}\cdot\frac{2}{K_j(K_j-1)}\cdot\sum_{g=1}^{K_j-1}\sum_{h=g+1}^{K_j} r_{jg,jh}\right)^{\frac{1}{2}}}_{\substack{\text{geometric mean of the average monotrait–heteromethod}\\ \text{correlation of construct } \xi_i \text{ and the average}\\ \text{monotrait–heteromethod correlation of construct } \xi_j}} . \quad [1]$$

Formula [1] showing the modified MTMM with hetero-methods

where *i* and *j* are the constructs, *Ki* and *Kj* are reflective indicators of constructs *i* and *j* respectively, *rig, jh* is the correlation matrix of items within constructs *i* and *j*. This would mean that HTMT*ij* is the average of correlation coefficient matrix of the constructs *i* and *j* divided by the geometric mean of the correlation coefficient matrix of the constructs *i* and *j*.

Since there was only one method (mono-method) to collect data and to generate correlation matrix, and the PedsQL™ 4.0 being multi-construct (multi-trait), the multitrait-monomethod ratio was derived as follows:

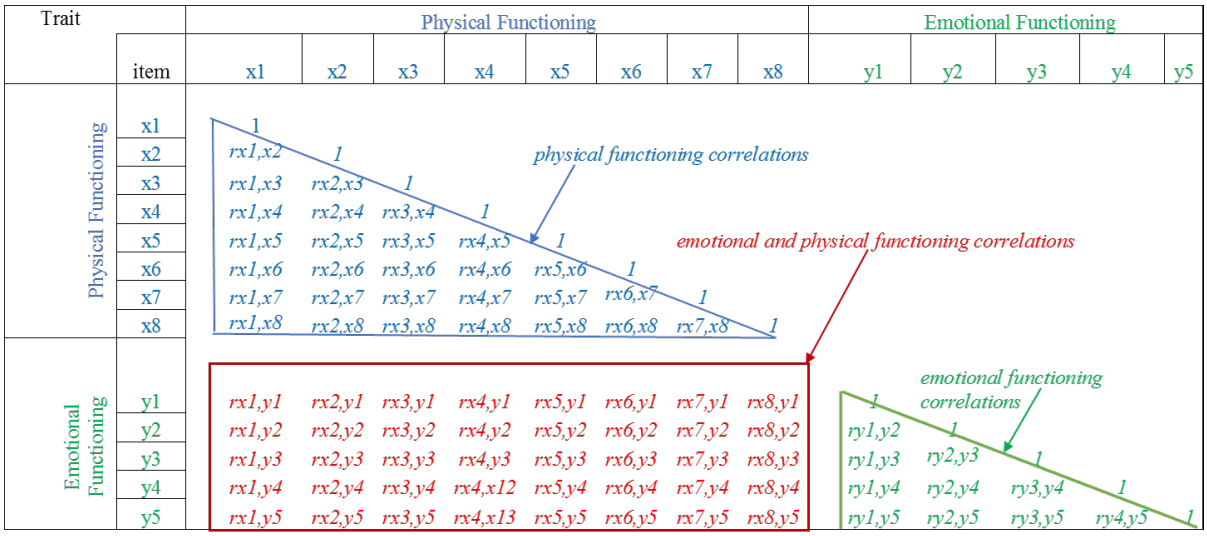
$$\frac{\textit{average multitrait} - \textit{monomethod correlation for construct ij}}{\sqrt{(\textit{average monotrait} - \textit{monomethod correlation for construct i}) \times (\textit{average monotrait} - \textit{monomethod correlation for construct j})}} \quad [2]$$

Formula [2] showing a modified MTMM leaving out the methods factor

where *i* and *j* are the constructs.

This modified MTMM ratio for the PedsQL™ 4.0 as represented in formula [2] above was derived for PedsQL™ 4.0 as shown in Figure 5.1 and in formula [3] below.

Figure 5.1 Illustrative example of the PedsQL™ 4.0 reduced MTMM matrix (adapted from Henseler et al)



The modified MTMM ratios for the PedsQL™ 4.0 scale were calculated as follows:

$$\frac{average\ correlation\ matrix\ for\ emotional\ and\ physical\ functioning}{[(average\ correlation\ for\ physical\ functioning)x(average\ correlation\ for\ emotional\ functioning)]} \quad [3]$$

Formula [3] showing a modified MTMM for PedsQL™ 4.0scale

where *x1…x8, y1…y5* are items within Physical Functioning and Emotional Functioning respectively, *rx1,x2…rx7,x8* and *ry1,y2…ry4,y5* are intra correlation coefficients within Physical Functioning and Emotional Functioning respectively, and *rx1,y1…rx8,y5* is the inter correlation coefficients of the items within Physical Functioning and Emotional Functioning sub-scales.

If the ratio is below 0.85 (conservative approach) the scale is said to have met the criteria for discriminant validity.(255) It was hypothesized that all the between sub-scale ratios would be below the 0.85 threshold to support evidence for discriminant validity.

5.4.3.3.4 Known-groups validity

It was hypothesized, based on previous findings, (267) that all PedsQL™ 4.0 sub-scale and Total scale Scores would be worse for sick compared with healthy children.

This was investigated using a t-test to establish whether there was a significant difference between the two groups. Additionally, to assess the magnitude of the difference, an effect size was calculated using the mean difference divided by standard deviation of the reference (healthy) group as follows:

$$\frac{mean\ of\ healthy\ children - mean\ of\ sick\ children}{standard\ deviation\ of\ healthy\ children}$$

Cohen's criterion was used to determine the magnitude of effect sizes (<0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, and >0.8 large). (150, 231)

Each analysis described above was conducted separately for the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report. Further, all the analyses were conducted on the whole sample and then separately for the healthy and sick samples (i.e. healthy, chronic and acute groups).

5.5 Results

5.5.1 Participant and recruitment

In total, 289 participants completed the PedsQL™ 4.0 aged 8-17 years (mean 13.6, median 14). (Table 5.1) Of these 191 completed the PedsQL™ 4.0 teen self-report, and 98 completed the PedsQL™ 4.0 child self-report.

Table 5.1 Participant characteristics for the PedsQL™ 4.0 child self-report and teen self-report.

Characteristic	N (%)	Child self-report (age 8-12yrs)	Teen self-report (age 13-17yrs)
Participants	289	98	191
Gender*			
Male	121 (44%)	39	82
Female	153 (56%)	51	102
Health condition			
healthy	95 (33%)	12	83
acute	155 (54%)	85	70
chronic	39 (13%)	1	38
Grade#			
1-5	71 (25%)	53	18
6-8	97 (35%)	40	57
9-12	111 (40%)	0	111

*missing data: 15 (child self-report=8, teen self-report=7); #missing data: 10 (5 in each age groups)

5.5.2 Item performance

Table 5.2 contain item performance for both the PedsQL™ 4.0 child and PedsQL™ 4.0 teen self-report versions. There was a large number (15 out of 23) of item-level missing data (>5%) amongst the younger participants (age 8-12 years) who completed the PedsQL™ 4.0 child version. All the four PedsQL™ 4.0 sub-scales had some missing data. Amongst the adolescents (age 13-18 years) who completed the PedsQL™ 4.0 teen self-report, missing data was less of a problem (1 of 23 items).

All the items on PedsQL™ 4.0 child self-report and teen self-report versions had acceptable maximum endorsement frequency.

Most items (20/23), however, failed the aggregate adjacent endorsement frequency. Only three items each had no aggregate adjacent endorsement frequency issues for both the PedsQL™ 4.0 child self-report and teen self-report.

For both the PedsQL™ 4.0 child self-report and teen self-report there were no pairs of items with correlations > 0.75.

Table 5.2 Item-level analysis for all the data by age groups for the PedsQL™ 4.0 Child self-report and PedsQL™ 4.0 teen self-report

PedsQL™ 4.0			Missing Data		Maximum endorsement frequency*		Adjacent aggregate endorsement frequency		Item redundancy*	
Sub-scales	**Items**		**Child**	**Teen**			**Child**	**Teen**		
Physical Functioning	1.1	To walk more than one block	x				x	x		
	1.2	To run					x	x		
	1.3	Sports activity or exercise						x		
	1.4	To lift something heavy	x							
	1.5	To take a bath or shower by myself	x				x	x		
	1.6	To do chores around the house					x	x		
	1.7	Hurt or ache					x	x		
	1.8	Low energy	x				x	x		
Emotional Functioning	2.1	Afraid or scared					x			
	2.2	Sad or blue					x	x		
	2.3	Angry	x				x	x		
	2.4	Trouble sleeping	x				x	x		
	2.5	Worry about what will happen to me	x				x	x		
Social Functioning	3.1	Trouble getting along with other children	x				x	x		
	3.2	Other children do not want to be my friend	x				x	x		
	3.3	Other children tease me	x				x	x		
	3.4	Cannot do things that other children can do	x				x	x		
	3.5	Keep up when I play with other children	x					x		
School Functioning	4.1	Pay attention in class					x	x		
	4.2	Forget things	x	x			x			
	4.3	Trouble keeping up with my schoolwork	x				x	x		
	4.4	Miss school because of not feeling well	x				x	x		
	4.5	Miss school to go to doctor or hospital					x	x		

* no item failed for maximum endorsement frequency and item redundancy; x indicates a fail on a criteria

Table 5.3 presents the item-level analysis based on the health status of the participants. There was a considerable amount of missing item-level data amongst the healthy children, especially in the Physical Functioning sub-scale. Missing data was not a problem among the healthy participants but was more an issue in the acutely ill. Among adolescent participants, there were higher levels of missing data in the chronically than in acutely ill, but not in the healthy participants.

Although the evidence for items with <10% summed response within the middle of response was mixed it was more of an issue in acute children. Similarly, this was absent in the healthy children in all PedsQL™ 4.0 sub-scales for the child self-report, but was an issue in all sub-scales among the acute ill: 6 out of 8 items in the Physical Functioning, 5 out of 5 items in the Emotional Functioning, 2 out of 5 items in the Social Functioning, and 4 out of 5 items in the School Functioning. The PedsQL™ 4.0 teen self-report on the other hand reported <10% summed response in all the sub-scales of health status groups. Much of the aggregate adjacent endorsement frequency were among the acutely ill in the Physical Functioning sub-scale (7 out of 8 items) and Emotional Functioning (3 out 5 items). In only 7 out of 23 items, and 4 out of the 23 items was there evidence of <10% summed response within the scale among the healthy and chronically ill respectively.

For this analysis based on health conditions, item redundancy was mostly evident for PedsQL™ 4.0 child self-report among the healthy children in the Physical Functioning sub-scale (items: to run/to walk more than one block; to lift something heavy/ to walk more than one block; to lift something heavy / to run; to take a bath or shower by myself/ to walk more than one block; to do chores around the house / to take a bath or shower by myself; hurt or

ache/ to walk more than one block; hurt or ache/ to run; hurt or ache/ to lift something heavy). Additionally, item redundancy was observed in one item amongst the acutely ill children in School Functioning sub-scale (items miss school to go to a doctor/miss school because of not feeling well). Items with correlation >0.75 for PedsQL™ 4.0 teen self-report were high among the chronically ill children in the Physical Functioning (items: sports or activity or exercise/ to run; to lift something heavy / to walk more than one block; to take a bath or shower by myself / sports or activity or exercise; to do chores around the house/to take a bath or shower by myself; hurt or ache with all other items except to lift something heavy and low energy; and item low energy with all the other items except to walk more than one block), Social Functioning (items: other children don't want to be my friend/trouble getting along with other children; other children tease me/ other children don't want to be my friend; cannot do things that other children can do/trouble getting along with other children; cannot do things that other children can do/ with all other items except keep up when I play with other children; and keep up when I play with other children with all other items) and School Functioning (items: trouble keeping up with my schoolwork/forget things; miss school because of not feeling well/forget things; miss school to go to doctor or hospital/pay attention in class; and miss school to go to doctor or hospital / miss school because of not feeling well) sub-scales. There was no correlation > 0.75 in all the four sub-scales among healthy population, and only one item that did not meet criteria for item redundancy in both the Physical and School Functioning sub-scales in the acutely ill children.

Table 5.3 PedsQL™ 4.0 child self-report (8-12 years) and PedsQL™ 4.0 teen self-report (13-18 years) item level analysis based on health conditions

PedsQL™ 4.0*			Missing Data					Maximum endorsement frequency#					Adjacent aggregate endorsement frequency					Item redundancy				
			Child		Teen			Child		Teen			Child		Teen			Child		Teen		
Sub-scales	**Items**		healthy	acute	healthy	chronic	acute	healthy	acute	healthy	chronic	acute	healthy	acute	healthy	chronic	acute	healthy	acute	healthy	chronic	acute
Physical Functioning	1.1	To walk more than one block	x			x	x							x	x		x					
	1.2	To run	x			x								x			x	x				
	1.3	Sports activity or exercise				x									x		x				x	x
	1.4	To lift something heavy	x	x		x											x	x			x	
	1.5	To take a bath or shower by myself	x	x		x	x							x			x	x			x	
	1.6	To do chores around the house	x			x								x			x	x			x	
	1.7	Hurt or ache	x			x								x				x			x	
	1.8	Low energy	x	x		x								x			x				x	
Emotional Functioning	2.1	Afraid or scared				x	x							x			x					
	2.2	Sad or blue				x	x							x			x					
	2.3	Angry	x	x		x	x							x			x					
	2.4	Trouble sleeping	x			x	x							x				x			x	
	2.5	Worry about what will happen to me		x		x	x							x	x	x						
Social Functioning	3.1	Trouble getting along with other children		x		x										x	x					
	3.2	Other children do not want		x		x	x														x	

		to be my friend																				
	3.3	Other children tease me		x		x								x			x	x			x	
	3.4	Cannot do things that other children can do		x		x	x							x	x		x				x	
	3.5	Keep up when I play with other children		x			x								x	x					x	
School Functioning	4.1	Pay attention in class				x								x								
	4.2	Forget things		x		x								x	x							
	4.3	Trouble keeping up with my schoolwork		x		x										x					x	
	4.4	Miss school because of not feeling well		x		x								x							x	
	4.5	Miss school to go to doctor or hospital		x		x								x	x			x	x		x	x

*n=1 with chronic health condition for child self-report; x indicates a fail on a criteria
no item failed for maximum endorsement frequency

5.5.3 Reliability

The findings of the reliability analysis and Cronbach's alpha coefficients are presented in Tables 5.4 and 5.5 for the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report respectively. All the PedsQL™ 4.0 child self-report sub-scales had Cronbach's alpha >0.80. A similarly high Cronbach's alpha >0.90 was observed for summary scores and the total scale score recommended for analysing patient scale scores. Most (of the corrected item-total correlations for each of the four sub-scales of the PedsQL™ 4.0 child self-report met the criterion. Two items (1.7 and 2.2) very slightly exceed the 0.8 criterion.

The PedsQL™ 4.0 teen self-report sub-scales demonstrated similar findings for group comparisons as well as summary scores and the total scale score. All the corrected item-total correlations for each of the four sub-scales were within the hypothesized acceptable range. There was similarly little deviation of the Cronbach alpha if the item deleted, range 0.947-0.950, to the standardized total scale score Cronbach's alpha of 0.951.

Table 5.4 PedsQL™ 4.0 child self-report (8-12 years) showing internal consistency reliability

Subscale (n), Items	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	change (standardized alpha- Cronbach alpha if item deleted)	Cronbach's Alpha Based on Standardized Items
Physical Functioning sub-scale (n=96, excluded 2)				0.919
1.1 To walk more than one block	0.772	0.905	0.014	
1.2 To run	0.652	0.914	0.005	
1.3 Sports activity or exercise	0.728	0.908	0.011	
1.4 To lift something heavy	0.715	0.909	0.010	
1.5 To take a bath or shower by myself	0.750	0.906	0.013	
1.6 To do chores around the house	0.721	0.909	0.010	
1.7 Hurt or ache	**0.806**	0.902	0.017	
1.8 Low energy	0.705	0.910	0.009	
Emotional Functioning sub-scale (n=94, excluded 4)				0.856
2.1 Afraid or scared	0.659	0.830	0.026	
2.2 Sad or blue	**0.809**	0.788	0.068	
2.3 Angry	0.633	0.836	0.020	
2.4 Trouble sleeping	0.616	0.841	0.015	
2.5 Worry about what will happen to me	0.644	0.834	0.022	
Social Functioning sub-scale (n=91, excluded 7)				0.865
3.1 Trouble getting along with other children	0.725	0.824	0.041	
3.2 Other children do not want to be my friend	0.619	0.851	0.014	
3.3 Other children tease me	0.687	0.833	0.032	
3.4 Cannot do things that other children can do	0.645	0.844	0.021	
3.5 Keep up when I play with other children	0.749	0.820	0.045	

School Functioning sub-scale (n=93, excluded 5)				0.836
4.1 Pay attention in class	0.609	0.811	0.025	
4.2 Forget things	0.540	0.830	0.006	
4.3 Trouble keeping up with my schoolwork	0.680	0.791	0.045	
4.4 Miss school because of not feeling well	0.659	0.798	0.038	
4.5 Miss school to go to doctor or hospital	0.706	0.783	0.053	
Psychosocial Health Summary Scale (n=91, excluded 7)				0.919
2.1 Afraid or scared	0.632	0.913	0.006	
2.2 Sad or blue	0.676	0.911	0.008	
2.3 Angry	0.480	0.918	0.001	
2.4 Trouble sleeping	0.558	0.915	0.004	
2.5 Worry about what will happen to me	0.662	0.912	0.007	
3.1 Trouble getting along with other children	0.668	0.912	0.007	
3.2 Other children do not want to be my friend	0.636	0.913	0.006	
3.3 Other children tease me	0.634	0.913	0.006	
3.4 Cannot do things that other children can do	0.698	0.911	0.008	
3.5 Keep up when I play with other children	0.719	0.910	0.009	
4.1 Pay attention in class	0.625	0.913	0.006	
4.2 Forget things	0.543	0.916	0.003	
4.3 Trouble keeping up with my schoolwork	0.666	0.912	0.007	
4.4 Miss school because of not feeling well	0.601	0.914	0.005	
4.5 Miss school to go to doctor or hospital	0.612	0.913	0.006	

Total Scale Score (n=90, excluded 8)				0.941
1.1 To walk more than one block	0.705	0.937	0.004	
1.2 To run	0.622	0.939	0.002	
1.3 Sports activity or exercise	0.601	0.939	0.002	
1.4 To lift something heavy	0.602	0.939	0.002	
1.5 To take a bath or shower by myself	0.672	0.938	0.003	
1.6 To do chores around the house	0.661	0.938	0.003	
1.7 Hurt or ache	0.738	0.937	0.004	
1.8 Low energy	0.610	0.939	0.002	
2.1 Afraid or scared	0.631	0.938	0.003	
2.2 Sad or blue	0.724	0.937	0.004	
2.3 Angry	0.536	0.940	0.001	
2.4 Trouble sleeping	0.585	0.939	0.002	
2.5 Worry about what will happen to me	0.669	0.938	0.003	
3.1 Trouble getting along with other children	0.610	0.939	0.002	
3.2 Other children do not want to be my friend	0.531	0.940	0.001	
3.3 Other children tease me	0.578	0.939	0.002	
3.4 Cannot do things that other children can do	0.713	0.937	0.004	
3.5 Keep up when I play with other children	0.667	0.938	0.003	
4.1 Pay attention in class	0.610	0.939	0.002	
4.2 Forget things	0.448	0.941	0	
4.3 Trouble keeping up with my schoolwork	0.679	0.938	0.003	
4.4 Miss school because of not feeling well	0.536	0.940	0.001	
4.5 Miss school to go to doctor or hospital	0.563	0.939	0.002	

Table 5.5 PedsQL™ 4.0 teen self-report (13-18 years) showing internal consistency reliability

Scale (n) Items	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Change (std alpha-Cronbach alpha if item deleted)	Cronbach's Alpha Based on Std Items
Physical Functioning sub-scale (n=189, excluded=2)				0.900
1.1 To walk more than one block	0.717	0.884	0.016	
1.2 To run	0.733	0.882	0.018	
1.3 Sports activity or exercise	0.766	0.880	0.020	
1.4 To lift something heavy	0.644	0.891	0.009	
1.5 To take a bath or shower by myself	0.680	0.888	0.012	
1.6 To do chores around the house	0.773	0.879	0.021	
1.7 Hurt or ache	0.621	0.893	0.007	
1.8 Low energy	0.562	0.898	0.002	
Emotional Functioning sub-scale (n=184, excluded=7)				0.846
2.1 Afraid or scared	0.647	0.816	0.030	
2.2 Sad or blue	0.772	0.780	0.066	
2.3 Angry	0.578	0.835	0.011	
2.4 Trouble sleeping	0.627	0.821	0.025	
2.5 Worry about what will happen to me	0.650	0.816	0.030	
Social Functioning sub-scale (n=187, excluded=4)				0.865
3.1 Trouble getting along with other teens	0.697	0.833	0.032	
3.2 Other teens do not want to be my friend	0.679	0.838	0.027	
3.3 Other teens tease me	0.613	0.854	0.011	

3.4 Cannot do things that other teens can do	0.724	0.826	0.039	
3.5 Keep up when I play with my peers	0.721	0.827	0.038	
School Functioning sub-scale (n=184, excluded=7)				0.890
4.1 Pay attention in class	0.697	0.865	0.025	
4.2 Forget things	0.609	0.885	0.005	
4.3 Trouble keeping up with my schoolwork	0.762	0.850	0.040	
4.4 Miss school because of not feeling well	0.764	0.849	0.041	
4.5 Miss school to go to doctor or hospital	0.778	0.845	0.045	
Psychosocial Health Summary scale (n=182, excluded 9)				0.929
2.1 Afraid or scared	0.641	0.925	0.004	
2.2 Sad or blue	0.686	0.923	0.006	
2.3 Angry	0.539	0.927	0.002	
2.4 Trouble sleeping	0.665	0.924	0.005	
2.5 Worry about what will happen to me	0.626	0.925	0.004	
3.1 Trouble getting along with other teens	0.688	0.923	0.006	
3.2 Other teens do not want to be my friend	0.608	0.925	0.004	
3.3 Other teens tease me	0.622	0.925	0.004	
3.4 Cannot do things that other teens can do	0.698	0.923	0.006	
3.5 Keep up with my peers	0.630	0.925	0.004	
4.1 Pay attention in class	0.705	0.923	0.006	
4.2 Forget things	0.604	0.926	0.003	
4.3 Trouble keeping up with my schoolwork	0.754	0.921	0.008	
4.4 Miss school because of not feeling well	0.687	0.923	0.006	
4.5 Miss school to go to doctor or hospital	0.704	0.923	0.006	

Total Scale Score (n=181, excluded=10)				**0.951**
1.1 To walk more than one block	0.683	0.948	0.003	
1.2 To run	0.654	0.948	0.003	
1.3 Sports activity or exercise	0.702	0.948	0.003	
1.4 To lift something heavy	0.662	0.948	0.003	
1.5 To take a bath or shower by myself	0.620	0.949	0.002	
1.6 To do chores around the house	0.736	0.947	0.004	
1.7 Hurt or ache	0.639	0.948	0.003	
1.8 Low energy	0.629	0.948	0.003	
2.1 Afraid or scared	0.646	0.948	0.003	
2.2 Sad or blue	0.670	0.948	0.003	
2.3 Angry	0.509	0.950	0.001	
2.4 Trouble sleeping	0.703	0.948	0.003	
2.5 Worry about what will happen to me	0.615	0.949	0.002	
3.1 Trouble getting along with other teens	0.658	0.948	0.003	
3.2 Other teens do not want to be my friend	0.591	0.949	0.002	
3.3 Other teens tease me	0.588	0.949	0.002	
3.4 Cannot do things that other teens can do	0.682	0.948	0.003	
3.5 Keep up with my peers	0.624	0.949	0.002	
4.1 Pay attention in class	0.707	0.947	0.004	
4.2 Forget things	0.590	0.949	0.002	
4.3 Trouble keeping up with my schoolwork	0.769	0.947	0.004	
4.4 Miss school because of not feeling well	0.714	0.947	0.004	
4.5 Miss school to go to doctor or hospital	0.726	0.947	0.004	

5.5.4 Validity

5.5.4.1 Convergent validity

Results of convergent validity are summarised in Tables 5.6 and 5.7.

All the intra scale correlation coefficients were above 0.40 for all PedsQL™ 4.0 child self-report sub-scales: Physical Functioning 0.40-0.74, Emotional Functioning 0.41-0.70, Social Functioning 0.42-0.66, and School Functioning 0.40-0.74. (Table 5.6) The PedsQL™ 4.0 teen self-report similarly had most of the intra scale correlation coefficients above 0.40: Physical Functioning 0.39-0.79, Emotional Functioning 0.39-0.67, Social Functioning 0.48-0.72, and School Functioning 0.49-0.87. (Table 5.7).

Table 5.6 PedsQL™4.0 Child self-report (8-12 years) Item correlation matrix

Scale	Items	Physical Functioning								Emotional Functioning					Social Functioning					School Functioning				
		1.1 To walk more than one block	1.2 To run	1.3 Sports activity or exercise	1.4 To lift something heavy	1.5 To take a bath or shower by myself	1.6 To do chores around the house	1.7 Hurt or ache	1.8 Low energy	2.1 Afraid or scared	2.2 Sad or blue	2.3 Angry	2.4 Trouble sleeping	2.5 Worry about what will happen to me	3.1 Trouble getting along with other children	3.2 Other children do not want to be my friend	3.3 Other children tease me	3.4 Cannot do things that other children can do	3.5 Keep up when I play with other children	4.1 Pay attention in class	4.2 Forget things	4.3 Trouble keeping up with my schoolwork	4.4 Miss school because of not feeling well	4.5 Miss school to go to doctor or hospital
Physical Functioning	1.1 To walk more than one block																							
	1.2 To run	0.61																						
	1.3 Sports activity or exercise	0.57	0.64																					
	1.4 To lift something heavy	0.58	0.56	0.57																				
	1.5 To take a bath or shower by myself	0.74	0.47	0.54	0.59																			
	1.6 To do chores around the house	0.60	0.40	0.50	0.56	0.70																		
	1.7 Hurt or ache	0.67	0.61	0.64	0.69	0.58	0.66																	
	1.8 Low energy	0.55	0.45	0.63	0.47	0.60	0.66	0.63																
Emotional Functioning	2.1 Afraid or scared	0.37	0.39	0.46	0.43	0.27	0.33	0.59	0.41															
	2.2 Sad or blue	0.54	0.55	0.53	0.57	0.44	0.43	0.60	0.44	0.68														
	2.3 Angry	0.37	0.35	0.52	0.41	0.33	0.40	0.37	0.43	0.44	0.70													
	2.4 Trouble sleeping	0.42	0.36	0.35	0.32	0.46	0.41	0.49	0.46	0.45	0.54	0.51												
	2.5 Worry about what will happen to me	0.51	0.38	0.44	0.37	0.41	0.43	0.48	0.40	0.57	0.62	0.41	0.52											
Social Functioning	3.1 Trouble getting along with other children	0.36	0.30	0.16	0.24	0.43	0.43	0.29	0.34	0.28	0.40	0.35	0.45	0.45										
	3.2 Other children do not want to be my friend	0.18	0.27	0.18	0.10	0.23	0.32	0.24	0.24	0.47	0.37	0.23	0.23	0.44	0.66									
	3.3 Other children tease me	0.40	0.42	0.14	0.20	0.40	0.19	0.42	0.26	0.28	0.42	0.31	0.51	0.45	0.59	0.42								
	3.4 Cannot do things that other children can do	0.44	0.52	0.42	0.53	0.47	0.42	0.58	0.33	0.51	0.54	0.36	0.38	0.51	0.47	0.46	0.60							
	3.5 Keep up when I play with other children	0.40	0.49	0.28	0.22	0.43	0.33	0.38	0.36	0.36	0.44	0.27	0.40	0.41	0.64	0.53	0.64	0.61						
School Functioning	4.1 Pay attention in class	0.47	0.36	0.27	0.29	0.36	0.37	0.39	0.38	0.32	0.38	0.21	0.28	0.35	0.62	0.42	0.41	0.49	0.64					
	4.2 Forget things	0.21	0.15	0.12	0.19	0.11	0.24	0.17	0.14	0.37	0.25	0.17	0.27	0.42	0.46	0.46	0.40	0.37	0.34	0.51				
	4.3 Trouble keeping up with my schoolwork	0.48	0.40	0.38	0.34	0.53	0.49	0.43	0.36	0.32	0.41	0.28	0.39	0.41	0.43	0.42	0.45	0.49	0.64	0.59	0.42			
	4.4 Miss school because of not feeling well	0.25	0.24	0.16	0.28	0.17	0.29	0.37	0.17	0.51	0.33	0.09	0.23	0.36	0.29	0.46	0.39	0.49	0.48	0.44	0.42	0.48		
	4.5 Miss school to go to doctor or hospital	0.32	0.24	0.23	0.23	0.27	0.43	0.32	0.20	0.40	0.33	0.18	0.24	0.43	0.35	0.57	0.25	0.42	0.48	0.40	0.41	0.65	0.74	

Yellow represents correlation of Physical functioning; Green represents correlation between Emotional functioning and Physical functioning; Blue represents correlation of Social functioning with Physical functioning (sky blue) and Emotional functioning (light blue); Orange represents School functioning correlation with Physical functioning (apricot shade), Emotional functioning (tangerine shade) and Social functioning (cantaloupe shade)

Table 5.7 PedsQL™4.0 Teen self-report (13-18 years) Item correlation matrix

Scale	Item	Physical Functioning								Emotional Functioning					Social Functioning					School Functioning				
		1.1 To walk more than one block	1.2 To run	1.3 Sports activity or exercise	1.4 To lift something heavy	1.5 To take a bath or shower by myself	1.6 To do chores around the house	1.7 Hurt or ache	1.8 Low energy	2.1 Afraid or scared	2.2 Sad or blue	2.3 Angry	2.4 Trouble sleeping	2.5 Worry about what will happen to me	3.1 Trouble getting along with other teens	3.2 Other teens do not want to be my friend	3.3 Other teens tease me	3.4 Cannot do things that other teens can do	3.5 Keep up with my peers	4.1 Pay attention in class	4.2 Forget things	4.3 Trouble keeping up with my	4.4 Miss school because of not feeling well	4.5 Miss school to go to doctor or hospital
Physical Functioning	1.1 To walk more than one block																							
	1.2 To run	0.66																						
	1.3 Sports activity or exercise	0.68	0.79																					
	1.4 To lift something heavy	**0.42**	0.49	0.51																				
	1.5 To take a bath or shower by myself	0.59	0.53	0.60	**0.42**																			
	1.6 To do chores around the house	0.63	0.57	0.58	0.58	0.70																		
	1.7 Hurt or ache	0.44	0.48	0.49	0.57	0.45	0.56																	
	1.8 Low energy	0.44	**0.41**	**0.44**	0.50	**0.39**	**0.51**	0.41																
Emotional Functioning	2.1 Afraid or scared	0.40	0.40	0.43	0.50	0.31	0.45	0.39	0.52															
	2.2 Sad or blue	0.41	0.33	0.39	0.52	0.37	0.48	0.43	0.45	0.67														
	2.3 Angry	0.29	0.26	0.27	0.37	0.27	0.35	0.37	0.27	**0.39**	**0.54**													
	2.4 Trouble sleeping	0.49	0.49	0.49	0.49	0.56	**0.63**	0.53	0.45	0.46	0.60	0.50												
	2.5 Worry about what will happen to me	0.41	0.40	0.47	0.43	0.30	0.42	0.33	0.42	0.55	0.60	**0.46**	0.47											
Social Functioning	3.1 Trouble getting along with other teens	0.42	0.40	0.42	0.41	0.38	0.45	0.36	0.37	0.43	0.50	0.44	0.44	0.42										
	3.2 Other teens do not want to be my friend	0.37	0.32	0.40	0.38	0.41	0.40	0.30	0.41	0.38	0.38	0.28	0.46	0.43	0.60									
	3.3 Other teens tease me	0.32	0.25	0.41	0.44	0.33	0.34	0.36	0.30	0.44	0.45	0.36	0.38	0.49	**0.48**	0.58								
	3.4 Cannot do things that other teens can do	0.46	0.36	0.39	0.38	0.36	0.43	0.42	0.47	0.50	0.50	0.28	0.43	0.47	0.57	0.56	**0.49**							
	3.5 Keep up with my peers	0.48	0.37	0.40	0.35	0.40	0.43	0.42	0.41	0.36	0.36	0.29	0.40	0.42	0.62	**0.50**	0.49	0.72						
School Functioning	4.1 Pay attention in class	0.50	0.57	0.51	0.44	0.46	0.54	0.43	0.40	0.48	0.53	0.35	0.54	0.45	0.56	0.39	0.41	0.42	0.40					
	4.2 Forget things	0.33	0.33	0.36	0.47	0.26	0.40	0.35	0.50	0.44	0.41	0.43	0.36	0.40	0.36	0.39	0.40	0.41	0.33	0.54				
	4.3 Trouble keeping up with my schoolwork	0.54	0.58	0.59	0.51	0.46	0.58	0.50	0.49	0.57	0.50	0.40	0.49	0.45	0.54	0.40	0.54	0.52	0.47	0.73	0.57			
	4.4 Miss school because of not feeling well	0.53	0.47	0.51	0.45	0.49	0.59	0.59	0.46	0.38	0.42	0.37	0.54	0.34	0.46	0.39	0.37	0.51	0.44	0.54	0.49	0.62		
	4.5 Miss school to go to doctor or hospital	0.52	0.53	0.55	0.45	0.49	0.56	0.52	0.42	0.39	0.42	0.33	0.55	0.36	0.44	0.41	0.41	0.53	0.50	0.55	0.51	0.63	0.87	

Yellow represents correlation of Physical functioning; Green represents correlation between Emotional functioning and Physical functioning; Blue represents correlation of Social functioning with Physical functioning (sky blue) and Emotional functioning (light blue); Orange represents School functioning correlation with Physical functioning (apricot shade), Emotional functioning (tangerine shade) and Social functioning (cantaloupe shade)

5.5.4.2 Discriminant validity

Discriminant validity results on a scale level by gender, age and school grade are shown in Tables 5.8 to 5.10 respectively.

Table 5.8 show that the mean difference in the total scale (as well as summary health) score between male and female children who completed the PedsQL™ 4.0 child self-report was not statistically significant (mean = 6.26, $t = 1.716$, $p < 0.90$). (Table 5.8) The mean difference in the total scale (as well as summary health) score between male and female adolescents who completed the PedsQL™ 4.0 teen self-report was similarly not statistically significant (mean = 2.71, $t = 1.062$, $p < 0.290$).

Across most of the sub-scales the mean difference in scores between male and female participants were similarly not statistically significant for PedsQL™ 4.0 child self-report and teen self-report. The only sub-scale where there was a statistically significant mean score difference between male and female participants was in the Social Functioning sub-scale for the PedsQL™ 4.0 child self-report (mean = 11.23, $t = 2.761$, $p < 0.007$). This was not the case with PedsQL™ 4.0 teen self-report (mean = 5.20, $t = 1.748$, $p < 0.082$).

Table 5.8 Discriminant validity by gender for PedsQL™ 4.0 Child self-report (8-12 years) and Teen self-report (13-18 years)

Scale	Child self-report (n= 98, male= 38, female=50)			Teen self-report (n= 182, male= 81, female=101)		
		t-test			t-test	
	mean difference	t-statistic*	p-value	mean difference	t-statistic*	p-value
Total Scale score	6.26	1.716	0.090	2.71	1.062	0.290
Psychosocial Summary Health	7.11	1.937	0.056	2.86	1.077	0.283
Physical Summary Health	2.18	0.481	0.632	1.90	0.715	0.477
Emotional Functioning sub-scale	3.41	0.762	0.448	2.52	0.911	0.364
Social Functioning sub-scale	11.23	2.761	**0.007**	5.20	1.748	0.082
School Functioning sub-scale	5.40	1.269	0.208	1.89	0.571	0.569

*assume equal variance

Table 5.9 shows correlation between age and PedsQL™ 4.0 self-report scores. The correlation between age and PedsQL™ 4.0 Child self-report Total scale score, Summary Health score and sub-scale score was 0.101-0.153. Similarly, the correlation between age and PedsQL™ 4.0 Teen self-report scores ranged between 0.007-0.162 which was below the hypothesized 0.2 threshold.

Table 5.9 Discriminant validity by age for PedsQL™ 4.0 child self-report (8-12 years) and teen self-report (13-18 years)

Scale	Correlation coefficient	
	Child self-report	Teen self-report
Total Scale score	-0.143	0.153
Psychosocial Health Summary	-0.146	0.116
Physical Health Summary	-0.105	0.162
Emotional Functioning sub-scale	-0.101	0.007
Social Functioning sub-scale	-0.153	0.149
School Functioning sub-scale	-0.153	0.119

The association between grade and PedsQL™ 4.0 self-report scores is contained in Table 5.10. The association between the total scale score and school grade for the PedsQL™ 4.0 child self-report was not statistically significant (F=1.078, sig 0.302). The trend was also evident in all the sub-scale and Summary Health scores for the PedsQL™ 4.0 child self-report.

There was however significant mean difference in the total scores of adolescents who completed PedsQL™ 4.0 Teen self-report and school grades (F=5.600, sig 0.004). (Table 5.10) There was similar significant association between school grade and Sub-scale scores except in the Emotional Functioning Sub-scale. When the age groups were combined, there was a similar association between school grade categories and the PedsQL™ 4.0 Total Scale score (F=8.786, sig 0.000). Association was also evident between grade and all the PedsQL™ 4.0 Sub-scale and Summary scores when age was combined.

Table 5.10 PedsQL™ 4.0 child self-report and teen self-report discriminant validity by grade

Scale		Child self-report (grade* group1=53, group2=40)				Teen self-report (grade* group1=18, group2=57, group3=111)				Child and Teen self-report combined (grade* group1=71, grade group 2=97, grade group3=111)			
		df	Mean Square	F	Sig.	df	Mean Square	F	Sig.	df	Mean Square	F	Sig.
Physical Functioning Sub-scale	Between Groups	1	15.6	0.035	0.851	2	1753.2	5.735	**0.004**	2	2606.8	7.461	**0.001**
	Within Groups	89	441.2			181	305.7			272	349.4		
	Total	90				183				274			
Emotional Functioning Sub-scale	Between Groups	1	518.3	1.248	0.267	2	318.5	0.942	0.392	2	1374.3	3.801	**0.024**
	Within Groups	87	415.3			176	338.1			265	361.5		
	Total	88				178				267			
Social Functioning Sub-scale	Between Groups	1	239.6	0.654	0.421	2	1981.2	5.231	**0.006**	2	2154.5	5.766	**0.004**
	Within Groups	84	366.3			179	378.7			265	373.7		
	Total	85				181				267			
School Functioning Sub-scale	Between Groups	1	1079.3	2.935	0.090	2	1851.0	3.969	**0.021**	2	2715.5	6.280	**0.002**
	Within Groups	86	367.8			176	466.3			264	432.4		
	Total	87				178				266			
Psychosocial Health Summary	Between Groups	1	633.3	2.296	0.133	2	1157.3	3.882	**0.022**	2	2025.7	7.011	**0.001**
	Within Groups	84	275.8			174	298.1			260	288.9		
	Total	85				176				262			
Physical Health Summary	Between Groups	1	15.6	0.035	0.851	2	1753.2	5.735	**0.004**	2	2606.8	7.461	**0.001**
	Within Groups	89	441.2			181	305.7			272	349.4		
	Total	90				183				274			
Total Scale Score	Between Groups	1	292.0	1.078	0.302	2	1504.4	5.600	**0.004**	2	2349.6	8.786	**0.000**
	Within Groups	83	270.8			173	268.6			258	267.4		
	Total	84				175				260			

*grade group1 = grade 1-5; group2 = grade 6-8; group3 = grade 9-12

5.5.4.3 Item convergent/discriminant validity

The MTMM matrix results in Table 5.11 suggest that the item convergent/discriminant validity criteria were generally met. The sub-scale inter-item relationship ratios are all within the criteria threshold of <0.85 as hypothesized. However, some of the sub-scales have their ratio close to this set threshold. For example, among the children who self-completed PedsQL™ 4.0 child self-report, the highest MTMM ratio (0.841) is in the Social Functioning / School Functioning matrix. Among adolescents who completed PedsQL™ 4.0 child teen-report the highest matrix ratio (0.843) is in Physical Functioning / School Functioning ratio. These high ratios would indicate that even though the criteria for discriminant validity was met the items belonging to these sub-scales do not completely establish discriminant validity.

Table 5.11 Item convergent/ discriminant validity for PedsQL™4.0 child self-report and teen self-report by Sub-scale

Sub-scale	Child self-report				Teen self-report			
	Physical Functioning	Emotional Functioning	Social Functioning	School Functioning	Physical Functioning	Emotional Functioning	Social Functioning	School Functioning
Physical Functioning	-				-			
Emotional Functioning	0.761	-			0.785	-		
Social Functioning	0.579	0.711	-		0.710	0.760	-	
School Functioning	0.541	0.606	**0.841**	-	**0.843**	0.774	0.755	-

The numbers are the MTMM ratios

Further inspection of the data revealed there were a few items that appear not to have clearly met the criteria for item convergent/discriminant validity. The lack of item convergent/discriminant validity by individual items was explored by comparing the lowest

intra scale correlations that had higher inter scale correlations and this showed some inconsistencies. For example, the 0.395 intra-item correlation for 'To run' had higher inter-item correlations with 'Sad or blue' (0.546) in the Emotional Functioning Sub-scale. (Table 5.12) This trend was seen between sub-scales such that on average the Emotional Functioning Sub-scale was more correlated with the School Functioning Sub-scale for the PedsQL™ 4.0 teen self-report.

Table 5.13 contain results for the PedsQL™ 4.0 teen self-report. On average, the Physical Functioning Sub-scale had more correlation with School Functioning.

Table 5.12 showing highest within and between correlations (in bold) of the PedsQL™4.0 Child self-report (8-12 years)

Scale		Physical Functioning								Emotional Functioning					Social Functioning					School Functioning				
	Item	1.1 To walk more than one block	1.2 To run	1.3 Sports activity or exercise	1.4 To lift something heavy	1.5 To take a bath or shower by myself	1.6 To do chores around the house	1.7 Hurt or ache	1.8 Low energy	2.1 Afraid or scared	2.2 Sad or blue	2.3 Angry	2.4 Trouble sleeping	2.5 Worry about what will happen to me	3.1 Trouble getting along with other children	3.2 Other children do not want to be my friend	3.3 Other children tease me	3.4 Cannot do things that other children can do	3.5 Keep up when I play with other children	4.1 Pay attention in class	4.2 Forget things	4.3 Trouble keeping up with my schoolwork	4.4 Miss school because of not feeling well	4.5 Miss school to go to doctor or hospital
Physical Functioning	1.1 To walk more than one block													**.510****						**.468****		**.484****		
	1.2 To run					.466**	.395**		.449**		**.546****							**.521****						
	1.3 Sports activity or exercise											**.523****		**.438****										
	1.4 To lift something heavy										**.565****	**.407****						**.528****						
	1.5 To take a bath or shower by myself							.584**					**.456****	**.412****				**.465****				**.525****		
	1.6 To do chores around the house		.395**	.503**										**.425****								**.489****		**.429****
	1.7 Hurt or ache										**.599****		**.487****	**.483****				**.582****				**.432****		
	1.8 Low energy	.553**			**.470****							**.427****	**.458****											
Emotional Functioning	2.1 Afraid or scared							**.586****					.451**			**.473****		**.508****					**.508****	
	2.2 Sad or blue		**.546****	**.531****	**.565****		**.426****	**.599****										**.542****						
	2.3 Angry			**.523****						.440**				.406**										
	2.4 Trouble sleeping						**.409****				.544**						**.513****							
	2.5 Worry about what will happen to me						**.425****					.406**				**.438****	**.453****	**.509****			**.420****			**.433****
Social Functioning	3.1 Trouble getting along with other children						**.430****							**.447****						**.617****	**.463****	**.434****		
	3.2 Other children do not want to be my friend									**.473****				**.438****			.424**	.457**	.525**	**.419****	**.461****	**.424****	**.459****	**.567****
	3.3 Other children tease me		**.418****										**.513****	**.453****		.424**				**.411****		**.446****		
	3.4 Cannot do things that other children can do		**.521****		**.528****		**.422****			**.508****				**.509****	.466**					**.486****		**.491****	**.493****	**.417****
	3.5 Keep up when I play with other children		**.485****											**.410****						**.643****		**.636****	**.483****	**.484****
School Functioning	4.1 Pay attention in class														**.617****			**.486****	**.643****					.403**
	4.2 Forget things													**.420****		**.461****						.421**	.420**	
	4.3 Trouble keeping up with my schoolwork		**.399****			**.525****	**.489****							**.413****		**.424****	**.446****	**.491****	**.636****					
	4.4 Miss school because of not feeling well									**.508****						**.459****		**.493****						
	4.5 Miss school to go to doctor or hospital						**.429****							**.433****		**.567****				.403**	.412**			

** all correlation significant at the 0.01 level (2-tailed)

Table 5.13 showing highest within and between correlations (in bold) of the PedsQL™4.0 Teen self-report (13-18 years)

Scale		Physical Functioning								Emotional Functioning					Social Functioning					School Functioning				
	Item	1.1 To walk more than one block	1.2 To run	1.3 Sports activity or exercise	1.4 To lift something heavy	1.5 To take a bath or shower by myself	1.6 To do chores around the house	1.7 Hurt or ache	1.8 Low energy	2.1 Afraid or scared	2.2 Sad or blue	2.3 Angry	2.4 Trouble sleeping	2.5 Worry about what will happen to me	3.1 Trouble getting along with other teens	3.2 Other teens do not want to be my friend	3.3 Other teens tease me	3.4 Cannot do things that other teens can do	3.5 Keep up with my peers	4.1 Pay attention in class	4.2 Forget things	4.3 Trouble keeping up with my schoolwork	4.4 Miss school because of not feeling well	4.5 Miss school to go to doctor or hospital
Physical Functioning	1.1 To walk more than one block									.396**			.492**										.532**	.524**
	1.2 To run									.398**			.492**						.566**		.582**			.526**
	1.3 Sports activity or exercise									.429**			.491**	.473**							.589**	.507**	.552**
	1.4 To lift something heavy	**.423****								.501**			.493**											
	1.5 To take a bath or shower by myself				**.422****				**.392***				.556**									.493**		
	1.6 To do chores around the house									.450**			.628**								.576**	.589**	.558**	
	1.7 Hurt or ache												.525**									.586**	.523**	
	1.8 Low energy		**0.408****	**0.444****		**0.392****	**0.509****	**0.410****		.520**										.495**				
Emotional Functioning	2.1 Afraid or scared				.501**				.520**			**.393****	**.458****					.499**			.571**			
	2.2 Sad or blue				.523**			.429**	.453**						.503**			.502**						
	2.3 Angry									**.393****	**537****			**.462****										
	2.4 Trouble sleeping		.492**	.491**	.493**	.556**	.628**	.525**	.454**										.536**			.541**	.550**	
	2.5 Worry about what will happen to me			.473**	.432**				.422**								.494**							
Social Functioning	3.1 Trouble getting along with other teens									.428**		.443**					**.481****		.556**					
	3.2 Other teens do not want to be my friend					.409**			.413**				.461**											
	3.3 Other teens tease me				.440**					.436**				.494**	**.481****			**.490****	**.493****					
	3.4 Cannot do things that other teens can do	.463**						.418**	.473**	.499**				.468**								.511**	.528**	
	3.5 Keep up with my peers	.479**				.395**		.422**	.414**							**.499****								
School Functioning	4.1 Pay attention in class	.504**	.566**	.514**	.443**	.460**	.535**	.429**	.397**	.484**			.536**		.556**									
	4.2 Forget things				.474**				.495**	.435**		.427**								**.536****		**.566****	**.492****	**.510****
	4.3 Trouble keeping up with my schoolwork	.544**	.582**	.589**	.505**	.458**	.576**	.503**	.491**	.571**		.398**	.485**		.543**		.540**	.522**						
	4.4 Miss school because of not feeling well	.532**		.507**	.448**	.493**	.589**	.586**	.459**				.541**					.511**			**.492****			
	4.5 Miss school to go to doctor or hospital	.524**	.526**	.552**	.452**	.488**	.558**	.523**	.416**				.550**					.528**	.504**					

** all correlation significant at the 0.01 level (2-tailed)

5.5.4.4 Known-groups validity

The overall score effect sizes between the healthy and sick children were variable. Generally, the PedsQL™ 4.0 child self-report displayed poor effect size for all the sub-scale scores and total scale score except for a small effect size in the Physical Functioning (0.32) sub-scale. (Table 5.14) For the PedsQL™ 4.0 teen self-report, the effect size ranged from poor (0.21) to large (0.93) for the sub-scale scores. The sub-scale with poor effect size was Emotional Functioning (0.21), while all the other sub-scale score effect sizes, as well as that for the total scale score, were moderate to large for Physical Functioning (0.93).

Table 5.14 PedsQL™4.0 child self-report (8-12 years) and teen self-report (13-18 years)

Scale	Child self-report (N=98, healthy=12, 81=sick)					Teen self-report (N=191, healthy=83, 106= sick)				
	t-statistic #					t-statistic #				
	t	p-value	Mean Difference	Std. Deviation	Effect size *	t	p-value	Mean Difference	Std. Deviation	Effect size*
Total Scale Score	0.751	0.455	4.14348	20.75426	0.20	3.720	0.000	9.0645	11.41292	0.79
	0.634	0.538	4.14348	16.61964		3.915	0.000	9.0645	19.37174	
Physical Health Summary	0.914	0.363	6.30889	19.70087	0.32	4.271	0.000	10.7565	11.58783	0.93
	0.987	0.341	6.30889	21.74236		4.550	0.000	10.7565	20.51927	
Psychosocial Health Summary	0.578	0.565	3.08368	20.80604	0.15	3.374	0.001	8.6298	13.07361	0.66
	0.490	0.632	3.08368	16.66462		3.524	0.001	8.6298	19.81541	
School Functioning Sub-scale	0.700	0.485	4.21811	22.03647	0.19	3.712	0.000	11.6917	15.29037	0.76
	0.629	0.540	4.21811	19.09115		3.919	0.000	11.6917	24.88628	
Social Functioning Sub-scale	0.381	0.704	2.33650	19.36003	0.12	3.932	0.000	11.1282	15.04293	0.74
	0.388	0.703	2.33650	19.85765		4.097	0.000	11.1282	21.99724	
Emotional Functioning Sub-scale	0.394	0.694	2.55081	22.05365	0.12	1.237	0.218	3.3789	16.42370	0.21
	0.377	0.712	2.55081	20.76237		1.262	0.208	3.3789	19.87847	
Physical Functioning Sub-scale	0.914	0.363	6.30889	19.70087	0.32	4.271	0.000	10.7565	11.58783	0.93
	0.987	0.341	6.30889	21.74236		4.550	0.000	10.7565	20.51927	

assuming equal variance

*effect size designated as <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, and >0.8 large

5.6 Discussion

The Chichewa versions of the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report demonstrated mixed psychometric results. Both instruments had mixed performance at item level. Both instruments performed well at item level (maximum endorsement frequency, and item redundancy), reliability, convergent validity, and discriminant validity by age. However, the PedsQL™ 4.0 child self-report version failed on completion of all responses without missing out some responses (item missing data), utilization of all responses (aggregate endorsement frequency) and to distinguish scores between the sick and the health population (known-group validity). The PedsQL™ 4.0 teen self-report similarly failed to meet criteria for utilization of all item responses and to discriminate by age. In general, the Chichewa versions of the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report performed well for use in older (>12years) children and adolescents although test-retest and responsiveness need to be evaluated. Further, children ≤12 years would benefit from use of an interviewer administered instrument instead of a self-completion.

The PedsQL™ 4.0 child self-report had problems with missing data which might point to the struggle of self-completing of questionnaires by younger participants. A similar observation was made when the PedsQL™ 4.0 was initially evaluated after development. (132, 285) This may indicate the challenge of obtaining good quality assessment of HRQoL from younger children using the self-completed PedsQL™ 4.0 child self-report. The analysis based on health conditions also showed evidence of missing data for both the PedsQL™ 4.0 child and teen self-report. This finding might point to problems with the content validity or to use of self-report generic health instrument in this setting. The fact that missing data was an issue across all health conditions raises issues of measuring health and HRQoL in this setting. (267, 285) This is explored in chapter 7 where health is defined from children and adolescents from a Malawian

perspective.

The evidence of lack of utilization of all item responses (adjacent aggregate endorsement frequency) for both the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report in all the health conditions would indicate that the children found some of the responses not relevant or difficult to understand and use. This could possibly be expected since the PedsQL™ 4.0 was developed to measure HRQoL across heterogenous populations. The sample in this study was relatively healthy, even among the sick population, as was evident in responses mainly in the 'never' and 'almost never' a problem options. Other studies have observed this problem when utilised in healthy populations. (290) The findings in this study indicate that the combination of item wording and response options might not be appropriate for this setting. Elsewhere, some have explored need to examine the response options perhaps by reducing the number of responses. (291). Evidence from this present study may support the need to examine the response options in order to minimize this problem.

The child and teen versions of the PedsQL™ 4.0 performed well with regards to internal consistency reliability. This shows that the Chichewa versions have potential of producing consistent results when used in a Malawian setting. The excellent internal reliability of above Cronbach's alpha of 0.9 for both instruments, provides evidence that total scale scores may be suitable for patient group level analysis. (230) Similarly, for both versions, the Cronbach's acceptable level indicates homogeneity of items within the instruments. This is consistent with high reliability reported for the original PedsQL™ 4.0. (285) This provides evidence that the total scale score as well as respective sub-scales could be used to examine specific domains of functioning in this setting. This is similar to the all the sub-scales except Emotional Functioning. The lower Cronbach's alpha for the Emotional Functioning sub-scale in the

PedsQL™ 4.0 teen self-report could perhaps indicate that adolescents are more self-conscious about their feelings. Adolescents more so than children tend to hide their feeling and might not express these in an instrument. Further, the fact that the sample was generally made up of a healthier population might imply that the items in this Emotional Functioning sub-scale were not relevant them. Previous findings have shown that the Emotional Functioning sub-scale performs much better among participants with some mental health condition for example. (286) On the other hand, a lower (but acceptable) Cronbach's alpha in the School Functioning sub-scale in the PedsQL™ 4.0 child self-report indicates that items relating to school may not be appropriate to children in this setting. It is possible this may be because the concept of school might not be understood by sample of children which was surprising since all of them were school going.

Both the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report performed as expected in terms of convergent validity. However, the negative correlation (though not significant) between age and PedsQL™ 4.0 total scale scores in younger children might be of concern. This could indicate that the sample of participants probably did not understand the items which raises concerns of obtaining HRQoL scores from children in this setting. There was also inconsistency with lack of discriminant validity by grade in the PedsQL™ 4.0 teen self-report in all sub-scales except Emotional Functioning. This could be due to a small number of children in the grade for standards 1-5 group which may have skewed the association. However, the lack of discriminant validity when data were combined in all sub-scales would perhaps indicate that the grade is a factor in self-completion of questionnaire. This may well support that grade or being in school is necessary to better understanding the questionnaire and obtaining better HRQoL data.

The adapted MTMM analyses generally established item convergent/discriminant validity between the sub-scale scores for both the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report. This is consistent with other MTMM findings for PedsQL™ 4.0 to measure multidimensional constructs. (285) However, the high MTMM ratio in some matrix would indicate that while both instruments achieved the acceptable threshold, they did not completely establish item convergent/discriminant validity. A further inspection of the correlation matrix revealed several items had higher inter-scale than intra-scale correlations especially in the same sub-scales that had high MTMM ratio. This detailed level of item analysis was only possible because the data were from one time point and from single completion. However, this finding does reveal problems identified with some individual items in this setting. It might be that the way some items like 'worry about what will happen to me', 'cannot do things that other children do', 'trouble keeping up with my schoolwork' and 'to do chores around the house' are expressed in this setting may be problematic. In turn this impacts on how these items are interpreted/understood meaning that they have a relatively high correlation with a domain that is not its own. The understanding and therefore interpretation of the items could be due to different factors including the setting, characteristics of the participants and the health care system from which the sample was taken among others. This study is not able to ascertain the possible cause since the sample was from the same area. However, it may likely be that since most participants were relatively healthy, some of these questions did not apply to them. Future studies could look at the role of the setting in order to ascertain this.

The PedsQL™ 4.0 teen self-report was able to distinguish between the healthy and sick populations. The biggest effect size was in the Physical Functioning sub-scale which is consistent with extant literature (290) and the smallest in the Emotional and Social Functioning sub-scales. This study was unable to establish why this was so and a qualitative follow up study

would have assisted to answer this. On the other hand, the PedsQL™ 4.0 child self-report was unable to distinguish between the healthy and sick population except in the Physical Functioning sub-scale. In general, the effect sizes were much lower among the children compared to adolescent participants. One possibility for this difference was probably the small sample of healthy compared to sick children which may have amplified the effect size in the former.

Having done a short-term validation of the measures, there are points that need further reflection. A question arises as to whether there is a need for long-term valuation of patient-reported outcome measures or person-reported outcomes like PedsQL™ 4.0, and if that were the case perhaps adults would be best suited at understanding long-term consequences. As discussed in chapter 3, it is a fact that adults have once been children and therefore better positioned to determine needs and make choices pertaining to health for children. This might be justified from the viewpoint of decision makers to have an older person rate health for a child. Afterall, children and adolescents are generally not involved in the financing and decision making of the healthcare system. (292) Furthermore, children and adolescents may be inexperienced in making medical decisions for themselves because these decisions are ultimately made by the person responsible for caring for them. (293) But as brought out in chapter 3.7.1, it can be argued (293, 294) that it is the child or adolescent who experiences the health states and lives with the health impairments. (138) As such preferences of children and adolescents are directly relevant. (295) Further, the argument of having long-term valuation by adults could be extended to adult instruments too. Adult instruments transcends across all adult age groups but the needs of an 18year-old may be different from that of a 40 year-old which may further be different from that of a 60 year-old. Yet, adult instruments are completed across the whole adult age irrespective of the age band.

Secondly, there is a role of 'social expectations' on subjective measures of health. Evidence shows that LMICs have lowered expectations and therefore might rank their health better than HICs. At the same time there is an association between education, income/social status etc those in higher SES tend to have higher HRQoL. (296) Children are also embedded within various social aspects including peers, the classroom, and the community (297) that can influence their subjective measure of health. The role of gender could also influence how health is perceived. But as this study showed there was no association between gender and health scores. This is important since, like adult instruments, children instruments should also go beyond the gender divide. Were this to be a factor there would be need to develop several instruments targeting specific gender.

Limitations of the study need to be considered. First, as explained in chapter 4, the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report were adapted by interviewing the healthy population only. It could not be established therefore if the concepts would have changed if patient population had been involved at the adapting stage, although it is highly unlikely that it would have changed the translation text. Another limitation is to do with data for the psychometric assessment which were collected from one time point only and so neither test-retest reliability nor responsiveness could be performed in this study. Test-retest is necessary to establish if the instrument is stable over time and responsiveness ensures that the instrument detects clinical meaningful changes over time. This would further cement the usefulness of the instrument in clinical settings.

5.7 Conclusion

The findings of this study demonstrate the Chichewa (Malawi) PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report established reliability, but validity was mixed. The Chichewa (Malawi) PedsQL™ 4.0 teen self-report will be valuable for HRQoL measurement. The lack of discriminant validity by grade need to be further investigated as well as aggregate endorsement frequency. While the PedsQL™ 4.0 child self-report met most psychometric criteria, the missing data and lack of known-groups validity means this should be used with caution in younger children and in those with different health conditions. Although there are some difficulties with the PedsQL™ 4.0 self-report especially with the younger version, it's still demonstrated robust internal consistency and some validity. The psychometric evidence presented in this chapter show that the PedsQL™ 4.0 self-report is adequate for use as a validating measure in the evaluation of the EQ-5D-Y.

5.8 Summary of chapter 5

Chapter 5 evaluated the psychometric performance of the Chichewa (Malawi) versions of the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report. Psychometric evaluation is necessary in establishing an instrument reliability and validity across different population groups. The evidence shows both instruments performed generally well in terms of acceptability and reliability. Validity even though the criteria was met, showed that some items might not be suited for this setting. Conclusively, the results confirm that the Chichewa PedsQL™ 4.0 versions are appropriate health measures in Malawi although this must be done with caution among children <12 years.

6 Chapter 6- Psychometric evaluation/validation of the EQ-5D-Y and EQ-5D-Y-5L Chichewa versions

With:

Ngwira, L.G; Verstraete, J.; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S
2nd African Regional EuroQol Academy Meeting 2022, Virtual. (Oral)

Ngwira, L.G; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S
5th EuroQol Academy Meeting 2021, Virtual. (Poster)

Ngwira, L.G.; Verstraete, J.; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S
(manuscript drafted)

Verstraete, J.; Ngwira, L.G.; Maheswaran, H.; Petrou, S.; Niessen, L.; Smith, S
(manuscript drafted)

6.1 Overview of chapter

This chapter evaluates the psychometric performance of the Chichewa versions of EQ-5D-Y and EQ-5D-Y-5L. The chapter will assess the acceptability, reliability and validity of both the EQ-5D-Y and EQ-5D-Y-5L for use in Malawi. The validity assessment will consider different aspects including construct, known-groups and empirical validity. In order to achieve construct validity, the EQ-5D-Y and EQ-5D-Y-5L will not only be compared against each other but also with the PedsQL™ 4.0 Child report and teen report.

6.2 Introduction

The adult EQ-5D-3L, introduced in chapter 4, is one of the most widely used preference-based HRQoL measures in health economic evaluations. (122) The prominence of the EQ-5D instruments can be attributed to its simple descriptive system comprising of only five items (representing five dimensions, one item per dimension) each with three (EQ-5D-3L) or five

responses (EQ-5D-5L). This simple descriptive system and corresponding sets of response levels has made it relatively easy for valuation exercises to generate utility values but also for self-completion by participants. The EQ-5D-3L has been criticized for being too simplistic and less sensitive to small changes in health status, which led to the development of the five response level EQ-5D-5L. (256)

As discussed in chapter 4, EQ-5D youth versions emerged from the adult EQ-5D versions. (159, 259) The youth versions include the EQ-5D-Y three-level and the extended five-level EQ-5D-Y-5L, currently undergoing piloting before becoming an official version. Research suggests the adult five-response level EQ-5D-5L version performs better, less affected by the ceiling effect and improved known-groups validity compared to the three- level EQ-5D-3L version. (298-300). The EQ-5D-Y-5L has been developed on the same premise as the adult EQ-5D-5L version to increase its sensitivity and to reduce ceiling effects. (256)

Psychometric performance of the EQ-5D-Y has been reported in a limited number of studies involving healthy children as well as those with different health conditions. (160, 261, 301) Further, the EQ-5D-Y has been widely used in different studies including interventional, observational, and randomised controlled trials, routine data collection as well as across therapeutic areas. (261) To a large extent, the EQ-5D-Y has demonstrated good reliability, with acceptable levels of convergent, discriminant and known-group validity around the globe. (181, 302, 303) One systematic review showing use of the EQ-5D-Y between 2000-2010 reported that most studies have reported problems with missing values associated with the instrument. (304) The performance of the EQ-5D-Y-5L has also been reported but in only a few studies. (260, 279, 280) While the EQ-5D-Y-5L has demonstrated to be feasible and to reduce ceiling

effect, it has not performed differently on other psychometric properties than the EQ-5D-Y. (260, 279, 305)

Neither the EQ-5D-Y nor the EQ-5D-Y-5L have been psychometrically evaluated in Malawi. Additionally, since both the EQ-5D-Y and EQ-5D-Y-5L are preference-based measures, their utility values need to be evaluated as well as their psychometric properties. (154) Unfortunately, the utility algorithms for the EQ-5D-Y have just been developed more recently and due to this, researchers have tended to apply adult tariffs to EQ-5D-Y health states. (100, 304) This approach has, however, not been encouraged by the EuroQol Group. (306) The pros and cons of using adult or proxy tariffs, as well as de novo valuation techniques have been discussed at length in chapter 2 of this thesis and elsewhere and are not the subject of this chapter. (126, 141, 261, 307) In general, there is a discrepancy between children and adult/proxy utility values, and this is the case even among adults themselves. It has further been argued that some of the valuation techniques are either unethical or difficult for children cognitive abilities. The sensitivity of the EQ-5D-Y and EQ-5D-Y-5L as preference-based HRQoL measures for use in economic evaluation have not yet been assessed in Malawi.

The psychometric performance of the cross-culturally adapted versions of the EQ-5D-Y and EQ-5D-Y-5L has not been conducted elsewhere. Additionally, empirical validity of the EQ-5D-Y vis-à-vis that of the EQ-5D-Y-5L has not been conducted elsewhere owing to the latter being a beta version. This study therefore aims to directly compare the specific psychometric properties of the Chichewa versions of the EQ-5D-Y and EQ-5D-Y-5L in general population, chronic and acutely ill children and adolescents in Malawi. Secondly, the study compares the empirical validity of the two measures.

6.3 Aims and objectives

The aim of this analysis was to psychometrically evaluate the Chichewa (Malawi) language versions of the EQ-5D-Y and EQ-5D-Y-5L. Specifically, this was to explore i) item analysis (level of missing data and response distribution and item redundancy); ii) reliability (scales are homogenous); and iii) validity (scales measure what they purport to measure). As both EQ-5D-Y and EQ-5D-Y-5L are preference-based HRQoL measures, their empirical validity was also assessed, by comparing utility scores and stated general health.

6.4 Methods

6.4.1 Participants, recruitment and procedure

The participant and recruitment procedure outlined in chapter 4 section 4.1 was followed for this part of the research. The sample of healthy, chronic and acute ill participants that were administered the EQ-5D-Y and EQ-5D-Y-5L, were the same ones that also completed the PedsQL™ 4.0 child and PedsQL™ 4.0 teen self-report, as outlined in chapter 5 section 4.1.

In one previous study, it was found that patients avoided the intermediate response level in the equivalent adult five level EQ-5D-5L questionnaire if the EQ-5D-3L was administered first. (298) For this reason, the five response level EQ-5D-Y-5L was administered first before the three-response level EQ-5D-Y. After which the PedsQL™4.0 Child self-reports and PedsQL™ 4.0 Teen self-reports were administered to those aged 8-12 years and 13-17 years respectively.

6.4.2 The instruments

The EQ-5D-Y and EQ-5D-Y-5L were presented in chapter 4 of this thesis. The EQ-5D-Y-5L included a cover page with questions to collect demographic and medical condition (Appendix

19) The rest of the questionnaire maintained the structure and content as per developers' format. Another important feature of the EQ-5D instruments is the EQ VAS. As discussed in chapter 4, there are some limitations with the EQ VAS but it is an integral part of any of the EQ-5D instruments. Generally VAS has advantages of simplicity, ease of administration and scoring, as well suitability for frequent and repeated use. (277) Some studies have found VAS to be valid and reliable (308) but not others. (237) Nevertheless the EQ VAS has been used to compare scores with the EQ-5D profile although with some limitations. (276) A self-rated general health was also included through the question: How would you rate your health today? Excellent, very good, good, fair, or poor? Self-rated health questions are considered a reliable and valid, although limited since they are single item and therefore lack content validity, measure of HRQoL. (309) Nevertheless, they have been used in various studies involving children and adolescents as a crude measure for comparison. (279)

The EQ-5D-Y and EQ-5D-Y-5L were scored in two different ways to assess psychometric scale analysis and utility performance (empirical validity). Firstly, for psychometric scale analysis, the scores were generated by summing the responses. A health state (represented by responses) '11111' (denoting a one for each of the five dimensions) had a sum score of 5. Similarly, a health state of '33333' and '55555' had sum scores of 15 and 25 for the EQ-5D-Y and EQ-5D-Y-5L respectively. The other health states had a sum score ranging between 5 and 15 for the EQ-5D-Y and between 5 and 25 for the EQ-5D-Y-5L. Both the EQ-5D-Y and EQ-5D-Y-5L are a one item only for each dimension, therefore the dimension scores are the same as the 5 item sum scores. This is a crude measure with some limitations but gives a better indication of the dimension performance. (310)

For the purpose of assessing construct validity, correlation of EQ-5D-Y and EQ-5D-Y-5L sum scores and PedsQL™ 4.0 total scale scores were evaluated. The PedsQL™ 4.0 scores (health summary and total scale scores) were calculated as outlined in chapter 4.4.3.1. The sample for these analyses was divided into two groups, those aged 8-12 years and those 13-17 years. This was to reflect the age ranges for the PedsQL™4.0 child self-reports (for age 8-12 years) and PedsQL™ 4.0 teen self-reports (for age 13-18 years).

Secondly, to assess utility performance (empirical validity), since both the EQ-5D-Y and EQ-5D-Y-5L are utility instruments, utility scores were generated. This aspect was done to assess the degree to which utility scores of the EQ-5D-Y and EQ-5D-Y-5L predict self-reported general health. Youth tariff sets were not available at the time of doing this research. In addition, few countries have adult utility values for both the EQ-5D-3L and EQ-5D-5L, and none of these is in Africa. (311) The utility values for the EQ-5D-Y and EQ-5D-Y-5L were therefore generated from the USA tariffs sets developed by Shaw et al (312) and Pickard et al (313), respectively on recommendation from the EuroQol Group.

6.4.3 Psychometric analyses

Data analysis were performed using IBM SPSS 26.0.0. for Mac (IBM Corp. Armonk, New York, USA). (257) Psychometric analyses were conducted using gold standard techniques and based on the guidelines outlined in previous chapters 2 and 5. Item performance, reliability (internal consistency) and validity (convergent, discriminant and known-groups) were evaluated. In addition, utility performance for evaluating empirical validity was also performed. The detailed analyses are outlined below.

6.4.3.1 Item performance

The proportion of missing responses and distribution of responses in each EQ-5D-Y and EQ-5D-Y-5L dimension was calculated. As discussed in chapter 5.4.3.1, there are four aspects of item analysis that were evaluated: missing data, MEF, AEF and item redundancy. The same criteria for assessing each of these as outlined there were also applied here:

Missing data: EQ-5D-Y and EQ-5D-Y-5L dimensions that had ≥ 5% missing data were flagged as potentially problematic.

MEF: Dimensions were considered problematic if they had >80% endorsement at either end of the scale.

AEF: Dimensions were considered problematic if any two or more adjacent response options summed to <10%.

Item redundancy: Pairs of items with correlations > 0.75 were considered problematic.

A final component that was assessed was the evidence for cluster responding for the EQ VAS. There is no universally accepted criterion for this and in this study, this was arbitrarily defined as highly likely if responses at any point were ≥5%.

6.4.3.2 Reliability

This was assessed using Cronbach's alpha ≥0.70 to indicate acceptable internal consistency. The corrected item-total correlations were hypothesized to be in the range 0.25-0.8. Further, it was hypothesized that there would be little difference between the standardized alpha for the

overall scale and the Cronbach's alpha if an item is deleted.

6.4.3.3 Validity

6.4.3.3.1 Convergent validity

This was evaluated by comparing the EQ-5D-Y and EQ-5D-Y-5L sum scores with the PedsQL™ 4.0 total scale scores. Although the PedsQL™ 4.0 is a non-preference-based HRQoL instrument, it has been used extensively for convergent validity against both versions of the EQ-5D-Y. (160, 204, 206-208) It was hypothesized that the EQ-5D-Y and EQ-5D-Y-5L sum scores would be correlated with PedsQL™ 4.0 total scale scores. Additionally, it was hypothesized that the EQ-5D-Y and EQ-5D-Y-5L dimensions would be correlated with similar PedsQL™ 4.0 health summary scores. That means, the EQ-5D-Y mobility and feeling worried, sad or unhappy would be correlated with PedsQL™ 4.0 physical and psychosocial health summary scores respectively. The correlation should be negative because lower sum scores for EQ-5D-Y and EQ-5D-Y-5L means better HRQoL whereas higher PedsQL™ 4.0 score means better HRQoL. Generally, a correlation ≥0.4 is considered moderate to strong.

6.4.3.3.2 Discriminant validity

It was hypothesized that there would be no association between EQ-5D-Y and EQ-5D-Y-5L sum scores with gender (t-test), age (correlations) and grade (ANOVA).

It was anticipated that there would be none to small correlation <0.2 between EQ-5D-Y, EQ-5D-Y-5L sum scores and age. Similarly, it was anticipated that there would be no association at the 5% significance level between EQ-5D-Y and EQ-5D-Y-5L sum scores, with gender and grade.

6.4.3.3.3 Known-groups validity

It was hypothesised that the EQ-5D-Y and EQ-5D-Y-5L sum scores would be worse for the sick children compared to that of their healthy counterpart. The relationship was investigated using a t-test to establish lack of significant difference between the two groups. The magnitude of the difference was evaluated using an effect size according to Cohen's criterion: <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, and >0.8 large. (150, 231)

6.4.3.4 Utility performance (Empirical validity)

It was hypothesised that the utility scores generated by the EQ-5D-Y and EQ-5D-Y-5L would detect differences in external indicators of health status. Additionally, it was expected that the EQ-5D-Y-5L would possess greater empirical validity compared to the EQ-5D-Y. This hypothesis was based on the fact that the EQ-5D-Y-5L, like the adult EQ-5D-5L counterpart, was developed on the premise to address ceiling effect as well as sensitivity. (256) As such the EQ-5D-Y-5L is expected to have better characteristics compared to the EQ-5D-Y. The EQ-5D-Y and EQ-5D-Y-5L ability to assess differences in external indicators of health status was investigated in two ways. Firstly, the utility scores were compared with a single five response level self-reported general health question. This self-completed question asks participants to rate their general health from excellent to poor. Secondly, the EQ-5D-Y and EQ-5D-Y-5L utility scores were compared against the PedsQL™ 4.0 total scale scores. Lastly, the utility scores from the EQ-5D-Y and EQ-5D-Y-5L were compared against the EQ VAS scores.

The relative ability of the utility scores to detect differences in external indicators of health status was estimated using the relative efficiency statistic. Relative efficiency is defined as 'the ratio of the square of the t-statistic of the comparator instrument over the square of the t-statistic of the reference instrument'. (314) The EQ-5D-Y-5L acted as the comparator instrument in this study and the EQ-5D-Y as the reference since the latter has been widely used and

psychometrically validated. (259) The following criteria for relative efficiency were used: Relative efficiency equal to 1.0 indicates that the EQ-5D-Y-5L has the same efficiency as the EQ-5D-Y at detecting differences in health status using the external indicator (self-reported general health or the PedsQL™ 4.0 total scale score in this study); Relative efficiency >1.0 indicates that the EQ-5D-Y-5L is more efficient than the EQ-5D-Y at detecting differences in health status using the external indicator; and Relative efficiency <1.0 indicates that the EQ-5D-Y-5L is less efficient than the EQ-5D-Y at detecting differences in health status using an external indicator.(314)

In order to calculate the relative efficiency statistic, self-reported general health and PedsQL™ 4.0 total scale scores were dichotomised. A frequency distribution determined the cut-off (314) for dichotomising self-reported general health status in two: i) excellent or very good versus good or fair or poor, and ii) excellent versus very good or good or fair or poor. The mean for the total scale scores provided a cut-off for the PedsQL™ 4.0 such that below mean was one category, and mean and above was another category.

All empirical validity analyses were based on participants who completed both the EQ-5D-Y-5L and EQ-5D-Y. All missing responses for the EQ-5D-Y-5L and EQ-5D-Y measures had a value of 9 imputed and were not included in this analysis as per EuroQol scoring algorithms. However, the scoring algorithm for PedsQL™ 4.0 allows for imputation of missing values to generate PedsQL™ 4.0 scores. Consequently, participants with some missing values on the PedsQL™ 4.0 were not excluded. In addition, there were concerns that the EQ-5D-Y would have much lower scores below the utility score of 0 compared to the EQ-5D-Y-5L since utility scores for the latter are based on more recent valuation model (313). Lower bound values would

lead to under predicting the value for the poorest heath states. To overcome this, the relative efficiency statistic was re-calculated by restricting utility scores between 0 and 1 only.

6.5 Results

6.5.1 Participant characteristics

In total, 289 participants completed the EQ-5D-Y, EQ-5D-Y-5L, and PedsQL™ 4.0, aged 8-17 years (mean 13.6, median 14) as presented in last chapter and re-presented in Table 6.1.

Table 6.1 Participant characteristics

Characteristic		N (%)	Age group 1 (8-12yrs)	Age group 2 (13-17yrs)
Participants		289	98	191
Gender*	Male	121 (44%)	39	82
	Female	153 (56%)	51	102
Health condition	healthy	95 (33%)	12	83
	acute	155 (54%)	85	70
	chronic	39 (13%)	1	38
Grade#	1-5	71 (25%)	53	18
	6-8	97 (35%)	40	57
	9-12	111 (40%)	0	111

*missing data: 15 (age group 1 =8, age group 2=7); #missing data: 10 (5 in each age groups)
Age group 1 completed EQ-5D-Y, EQ-5D-Y-5L and PedsQL 4.0 child self-report
Age group 2 completed EQ-5D-Y, EQ-5D-Y-5L and PedsQL 4.0 teen self-report

6.5.2 Item performance

Overall, all items for the EQ-5D-Y passed the criterion for missing data, two failed maximum endorsement (mobility and looking after oneself), all passed aggregate endorsement frequency and item redundancy. (Table 6.2) For the EQ-5D-Y-5L, all items passed the criterion for missing data, one item failed maximum endorsement (looking after oneself), all failed the aggregate endorsement frequency criterion and all passed item redundancy. (Table 6.2)

Table 6.2 Item performance for EQ-5D-Y and EQ-5D-Y-5L for combined sample

Measure	Dimension	Item performance*			
		MD	**MEF**	**AEF**	**IR**
EQ-5D-Y	Mobility		x		
	Looking after oneself		x		
	Doing usual activities				
	Having pain or discomfort				
	Feeling worried, sad or unhappy				
EQ-5D-Y-5L	Mobility			x	
	Looking after oneself		x	x	
	Doing usual activities			x	
	Having pain or discomfort			x	
	Feeling worried, sad or unhappy			x	

* x represent dimension failed on the criteria; MD: Missing data; MEF: Maximum endorsement frequency; AEF: Adjacent aggregate endorsement frequency; IR: Item redundancy

Table 6.3 shows analysis by age. All items failed criterion for missing data for the EQ-5D-Y among children compared to none failing among adolescents. While no items failed on maximum endorsement frequency among children, three items failed criterion among adolescents. All items passed aggregate endorsement frequency criterion among and only one item failed among adolescents. No item failed on the item redundancy criterion in both age groups.

For the EQ-5D-Y-5L, three items in children compared to one among adolescents failed criterion for missing data. None of the items failed maximum endorsement frequency criterion in children compared to two that failed in adolescents. While all items failed maximum endorsement frequency criterion, all items passed item redundancy in both age groups.

Table 6.3 Item performance for EQ-5D-Y and EQ-5D-Y-5L by age

Measure	Dimension	Item level performance *							
		MD		**MEF**		**AEF**		**IR**	
		8-12yrs	13-18yrs	8-12yrs	13-18yrs	8-12yrs	13-18yrs	8-12yrs	13-18yrs
EQ-5D-Y	Mobility	x			x				
	Looking after oneself	x			x		x		
	Doing usual activities	x			x				
	Having pain or discomfort	x							
	Feeling worried, sad or unhappy	x							
EQ-5D-Y-5L	Mobility				x	x	x		
	Looking after oneself				x	x	x		
	Doing usual activities	x				x	x		
	Having pain or discomfort	x	x			x	x		
	Feeling worried, sad or unhappy	x				x	x		

* x represent dimension failed on the criteria; MD: Missing data; MEF: Maximum endorsement frequency; AEF: Adjacent aggregate endorsement frequency; IR: Item redundancy

The item level analysis based on health status of the participants for both the EQ-5D-Y and EQ-5D-Y-5L are presented in Table 6.4. For the EQ-5D-Y, missing data was not observed in healthy and chronic children, but two items failed criterion among the acutely ill. While all items failed maximum endorsement frequency among the healthy and chronically ill, none failed among the acutely ill. None of the items failed aggregate endorsement frequency among healthy and acutely ill children. This was, however, observed in the chronically ill children. No item failed on item redundancy among the healthy and acutely ill, but this was observed in one item (having pain or discomfort) among the chronically ill.

For the EQ-5D-Y-5L, missing data was similarly not observed in healthy and chronic children, but one item failed criterion among the acutely ill. Three and five items failed maximum endorsement frequency among the healthy and chronically ill respectively, but none failed among the acutely ill. Whereas all items failed aggregate endorsement frequency among healthy and acutely ill, this was not observed in the chronically ill children. No item failed on item redundancy among the healthy and acutely ill, but this was observed in all items among the chronically ill.

Finally, item level performance was also assessed for the EQ VAS. For both the EQ-5D-Y and EQ-5D-Y-5L versions, there was clear evidence of clustering response around the five and ten cut off points. (Table 6.5)

Table 6.4 Item performance for EQ-5D-Y and EQ-5D-Y-5L by health status

Measure	Dimension	Item level performance*											
		Missing data			Maximum endorsement frequency			Adjacent endorsement frequency			Item redundancy		
		healthy	chronic	acute	healthy	chronic	acute	healthy	chronic	acute	healthy	chronic	acute
EQ-5D-Y	Mobility				x	x			x				
	Looking after oneself			x	x	x							
	Doing usual activities				x	x							
	Having pain or discomfort				x	x			x			x	
	Feeling worried, sad or unhappy			x	x	x							
EQ-5D-Y-5L	Mobility				x	x		x		x		x	
	Looking after oneself				x	x		x		x		x	
	Doing usual activities			x	x	x		x		x		x	
	Having pain or discomfort					x		x		x		x	
	Feeling worried, sad or unhappy					x		x		x		x	

* x represent dimension failed on the criteria

Table 6.5 EQ VAS distribution

	EQ-5D-Y		EQ-5D-Y-5L	
EQ VAS *	**Frequency**	**Percent**	**Frequency**	**Percent**
1	1	0.3	1	0.3
20	3	1.0	1	0.3
30	0	0	3	1.0
35	1	0.3	1	0.3
40	2	0.7	7	2.4
45	2	0.7	3	1.0
50	9	3.1	8	2.8
55	4	1.4	3	1.0
60	8	2.8	4	1.4
65	1	0.3	2	0.7
69	1	0.3	0	0
70	11	3.8	11	3.8
75	8	2.8	7	2.4
76	1	0.3	0	0
80	18	6.3	22	7.7
85	16	5.6	10	3.5
86	2	0.7	2	0.7
87	1	0.3	1	0.3
89	0	0	1	0.3
90	43	15.0	34	11.8
91	1	0.3	0	0
92	1	0.3	0	0
94	1	0.3	0	0
95	37	12.9	46	16.0
96	0	0	2	0.7
97	0	0	1	0.3
98	3	1.0	2	0.7
99	6	2.1	7	2.4
100	89	31.1	91	31.7
Sub-Total	270	94.4	270	94.1
Missing	16	5.6	17	5.9
Total	286	100.0	287	100.0

*orange colour represents 2-5% responses; yellow represent ≥5% responses

6.5.3 Reliability

Internal consistency coefficients are presented in Table 6.6. For the EQ-5D-Y, the overall Cronbach's alpha (0.766) was above the 0.70 threshold. All the dimensions were within the acceptable range of 0.25 – 0.8 for corrected item-total correlation. Additionally, the individual dimension Cronbach's alpha if item deleted ranged between 0.688 – 0.762 against the total scale Cronbach's alpha of 0.766

Similarly, for the EQ-5D-Y-5L, the overall Cronbach's alpha ((0.838) was above the 0.70 threshold. All the dimensions were within the acceptable range of 0.25 – 0.8. The individual dimension Cronbach's alpha if item deleted ranged between 0.783 – 0.819 against the total scale Cronbach's alpha 0.838. All this goes on to show the internal consistency reliability of using both the EQ-5D-Y and EQ-5D-Y-5L Chichewa versions in Malawi.

Table 6.6 Internal consistency

Measure	Dimension (n)	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	change (standardized alpha- Cronbach alpha if item deleted)	Cronbach's Alpha Based on Standardized Items
EQ-5D-Y	(n=264, excluded 22)				**0.766**
	Mobility	0.531	0.711	0.055	
	Looking after oneself	0.552	0.707	0.059	
	Usual activities	0.588	0.688	0.078	
	Pain or discomfort	0.414	0.762	0.004	
	Worried, sad or unhappy	0.576	0.691	0.075	
EQ-5D-Y-5L	(n=277, excluded 10)				**0.838**
	Mobility	0.569	0.819	0.019	
	Looking after oneself	0.648	0.798	0.040	
	Usual activities	0.701	0.783	0.055	
	Pain or discomfort	0.599	0.814	0.024	
	Worried, sad or unhappy	0.686	0.787	0.051	

6.5.4 Validity

6.5.4.1 Convergent validity

Results of convergent validity are summarised in Table 6.7 below. For the EQ-5D-Y, there was correlation above the hypothesized threshold between EQ-5D-Y sum scores and PedsQL™ 4.0 total scale score: -0.392 with PedsQL™ 4.0 child self-report and -0.399 with PedsQL™ 4.0 teen self-report. The correlation between EQ-5D-Y dimensions and PedsQL™ 4.0 health summary scores was mixed. The EQ-5D-Y mobility dimension and PedsQL™ 4.0 Physical health summary score correlation was lower than hypothesized: -0.254 with PedsQL™ 4.0 child self-report; and -0.343 with PedsQL™ 4.0 teen self-report. The correlation between EQ-5D-Y feeling worried, sad or unhappy dimension and PedsQL™ 4.0 Psychosocial health summary score was as hypothesized: -0.449 with PedsQL™ 4.0 child self-report; but lower: -0.259 with PedsQL™ 4.0 teen self-report .

The EQ-5D-Y-5L sum score and PedsQL™ 4.0 total scale score correlation was similarly as hypothesized-0.397 with PedsQL™ 4.0 child self-report and -0.375 with PedsQL™ 4.0 teen self-report. The correlation between EQ-5D-Y-5L dimensions and PedsQL™ 4.0 health summary scores was mixed. The EQ-5D-Y-5L mobility dimension and PedsQL™ 4.0 Physical health summary score correlation was lower than hypothesized: -0.265 with PedsQL™ 4.0 child self-report; but as hypothesized: -0.391 with PedsQL™ 4.0 teen self-report. The correlation between EQ-5D-Y-5L feeling worried, sad or unhappy dimension and PedsQL™ 4.0 Psychosocial health summary score was as hypothesized: -0.359 with PedsQL™ 4.0 child self-report but lower: -0.341 with PedsQL™ 4.0 teen self-report .

.

Table 6.7 Convergent validity of the EQ-5D-Y and EQ-5D-Y-5L with PedsQL™ 4.0 self-report sub-scale.

		PedsQL 4.0 child self-report			PedsQL 4.0 teen self-report		
Measure	Dimension / Sub-scale	Psychosocial Health Summary	Physical Health Summary	Total Scale score	Psychosocial Health Summary	Physical Health Summary	Total Scale score
EQ-5D-Y	Mobility	-.301**	-.254*	-.312**	-.371**	-.343**	-.380**
	Looking after oneself	-.240*	-.250*	-.276*	-.212**	-0.117	-.163*
	Doing usual activities	-.344**	-.316**	-.376**	-.266**	-.213**	-.252**
	Having pain or discomfort	-0.122	-0.138	-0.146	-.190**	-.155*	-.172*
	Worried, sad or unhappy	-.**449****	-.303**	-.432**	-.259**	-.283**	-.291**
	EQ-5D-Y sum score	-.347**	-.351**	**-.392****	-.406**	-.353**	**-.399****
EQ-5D-Y-5L	Mobility	-0.148	-.265*	-0.202	-.455**	**-.391****	-.440**
	Looking after oneself	-.330**	-.325**	-.370**	-.156*	-0.056	-0.092
	Doing usual activities	-.221*	-0.177	-.229*	-.309**	-.211**	-.260**
	Having pain or discomfort	-.328**	-.365**	-.377**	-.357**	-.267**	-.314**
	Worried, sad or unhappy	**-.359****	-.276**	-.353**	-.341**	-.365**	-.379**
	EQ-5D-Y-5L sum score	-.362**	-.366**	**-.397****	-.405**	-.324**	**-.375****

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

6.5.4.2 Discriminant validity

There was no statistically significant difference between EQ-5D-Y or EQ-5D-Y-5L sum scores and gender: EQ-5D-Y (mean = 0.036, $t = 1.799, p < 0.074$) and the EQ-5D-Y-5L (mean = 0.05, $t = 0.143, p < 0.886$) among adolescents. There was similar no significance between sum scores and gender in children. (Table 6.8)

Table 6.8 Discriminant validity by gender for EQ-5D-Y and EQ-5D-Y-5L sum score

Measure	Children (n= 98, male= 38, female=50)			Adolescents (n= 182, male= 81, female=101)		
	mean difference	t-test		mean difference	t-test	
		t-statistic	p-value		t-statistic	p-value
EQ-5D-Y	-0.27	-0.627	0.532	0.36	1.799	0.074
EQ-5D-Y 5L	0.39	0.558	0.578	0.05	0.143	0.886

There was low correlation between age and the sum scores of both the EQ-5D -Y, and EQ-5D-Y-5L (range 0.113 – 0.173) showing age was not associated with scores. The correlation between age and EQ-5D-Y in adolescents was borderline (-0.201) (Table 6.9)

Table 6.9 Discriminant validity by age for EQ-5D-Y and EQ-5D-Y-5L sum score

Measure	Score	Correlation coefficient	
		Children	Adolescents
EQ-5D-Y	Sum score	0.113	**-0.201**
EQ-5D-Y-5L	Sum score	0.173	-0.185

While there was no evidence of difference between the EQ-5D-Y and EQ-5D-Y-5L sum scores and grade categories in the younger age group, there was a statistically significant association of sum scores and grade categories in adolescents ($p<0.05$). When age was combined, the association was found to be statistically significant between EQ-5D-Y and EQ-5D-Y-5L sum scores with grade categories ($p<0.001$). The lack of association between grade and EQ-5D-Y and EQ-5D-Y-5L sum score in children would indicate that grade had no effect on scores but not among adolescents and when age was combined. (Table 6.10)

Table 6.10 EQ-5D-Y and EQ-5D-Y-5L sum score discriminant validity by grade

Measure		Age 8-12 years (grade* grp1=53, grade grp 2=40)				Age 13-18yrs years (grade* grp1=18, grade grp 2=57, grade grp3=111)				Age 7-18yrs years (grade* grp1=71, grade grp 2=97, grade grp3=111)			
		df	Mean Square	F	p-value	df	Mean Square	F	p-value.	Df	Mean Square	F	p-value.
EQ-5D-Y sum score	Between Groups	1	0.389	0.114	0.737	2	5.122	3.074	**0.049**	2	23.110	10.335	**0.000**
	Within Groups	79	3.420			171	1.667			252	2.236		
	Total	80				173				254			
EQ-5D-Y-5L sum score	Between Groups	1	1.539	0.146	0.703	2	28.585	4.741	**0.010**	2	56.435	7.491	**0.001**
	Within Groups	89	10.530			175	6.030			266	7.534		
	Total	90				177				268			

*grade group1 = grade 1-5; group2 = grade 6-8; group3 = grade 9-12

6.5.4.3 Known-group validity

In children, effect size was low (0.23) for EQ-5D-Y-5L compared to high (-1.15) for the EQ-5D-Y. In adolescents, effect sizes were generally much higher (>0.5) suggesting reasonably good known-groups validity. (Table 6.11)

Table 6.11 EQ-5D-Y and EQ-5D-Y-5L sum score known group validity

	Age 8-12 years (N=98, healthy=12, 81= sick) *					Age 13-18 years (N=191, healthy=83, 106= sick) *				
	t-statistic #					t-statistic #				
Measure	t	p-value	Mean Diff	SD	Effect size ¶	t	p-value	Mean Diff	SD	Effect size ¶
EQ-5D-Y 5L sum score	-0.850	0.398	-0.957	4.196	-0.23	-3.480	0.001	-1.262	1.220	**-1.03**
	-0.751	0.466	-0.957	3.556		-3.715	0.000	-1.262	3.088	
EQ-5D-Y sum score	-2.154	0.034	-1.391	1.206	**-1.15**	-2.757	0.006	-0.536	0.884	**-0.61**
	-3.172	0.005	-1.391	2.083		-2.856	0.005	-0.536	1.566	

*2 and 5 did not complete EQ-5D-Y and EQ-5D-Y-5L respectively
assuming equal variance
¶ effect size designated as <0.2 poor, 0.3-0.49 small, 0.5-0.8 moderate, and >0.8 large

6.5.4.4 Empirical validity

Table 6.12 presents the relative efficiency statistics for the EQ-5D-Y and EQ-5D-Y-5L over the dichotomous self-reported general health status, PedsQL™ 4.0 HRQoL measure, and EQ VAS. When the EQ-5D-Y was referenced at 1.0, the EQ-5D-Y-5L was 31% (relative efficiency statistic of 0.693 versus 1.0) to 91% less efficient than the EQ-5D-Y at detecting differences in self-reported general health. Again, when the EQ-5D-Y was referenced at 1.0, the EQ-5D-Y-5L was 5% to 44% less efficient than the EQ-5D-Y at detecting differences in the PedsQL™ 4.0 total scale score. Similarly, the EQ-5D-Y-5L was 14% to 67% less efficient than the EQ-5D-Y at detecting differences against the EQ VAS. For both outcomes (self-reported general health and PedsQL™ 4.0 total scale score), the EQ-5D-Y-5L did not perform as hypothesized.

Restricting the analyses to children reporting utility scores between 0 and 1 reduced the sample size by four. Despite this compression of the utility scale, the EQ-5D-Y-5L continued not to perform as hypothesized against the EQ-5D-Y. (Table 6.13) Except among adolescents when self-reported general health was dichotomised between excellent or very good versus good or fair did EQ-5D-Y-5L perform as hypothesized (736% more efficient than the EQ-5D-Y).

Table 6.12 Efficiency of the EQ-5D to detect differences in self-reported health status

Measure	Age	Categorisation	Utility score[#]		t-test*		Relative efficiency
			mean	(SD)	t-statistic	p-value	
		SRGH					
EQ-5D-Y	age 8-12yrs (n=81)	excellent or v. good	0.838	0.222	2.075	0.041	1.000
		good or fair	0.746	0.175			
EQ-5D-Y 5L		excellent or v. good	0.812	0.328	0.510	0.612	0.060
		good or fair	0.780	0.243			
EQ-5D-Y		excellent	0.872	0.200	2.197	0.033	1.000
		v. good, good, fair or poor	0.766	0.205			
EQ-5D-Y 5L		excellent	0.832	0.321	0.660	0.513	0.090
		v. good, good, fair or poor	0.783	0.284			
EQ-5D-Y	age 13-17yrs (n=172)	excellent or v. good	0.903	0.137	0.148	0.883	1.000
		good or fair	0.899	0.150			
EQ-5D-Y 5L		excellent or v. good	0.911	0.175	0.123	0.902	0.693
		good or fair	0.907	0.160			
EQ-5D-Y		excellent	0.933	0.116	2.205	0.029	1.000
		v. good, good, fair or poor	0.887	0.150			
EQ-5D-Y 5L		excellent	0.924	0.210	0.704	0.483	0.102
		v. good, good, fair or poor	0.902	0.147			
EQ-5D-Y	combined ages 7-17yrs (n=253)	excellent or v. good	0.883	0.169	1.733	0.085	1.000
		good or fair	0.844	0.175			
EQ-5D-Y 5L		excellent or v. good	0.881	0.235	0.702	0.484	0.164
		good or fair	0.862	0.202			
EQ-5D-Y		excellent	0.913	0.150	3.027	0.003	1.000
		v. good, good, fair or poor	0.848	0.178			
EQ-5D-Y 5L		excellent	0.895	0.253	0.945	0.346	0.098
		v. good, good, fair or poor	0.864	0.208			
		PedsQL score mean					
EQ-5D-Y	age 8-12yrs (n=81)	≥ 72.79	0.840	0.159	2.298	0.025	1.000
		< 72.79	0.727	0.249			
EQ-5D-Y 5L		≥ 72.79	0.865	0.223	2.237	0.030	0.948
		< 72.79	0.705	0.363			
EQ-5D-Y	age 13-17yrs (n=172)	≥ 78.68	0.946	0.087	3.837	0.000	1.000
		< 78.68	0.864	0.167			
EQ-5D-Y 5L		≥ 78.68	0.947	0.151	2.863	0.005	0.557
		< 78.68	0.872	0.177			
EQ-5D-Y	combined ages 7-17yrs (n=253)	≥ 76.81	0.918	0.114	4.716	0.000	1.000
		< 76.81	0.812	0.210			
EQ-5D-Y 5L		≥ 76.81	0.929	0.170	4.102	0.000	0.756
		< 76.81	0.808	0.265			
		EQ VAS mean					
EQ-5D-Y		≥ 84	0.862	0.134	4.881	<.001	1.000
		< 84	0.661	0.223			

EQ-5D-Y-5L	age 8-12yrs (n=78)	≥ 84 < 84	0.869 0.679	0.211 0.299	3.205	0.002	0.431
EQ-5D-Y	age 13-17yrs (n=171)	≥ 89 < 89	0.924 0.840	0.118 0.216	3.176	<.001	1.000
EQ-5D-Y-5L		≥ 89 < 89	0.935 0.840	0.125 0.287	2.936	0.002	0.855
EQ-5D-Y	combined ages 7-17yrs (n=253)	≥ 87 < 87	0.911 0.773	0.121 0.227	6.093	<.001	1.000
EQ-5D-Y-5L		≥ 87 <87	0.918 0.787	0.158 0.290	4.517	<.001	0.550

[#] US tariffs
*assuming equal variance
SRGH- self-reported general health

Table 6.13 Efficiency of the EQ-5D to detect differences in self-reported health status (utility set to between 0 and 1 only for both EQ-5D-Y and EQ-5D-Y-5L)

MEASURE	AGE	Categorisation	Utility score[#]		t-test*		Relative efficiency
			mean	(SD)	t-statistic	p-value	
		SRGH					
EQ-5D-Y	age 8-12yrs (n=78)	excellent or v. good	0.855	0.175	2.521	0.014	1.000
		good or fair	0.755	0.171			
EQ-5D-Y-5L		excellent or v. good	0.863	0.219	1.206	0.232	0.229
		good or fair	0.805	0.200			
EQ-5D-Y		excellent	0.867	0.202	1.688	0.099	1.000
		v. good, good, fair or poor	0.789	0.164			
EQ-5D-Y 5L		excellent	0.870	0.262	0.786	0.437	0.217
		v. good, good, fair or poor	0.824	0.185			
EQ-5D-Y	age 13-17yrs (n=171)	excellent or v. good	0.904	0.138	0.184	0.854	1.000
		good or fair	0.899	0.150			
EQ-5D-Y-5L		excellent or v. good	0.920	0.140	0.532	0.596	8.360
		good or fair	0.907	0.160			
EQ-5D-Y		excellent	0.935	0.116	2.306	0.023	1.000
		v. good, good, fair or poor	0.887	0.150			
EQ-5D-Y-5L		excellent	0.945	0.144	1.793	0.076	0.605
		v. good, good, fair or poor	0.902	0.147			
EQ-5D-Y	combined ages 7-17yrs (n=249)	excellent or v. good	0.890	0.151	1.900	0.059	1.000
		good or fair	5.000	0.171			
EQ-5D-Y 5L		excellent or v. good	0.904	0.169	1.393	0.165	0.538
		good or fair	0.871	0.181			
EQ-5D-Y		excellent	0.914	0.151	2.736	0.007	1.000

		v. good, good, fair or poor	0.856	0.160			
EQ-5D-Y-5L		excellent	0.921	0.191	1.744	0.084	0.406
		v. good, good, fair or poor	0.878	0.163			
		PedsQL score mean					
EQ-5D-Y	age 8-12yrs	≥ 73.10	0.866	0.129	3.255	0.002	1.000
		< 73.10	0.733	0.206			
EQ-5D-Y-5L	(n=78)	≥ 73.10	0.908	0.148	3.103	0.003	0.909
		< 73.10	0.754	0.252			
EQ-5D-Y	age 13-17yrs	≥ 78.64	0.947	0.086	3.917	0.000	1.000
		< 78.64	0.864	0.167			
EQ-5D-Y-5L	(n=171)	≥ 78.64	0.960	0.085	3.963	0.000	1.024
		< 78.64	0.872	0.177			
EQ-5D-Y	combined ages 7-17yrs	≥ 76.93	0.923	0.107	4.986	0.000	1.000
		< 76.93	0.819	0.191			
EQ-5D-Y-5L	(n=249)	≥ 76.93	0.945	0.109	5.096	0.000	1.045
		< 76.93	0.829	0.212			
		EQ VAS mean					
EQ-5D-Y	age 8-12yrs	≥ 84	0.863	0.137	4.305	<.001	1.000
		< 84	0.687	0.211			
EQ-5D-Y-5L	(n=78)	≥ 84	0.869	0.167	3.245	<.001	0.568
		< 84	0.679	0.267			
EQ-5D-Y	age 13-17yrs	≥ 89	0.921	0.119	2.809	0.006	1.000
		< 89	0.851	0.183			
EQ-5D-Y-5L	(n=171)	≥ 89	0.933	0.127	2.295	0.023	0.668
		< 89	0.874	0.186			
EQ-5D-Y	combined ages 7-17yrs	≥ 88	0.911	0.121	5.685	<.001	1.000
		< 88	0.790	0.202			
EQ-5D-Y-5L	(n=249)	≥ 88	0.924	0.136	4.536	<.001	0.637
		< 88	0.817	0.225			

[#] US tariffs
*assuming equal variance
SRGH- self-reported general health

6.6 Discussion

In this setting, the Chichewa versions demonstrated good internal consistency, and some mixed evidence of item performance and validity. Overall, among adolescents both the EQ-5D-Y and EQ-5D-Y-5L showed good evidence of reliability (internal consistency), convergent validity, discriminant validity and known-groups validity. The validity evidence was a little more mixed among children. The EQ-5D-Y outperformed the EQ-5D-Y-5L in empirical validity. Although reliability results are acceptable, there are some issues with item performance especially

endorsement (maximum endorsement for EQ-5D-Y and adjacent aggregate endorsement for EQ-5D-Y-5L) and redundancy for EQ-5D-Y-5L. This suggests that the response options are possibly not working as intended and also dimensions may not be making unique contributions to the scale. In general, both Chichewa versions demonstrated that they can be used with some limitations in children and adolescents in this setting. In addition, other psychometric properties like test-retest and responsiveness also need to be evaluated in this setting. There were also more issues with self-completion (from missing data) of the instruments especially in younger children aged ≤12years.

The item performance shows that both the EQ-5D-Y and EQ-5D-Y-5L had no item that is redundant. Adolescents in the combined sample ably self-completed the questionnaires with no missing data on all items. This generally lack of missing data for the EQ-5D-Y confirms what has been observed in several studies including the original EQ-5D-Y validation study. (160, 204, 301) The adult EQ-5D-3L and EQ-5D-5L versions have similarly demonstrated negligible missing values. (315) However, although this was the case in adolescents, missing data was found to be a problem in children and in acute participants in this study. This would indicate difficulty in providing good quality HRQoL data among the younger population when using self-completion questionnaires. The tendency to pick responses from the top end of an item (ceiling effect) has been less evident in adolescents compared to children in most studies, (316) although this study found the opposite (for both instruments maximum endorsement was observed in adolescents but not in children). The adult EQ-5D-5L has been found to reduce answering from top end of a scale compared to the EQ-5D-3L,(315, 317) and similarly in this study the EQ-5D-Y-5L edged the EQ-5D-Y. When stratified by age, the EQ-5D-Y performed better than the EQ-5D-Y-5L. Further, the EQ-5D-Y performed better than EQ-5D-Y-5L with no dimension failing the adjacent aggregate endorsement frequency. A possible explanation

for this is that since the EQ-5D-Y-5L has five response levels, some of the responses might not have been relevant to this population. Using the five response level, and related to this is the actual dimension content, among a healthy population seem to be problematic as some of the response options might not be relevant to them. This finding is comparable to what has been reported elsewhere. (204, 316) One previous study suggested that the use of a 3-point response scale is better to achieving reliability of a scale. (291) This, it was argued, gets rid of responses that are irrelevant in a scale. However, doing so would be in direct contrast to the intention of the EQ-5D-Y-5L which was developed precisely to increase sensitivity (by increasing response options) in health status reporting. (259) It might be reasoned that perhaps the EQ-5D-Y-5L is doing just that.

The clustering of EQ VAS observed around the five and ten cut off points would suggest that not all points on the EQ VAS are fully utilized. A suggestion would be to perhaps have the marks only around five and ten cut off points (e.g., 5, 10, 15, 20, 25 and so on) although a single item instrument has well been established in psychometrics to be problematic. (278) Recalibrating the cut-off points would therefore remove some but not all associated problems.

The evidence for construct validity shows that criteria for convergent validity were met at scale but not dimension level. The fact that PedsQL™ 4.0 Physical Functioning health summary scores had highest correlations with the EQ-5D-Y and EQ-5D-Y-5L sum score would suggest that even though there are emotional and pain dimensions, the EQ-5D-Y and EQ-5D-Y-5L are primarily physical measures. This might imply that the EQ-5D-Y and EQ-5D-Y-5L are best suited to assess physical functioning as opposed to other aspects of HRQoL. Additionally, while the adult EQ-5D-5L has been found to be consistently associated with high correlation with other health measure compared to the EQ-5D-3L, (318-320) this study found this not to

be the case between the five and three severity levels of the youth versions. Both the EQ-5D-Y and EQ-5D-Y-5L equally demonstrated convergent validity with PedsQL™ 4.0.

There were no differences between EQ-5D-Y and EQ-5D-Y-5L sum score association with both gender and age. The adult EQ-5D-3L and EQ-5D-5L have equally demonstrated discriminant validity with gender in young adults in a previous study. (321) Additionally, there has been observed little difference in the discriminant validity between the two measures. (317) This research confirms these findings although there was a negative correlation observed between scores and age among adolescents and a positive one in children. This study could not establish why this was the case and might be a point for further research. While age has been associated with different scores depending on the EQ-5D-3L or EQ-5D-5L, (317) this study did not find such a difference between the EQ-5D-Y and EQ-5D-Y-5L.

Although discriminant validity criteria between both the EQ-5D-Y and EQ-5D-Y-5L and grade was met in children aged 8-12 years, this was not the case among adolescents. The statistical significance association among adolescents, and when age was combined needs some consideration. It possibly indicates that there is a relationship between grade and getting good HRQoL data. It is likely the case that perhaps the fact that older children or education exposure contributes to expressing or general comprehension of health. This may be due confidence and experience with completing questionnaire and similar formats for older children.

As hypothesized, both the EQ-5D-Y and EQ-5D-Y-5L were able to distinguish between the healthy and sick scores using the known-groups method. There were differences between the scores of the healthy and sick children across different health conditions as evidenced by the moderate to large effect size. While the EQ-5D-Y had the largest effect size in children, the

EQ-5D-Y-5L was able to discriminate better in the adolescent population. The adult EQ-5D-5L five severity level ability to distinguish between groups based on age has been observed in another study. (322) The findings from this study confirms that the five severity level EQ-5D-Y-5L is best suited in older population probably since they have better comprehension of the severity levels.

Finally, it was hypothesized that the EQ-5D-Y-5L which has five severity level would perform better at predicting stated preference compared to the EQ-5D-Y. However, EQ-5D-Y outperformed the EQ-5D-Y-5L against this hypothesis in establishing empirical validity. This finding is rather surprising since the EQ-5D-Y-5L has five response level and therefore derives more health states compared to the EQ-5D-Y (3,125 compared to 243). Elsewhere, the SF-6D has performed better than the EQ-5D-3L because of the better sensitivity with the former. (314) Additionally, there were more adolescents and therefore should have been expected to understand the EQ-5D-Y-5L questionnaire better. That said, the fact that the participants were relatively healthy (even among those sick) could possibly have resulted in better response for the EQ-5D-Y compared to EQ-5D-Y-5L. In a head to head comparison of the adult versions of the EQ-5D-3L and EQ-5D-5L, the former was found to overestimate health problems leading to underestimation of utilities. (299) This overestimation can lead to underestimation of utilities which could result in biases for the EQ-5D-3L. This might be the reason why the EQ-5D-Y outperformed better than the EQ-5-Y-5L. In general, the EQ-5D-5L has been found to demonstrate better relative efficiency compared to the EQ-5D-3L. (323-325) Ultimately, the full understanding of why the EQ-5D-Y performed better than the EQ-5D-Y-5L in predicting health scores is beyond this study. Future studies could test the hypothesis stated above.

The psychometric evaluation of the Chichewa EQ-5D-Y and EQ-5D-Y-5L versions has some limitations that should be mentioned. Firstly, data was only collected from one time point and consequently test-retest reliability as well as responsiveness could not be evaluated. This might have an effect in ascertaining the reliability of the Chichewa versions of the EQ-5D-Y and EQ-5D-Y-5L in a clinical setting. Future studies should systematically evaluate these properties among children and adolescents. Second, the EQ-5D-Y is recommended for age 8-15 years. However, for like comparison the whole age range of children and adolescents (8-17 years) completed both the EQ-5D-Y and EQ-5D-Y-5L to account for PedsQL™ 4.0 measures. While this is a limitation, it is unlikely that it may have affected the findings since the dimensions are very similar to those of the adult EQ-5D-3L version.

Another limitation is to do with the utilities used for empirical validity. Currently, there are no utility values for both the EQ-5D-Y and let alone for the EQ-5D-Y-5L beta version. The empirical validity was therefore based on adult tariffs and these were also from a different geographical location. This may have affected the utility scores generated in this research. Ideally, the utilities for both the EQ-5D-3L and EQ-5D-5L should have been preferred from sSA region but no country utilities are available (at the time of this research). The EuroQol Group recommended use of adult US utilities instead for easy comparison as these are available for both the three and five response levels. However, even though these utilities are available for the USA, they were derived using two different approaches. This is an area that need further research not only in sSA region but also globally.

6.7 Conclusion

This is the first study to psychometrically validate the EQ-5D-Y-5L adapted for Malawi. Further, it is the first such work to undertake a psychometric performance comparison of the

EQ-5D-Y with the EQ-5D-Y-5L. The EQ-5D-Y established internal consistency among children and adolescents but was not able to assess test-retest reliability. The EQ-5D-Y had problems with item performance and known-groups validity. The EQ-5D-Y-5L similarly demonstrated internal consistency and validity except discriminant validity with grade. The study found both the EQ-5D-Y and the EQ-5D-Y-5L have problems with missing data in younger children. In addition, both the EQ-5D-Y and the EQ-5D-Y-5L have problems with endorsement suggesting that the response options are not working as intended in this setting. These issues are likely to be not specific to Malawi though as shown by evidence from elsewhere. However, some psychometric problems seen in Malawi are not caused by the translation/cultural but may be specific to the context.

In conclusion, there was not much difference between the performance of the EQ-5D-Y and the EQ-5D-Y-5L. Whereas the EQ-5D-Y-5L performed better than the EQ-5D-Y in discriminant validity, the latter outperformed the EQ-5D-Y-5L in the utilization of response options as well as empirical validity. Both instruments can be used among children and adolescents in Malawi although with some limitation. Further psychometric testing is required for both test re-test reliability and responsiveness which could not be carried out in this study.

6.8 Chapter summary

In this chapter the psychometric performance of the EQ-5D-Y and EQ-5D-Y-5L was performed. Psychometric evaluation is important in establishing whether an instrument is reliable and valid across population groups and culture. Both the EQ-5D-Y and EQ-5D-Y-5L performed well for internal consistency. Validity was mixed and item performance was varied between the two measures.

One other psychometric property not assessed in this quantitative evaluation is the content validity. This is assessed using qualitative methods and the following chapter 7, will now evaluate this in a Malawian setting. This will inform whether the EQ-5D instruments have all the necessary dimensions that children and adolescents consider important in this setting.

Findings based on the material in the chapter has been presented at an international conference.(326). A paper has been drafted for publication based on some of the material from this chapter. (327)

7 Chapter 7-Content validity of health-related quality of life scales: conceptual and cultural understanding of health amongst Malawian children

7.1 Chapter overview

In addition to the psychometric properties reviewed in chapter 6, a full psychometric appraisal needs to also evaluate content validity for which there is no statistical test. Content validity assesses whether the instrument adequately captures a complete range of conceptual ideas (represented by items) of the construct that the instrument claims to measure. The adult EQ-5D-3L (from which the EQ-5D-Y and EQ-5D-Y-5L have been developed) has generally struggled to demonstrate content validity as a generic measure in health research. (328, 329) While all the dimensions are relevant, there are other HRQoL aspects that have been found missing. Having emerged from the adult EQ-5D, with essentially the same content, it is likely that the EQ-5D-Y exhibits similar tendencies as regards content validity. The EQ-5D-Y has only five items and this is likely to lack conceptual breadth and validity.

Like most childhood preference-based HRQoL instruments, the EQ-5D-Y and EQ-5D-Y-5L have originated in HICs. The EQ-5D-Y has captured HRQoL from that perspective and need evaluation elsewhere. However, even in such HICs the EQ-5D instrument has been found to lack content of some HRQoL concepts. In a UK survey conducted in 2017, the general population (n=436) were asked to name some of dimensions missing from the EQ-5D if any, and 40.6% said that it was inadequate. (330) Sensory deprivation (vision and hearing), mental health (general or specific dementia), ability to communicate, ability to engage in relationships

and financial situation were some of the dimensions that were found to be missing. In a previous survey in New Zealand, among the 50% of the 1,360 participants who made comments regarding comprehensiveness of the EQ-5D, 29% mentioned that they found it inadequate. (331) The content validity of the EQ-5D has also been evaluated in some LMICs. In a most recent semi-structured interview in Singapore, (270) participants were asked to identify the most important domains to them, and also the relevance of the EQ-5D. While all the five EQ-5D dimensions were identified as important, suggestions were made to include social relationships, medical condition, treatment, health promotion and knowledge. It is possible that the issues to do with content and context may also arise in a sSA setting, like Malawi. The content of any of the EQ-5D instruments has to the best knowledge of this researcher, not been evaluated in the sSA region (at the time of doing this research).

Furthermore, having originated from the adult EQ-5D-3L version, the EQ-5D-Y may also fail to capture concepts from children and adolescent's perspective. A recent study with children and adolescents in Germany identified cognitive functioning as a dimension for inclusion in the EQ-5D-Y. (332) The findings in chapter 4 of this thesis also suggested that there could be other aspects of health that lie beyond those defined by the EQ-5D-Y. The issues raised above highlight the fact that the EQ-5D-Y exhibits similar tendencies as regards content validity, as the adult EQ-5D versions. All these issues points to the need for an investigation and a fuller discussion of the content and relevance of the EQ-5D-Y in Malawi.

This chapter explores whether the EQ-5D-Y and EQ-5D-Y-5L have all the necessary dimensions that children and adolescents consider important items in defining health in a Malawian setting. It reports a qualitative study to investigate the conceptual understanding of health from children and adolescent perspective in Malawi.

7.2 Introduction

What is known about health as a construct?

The concept of health is complex and as such takes many definitions, (80) as mentioned in the introductory part of chapter 2. The WHO 1947 definition of health as "a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity" (77) captures three dimensions: physical, mental and social. This definition has, however, created some debate as it has been suggested that it is impossible to define 'complete'. In parallel a slightly different set of concepts have also emerged in the literature known as QoL and HRQoL. (80) Like the concept of health these are generally considered to be multi-dimensional but usually involve the individual's subjective perception about the importance of each dimension. Thus, for the same objective amounts of function or health, HRQL can be rated very differently as two different people can experience the same set of circumstances differently. Like health, and as indicated in the same introductory section of chapter 2, there are many definitions for QoL and HRQL even though these concepts have been widely used in literature. (80) For example, the term "functional ability" is sometimes used interchangeably with "HRQL" even though there is a fairly clear distinction between the two terms. (114, 333) Some have suggested that the social dimension is best viewed outside of health and therefore should not be included in HRQoL. (333) Others contend that HRQoL is multi-dimensional and subjective, and therefore include the social dimension. As outlined in chapter 2 of this thesis, HRQoL may have similar attributes to QoL but the former principally focuses on how these affects someone's health.

In one of the earlier works on health status, life was defined as being made up of two components: quantity of life and quality of life. (334) The former was expressed by indicators

including average life expectancy and mortality rates. These indicators focus on biological functioning and not about the QoL. QoL was thus defined as going beyond just having a long healthy life and included issues around living standards, quality of housing, job satisfaction and health. The linkage between some aspects in life influencing a health condition and vice versa has led some to define these constructs by including factors that are not affected within the body and also to the construct of *health-related* quality of life; that is the aspects of QoL that are affected by a health condition. It is this perspective that has gained prominence in describing health as an outcome. (335)

Early health outcome measures

By the early 1960s, in the USA and other HICs, health care interventions for better QoL on chronic conditions such as cancer were gaining emphasis. (334) The national health surveys were being introduced to evaluate these different interventions and their impact on functional status as well as disease and treatment outcomes. The health surveys required questionnaire to measure the outcomes. A breakthrough came with the publication of a nominal 'blueprint' for health measurement by Fanshel and Bush in 1970. (336) Briefly, the paper tried to operationalize health definition on a continuum scale as a function of ability. A weight was assigned for the point on the scale belonging to an individual to produce a health state. This could be summed across each member of the population to give population health state. Multiplying the population health state over time would quantify health benefits from a programme or intervention. This was ground breaking work and provided theoretical basis for the development of such measures as Health Status Index (HSI), the Quality of Well-Being (QWB)(187), Rosser Kind index (RI)(337), Nottingham Health Profile (NHP)(338), and Sickness Index Profile (SIP) (339) among others. Another reason was that health economics was also an upcoming field for cost-effectiveness of health care interventions. (340) By 1987, the QWB, RI alongside 15D and SF-36, SIP, HUI, NHP were being used for health care

evaluation. (340)

Table 7.1 below shows a summary of the adult generic health measures developed by 1990. (see Appendix 20 for detailed description) A close examination shows these early measures were influenced (341) generally by the WHO definition of health. Health was captured as a multi-dimensional construct with several domains, with some measures also including behavior and attitudes. (342) To a large extent, these nine adult generic instruments measured health in terms of physical, mental and social aspects of health although there was variation in terminology used for descriptors. (340) Most of these measures had a profile index and varied in number of items from nine (The Dartmouth Cooperative Functional Assessment Charts (COOP)) to 136 (SIP). Technically there is a difference between profile and index in that the former has several items to capture a dimension score (profile), and in turn several dimension scores yield an overall scale score (index) to summarize HRQoL. (343)

Most of these instruments encompassed in the physical, mental and social trichotomy with some exceptions. The RI, for example, describes health in terms of disability and distress although these fall under physical and mental domains respectively. The 12D which was subsequently replaced by 15D had 12 domains, but these were described differently in comparison to ones outlined by the WHO health definition. The SIP and Duke Health Profile (DUHP) include both physical and social domains and a few additional categories aimed at capturing issues relating to sleep and rest, eating, work, home management, recreation, perceived health, and self-esteem which do not fall within the three predefined domains. In general, all adult measures include the physical functioning as a domain although this was described differently by the RI and 12D. Social domain is similarly included in all the other instruments except for the RI. Mental domain has a wide range of descriptors whether as part

of psychosocial domain (SIP), emotional reactions (Nottingham Health Profile (NHP)), emotional condition (COOP), social-emotional function (HUI1), distress (RI), anxiety and depression (DUHP). The QWB is the only instrument that does not have mental functioning whether at domain, sub-scale or item level.

In terms of conceptual framework, the adult measures were predominantly developed using a decomposed top-down approach. This means that developers were mostly guided by reviewing the health literature and seeking the opinions of experts in the field. The SIP and the NHP mark the beginning of the use of a more robust psychometric approach in health outcomes measurement and are the only two of the nine instruments that were developed using a bottom-up approach through qualitative work with patients. The bottom-up approach which include qualitative work is important to identify health aspects that the population of interest find important. As explained in chapter 3 of this thesis, qualitative work enhances comprehensiveness in understanding about a construct. For example, the SIP was developed by collecting statements that describe health from patients, carers and health care professionals. That process ensured that all relevant aspects of health from all involved or affected by illness were considered. The development also included steps to pilot test the tool to ensure only relevant items were retained. (339, 344) In addition, while there is scanty evidence of any psychometric assessment during initial development of some adult measures this was not the case with others. The SIP and NHP are two such measures that marked the beginning of more robust development of health instruments from a psychometric perspective. (345) As outlined in chapter 3, psychometrics ensures that there is a standardized way for assessing how an instrument perform. Additionally, psychometrics is necessary in establishing that the developed measure is robust and can lead to advances and practice. (346) In more recent years psychometric methods have been formulated into guidelines (145, 147) and subsequently used

in the actual selection and evaluation of items for health instruments. (146, 151)

Table 7.1 Pre-1995 adult generic health measures

Measure*	SIP	QWB	Rosser Kind	12 Dimensions	HUI1	NHP	COOP	DUHP	EQ-5D-3L
Ref	Bergner et al 1976; Bergner et al 1981	Kaplan et al 1976	Rosser 1976	Sintonen 1981	Torrance et al 1982	Hunt et al 1985	Nelson et al 1987; Nelson et al 1990	Parkerson et al 1990	The EuroQol Group 2000
Year developed	1976	1976	1976	1981	1982	1985	1987	1990	1990
Conceptual framework basis	qualitatively	literature review	expert review	Finnish medical records	literature review and epidemiologi-cal surveys	qualitative	reviewing clinical literature	Experts views.	literature review
Self-reported/ proxy reported/ interviewer administered	interviewer administered but participant answering questions only	interviewer administered	interviewer administered	self-reported	self-administered	self-reported	self-reported	self-reported	self-reported/ interviewer administered
Items	136	43 + symptoms	12	57	23	45	9	17	5
Domains#	2 dimensions (with 7 categories) + 5 independent categories	3	2	12	4 attributes	6 areas for part 1 (38 items) and daily life impact (7 statements) for part 2	9	6 health measures; 4 dysfunction measures (anxiety, depression, pain, and self-esteem)	5
	Independent categories: *sleep and rest*	**mobility** **physical activity**	**disability (8)** **distress (4)**	**perceived health** **breathing**	**physical function** **role function**	**sleep (5 items)** **physical mobility (8 items)**	**physical condition** **emotional condition**	**physical health** **mental health**	**mobility** **selfcare**

	eating	**social activity**		**sleeping**	**social-emotional function**	**energy (3 items)**	**daily work**	**social health**	**usual activities**
	work	**symptoms**		**speaking (communication)**	**health problems**	**pain (8 items)**	**social activities**	**general health (all 15 items, 5 each for physical, mental and social health)**	**pain/ discomfort**
	home management			**moving**		**emotional reactions (9 items)**	**pain**	**perceived health**	**anxiety/ depression**
	recreation and pastimes			**eating**		**social isolation (5 items)**	**change in condition**	**self-esteem**	
	1.Physical:			**incontinence**			**social support**	*anxiety*	
	ambulation			**hearing**		**how health impact daily life activities**	**quality of life (quality of life)**	*depression*	
	mobility			**seeing**				*pain*	
	body care and movement			**intellectual or mental functioning**				*disability*	
	2.Psychosocial:			**working**					
	social interaction			**social participation**					
	alertness								
	emotional behaviour								
	communication								

* SIP: sickness index profile; QWB: Quality well-being; HUI1: Health utility index mark 1; NHP: Nottingham health profile; COOP: The Dartmouth Cooperative Functional Assessment chart; DUHP: Duke Health Profile; #dimensions in bold

Differences in development between adult health and child health instruments

Development of adult measurements of health has preceded child measures due to among others the challenges of understanding children's perceptions about health and finding ways to elicit reliable and valid responses to questionnaires from children. (108) As discussed in chapter 2, development of children and adolescent health measures started to gain prominence in the early 1990s. (258) Table 7.2 outlines a summary (see Appendix 21 for detailed description) of childhood generic health measures developed before 1995. The cut-off of 1995 was chosen because firstly, this was the year the EQ-5D was officially named, and secondly, only two childhood measures (Child's Health Self-Concept Scale (CHSCS) and Children's Health Rating Scale (CHRS)), had been developed by 1990. It was reasoned that the later date would enable a few more childhood measures in order to make a meaningful comparison.

There is considerable similarity between childhood and adult health instruments in terms of the focus around capturing physical, mental and social aspects of health. Physical health is described variously by these measures as physical health (CHSCS), physical functioning (Child Health Questionnaire (CHQ)), functional capacity and physical state (Kinder Lebensqualität fragebogen (KINDL)). Social domain is also incorporated in the measures whether as role/social-physical concept (CHQ-PF50), social relationships (KINDL), or as play, friends (CHSCS). Mental well-being is similarly included in the three measures that provide detail of their content as emotional category (CHSCS); role/social-emotional/behavioral and, mental health (CHQ-PF50); or psychological well-being (KINDL). Further, there are also a few descriptors (personal grooming, dental health and nutrition) for childhood measures that do not directly fall within these three pre-defined dimensions of physical, mental and social categories. Notably some of these aspects of health such as personal grooming (CHSCS), bodily pain (CHQ) while described differently, have items that point to the physical health.

Table 7.2 Pre-1995 childhood generic health measures

Measure*	CHSCS	CHRS	CHQ-PF50	KINDL	EQ-5D-Y[$]
Ref	Hester NO et al 1994	Maylath et al 1990	Landgraf et al 1998; Landgraf and Abetz 1997	Ravens-Sieberer & Bullinger 1998	Wille et 2010
Year	1984	1990	1994	1994	2010
Conceptual framework	**qualitatively**	from existing measure.	literature review	semi-structured qualitatively	literature review
Self-reported/ Proxy/ Interviewer assisted	self-reported	self-reported	parent-completed version	self-reported	self-reported/ interviewer assisted
Items	45	17	50	40	5
Domains[#]	12 categories	not specified	2 (14 concepts: 13 + role/social-emotional/ behaviour counted as two concepts)	4	5
	Nutrition **Physical health** **Sleep** **Dental health** **Friends** **Healthiness** **Family** **Play** **Activity and exercise** **Personal grooming** **Emotional** **Nonspecific**		**Physical functioning** **Psychosocial functioning and well-being** *physical functioning (6)* *role/social-physical (2)* *general health perceptions (6)* *bodily pain (2)* *parental time impact (3)* *parental emotional impact (3)* *role/social-emotional/ behavioural (3)* *self-esteem (6)* *mental health (5)* *general behaviour (6)* *family activities (6)* *family cohesion (1)* *change in health (1)*	**Functional capacity in everyday life (11)** **Psychological well-being (11)** **Physical state (9)** **Social relationships (9)**	**Walking about** **Looking after myself** **Usual activities** **Pain or discomfort** **Worried, sad or unhappy**

*CHSCS: Child's Health Self-Concept Scale; CHRS: Children's health ratings scale; CHQ-PF50: Child Health Questionnaire-Parent Form 50; [$]developed in 2010 (from EQ-5D-3L); [#]dimensions in bold.

Content validity and utility instruments

Published conceptualization of health has differed, yet, there is some general consensus regarding what constructs a generic health status measure should contain. (347) The content of a measure to a large extent depends on what it purports to measure and the purpose for which it is to be developed. Likely to build on the WHO definition of health, almost all health instruments have primarily focused on measuring the physical functioning, mental and social functioning, as well as well behaviors and attitudes. (342) Not surprising is the fact that these concepts were common among these early measures.

Due to growing interest to measure the effectiveness of interventions, as previously described in chapter 2, health economists started to focus on measuring quality as well as quantity into a single unit. (348) Quantifying time and QoL into QALYs (also DALYs borne out of burden of diseases) have enabled a comparison of intervention effectiveness and thereby assisting in resource allocation decisions. Utility instruments have therefore not only assisted in measuring health improvement, but also in resource allocation. However, there has been ambiguity of developed instruments as to what they really measure, whether health, HRQoL, QoL or well-being. This ambiguity in the instruments can be seen in the constructs that they capture. For example, the SF-6D (derived from the SF-36) and EQ-5D, which are among the two most widely used generic preference based measures of health (144), have been described to measure health status, HRQoL and QoL. (80) Both the SF-6D and EQ-5D describe health in terms of index value which is not the same as health since the index refer to the value that individuals put on their health. (154)

The EuroQol Group initiated development of the EQ-5D content back in 1987. (349) Their objective was to come up with a single index score that would detect dimensional change but

also produce a net utility change across all dimensions. The group used knowledge from their respective work as well as a review of existing instruments available at the time: RI, NHP, SIP, QWB, 15-D, and Health and Measurement Questionnaire. (350) After the review, the EuroQol Group came up with the following six 'core' dimensions: mobility, self-care, main activity, social relationships, pain and mood energy. (349) Social relationships was taken out since it showed to contribute little to the valuation of health states and was therefore included in the main activity dimension. Following a further refinement of the descriptors, the remaining five dimensions were: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Attempt was made to separate pain and discomfort, and anxiety and depression but doing so would have led to a complex questionnaire for self-completion and valuation purposes. (350) Initially, the dimensions varied in the number of response levels (two to three) but later changed to having three levels across all dimensions. This was to give equal importance to all the dimensions. (351)

The EQ-5D was designed to have five single item dimensions in order to be quick to complete especially in large population studies. (352) However, there is still ongoing debate as to what the EQ-5D (and consequently the EQ-5D-Y) really measures. (340) A recent crude analysis of the PubMed citations found that EQ-5D was described as measuring: HRQoL (85%), QoL (81%), health status (29%), satisfaction (10%), patient-reported outcome (9%), and/or well-being (6%). (340) Whereas the early years of EQ-5D saw it being described as health status, QoL and HRQoL, the majority of literature in 2019 alone predominantly described the EQ-5D as measuring QoL or HRQoL. (340) The ambiguity in what the EQ-5D and subsequently EQ-5D-Y measure is further evident even within the EuroQol Group itself. For example, the EuroQol website, indicates in different places that it measures HRQoL and health status. (353) In the first paper sanctioned by the EuroQol Group, it is indicated that the group met to develop

a ‘standardised, non-disease-specific instrument for describing and valuing health states. However, the same paper indicates that the EuroQol Group aimed to develop ‘a new instrument for describing and valuing health-related quality of life’. (340)

What needs to be investigated

In this thesis, the concern is whether and to what extent the EQ-5D-Y captures health in the same way at it is conceptualized and thought about by children and adolescents in Malawi. The EQ-5D was developed for use with adults and its content was drawn from existing adult instruments. However, models of health for children are different to those developed for adults. (126, 130) Children have their own interests, needs and aspirations as well as the environment which shapes their conceptualization of health. (246)

Secondly, the EQ-5D was initiated in northern Europe (western culture) and it follows that the dimensions are likely biased towards this cultural setting. There may be important culturally specific aspects of health that are not currently represented in the items of the EQ-5D. It is well known that there are cultural differences in QoL perception and how health is expressed generally (188, 253, 339, 354) and there are particular differences between LMICs and HICs. (122) One key finding from a previous study (189) focused on HRQoL measures among young children (<8 years) in resource-constrained settings, found that understanding of health or concepts is different from western countries. For example, ‘social functioning’ was defined differently by parents in Malawi in comparison to the definition found in Western instruments. (355) The recommendation was for an understanding of the broader social-ecology aspect including its effects on child health. (189) Inclusion of environmental as well as faith and spirituality items or domains were also highlighted as important. (189) The systematic review, in chapter 2, similarly found that there has been no generic preference-based HRQoL measures

developed specifically among children, and their use in sSA setting has been limited. (199)

Further, it is possible that conceptualization of health by certain age groups may be determined by the underlying concept of health. For example, health may be understood to have the intrinsic value- such that it is important for its own sake, or that it has some instrumental value- a capability that enables one to do other things, or even more that it is driven by concept of Ubuntu- belonging to a group. The definition of health as suggested by the WHO framework broadly borders around the instrumental value of health, i.e., functioning and capabilities. While functioning looks at the individual's ability to take care of themselves -eating, shelter etc., capabilities is about freedom that enables someone to perform these different tasks including seeking healthcare or attending school. This framework has shaped most HRQoL instruments. On the other hand, children have got more sense of belonging to a group as opposed to individuals who may have already shaped their identity. However, it is a known fact that in most African cultures, even among adults, people have a sense of ubuntu- belonging to a group or community. This would mean that the domains that define health may go beyond an individual level but include a sense of communal belonging.

The qualitative study reported here is to investigate the concepts (and underlying philosophy) of health held by children in Malawi and to compare these with the content of the EQ-5D-Y.

7.3 Aims and objectives

The aim of this study was to gain insight into children's understanding of health in the Malawi setting. In order to achieve this, there were two specific objectives outlined:

i. To explore and construct a conceptual framework of concepts (or themes)

ii. To assess how the conceptual framework compare with dimensions of the existing EQ-5D-Y (and EQ-5D-Y-5L).

7.4 Methods

7.4.1 Participants and recruitment

Eligibility criteria of participants included being a child or adolescent aged 8-17years, and school going within Blantyre city. The procedure for consenting is similar to the one outlined in chapter 4.4.1 of this thesis. Except for the consent/assent recruitment procedure, this part of the research involved a different population from the participants in chapter 4. Information leaflets explaining the study and objectives were provided to prospective participants from primary and secondary schools within Blantyre city. (Appendix 22-24) A convenience sample of healthy children was conveniently sampled from Primary and Secondary schools. Upon consenting, FGDs were held separately for the following age groups: 8-10years, 11-12years and 13-17years. The FGDs were deliberately selected based on these age groups so that the participants could feel comfortable to discuss the topic with their peers. Each FGD composed of between 6-12 participants (214, 225) in order to allow diversity and inclusion of views from other participants.

7.4.2 Procedure

Four focus groups were conducted between 9th December 2019 and 25th February 2020. Prior to this, two pilot focus group discussions were conducted on 18th November and 29th November 2019. The pilots were not only aimed at enhancing interviewer interviewing skills but also logic flow. A minimum of six FGDs were planned to be conducted. The minimum number of FGDs was based on what has been established elsewhere that 80% and 90% of all

concepts are discoverable within two to three and three to six focus group discussions respectively. (356)

On pre-set dates FGDs were arranged. The FGDs started with some ice-breakers including some refreshments before proceeding to discuss topic of the day (health) with an aide of a topic guide. (Appendix 25) Before discussions could begin, all participants were assigned numbers to maintain confidentiality. Discussions were conducted in Chichewa language and were recorded after assigning the numbers. The discussions involved asking the participants their understanding of the construct of health using a topic guide. During the discussions, participants were also asked to identify what they thought were the principal concepts that made up health and to rank these in terms of what was important first. At the end of the discussion, all participants received a pen and an exercise book.

The recordings were later transcribed and translated from Chichewa into English by an experienced translator/transcriber. In order to ensure consistency of the translation, an independent experienced researcher reviewed 10% of the transcription. Any comments and suggested changes were fed back to translator/transcriber.

Use of a supervisory panel

A supervisory panel made up of two supervisors provided feedback on how the FGDs were conducted. This was a helpful exercise enabling this researcher to reflect and learn from the process. Additionally, other researchers within MLW with qualitative research skills were also asked for comments. First, reflexive notes were made by the interviewer which might have influenced in shaping what was said and how or what to avoid during the next round of discussions. These were shared and discussed with the supervisory panel for their comments.

The panel assisted with proposals and suggestions including how best to steer discussion in the next round of interviews as well as some points that needed further delving. All the reflexive notes were compiled and in addition to the topic guide assisted in the next round of FGDs. This same approach was repeated for all the four rounds of FGDs.

7.4.3 Underlying conceptual model and data analysis

7.4.3.1 Underlying conceptual model for development of the conceptual framework

The development of the framework was not shaped by a conceptual model but it aligned to the ecological model for health promotion. (357) Even though the ecological model is primarily for health promotion it had more in common with respect to how health was perceived by children and adolescents in the Malawian setting. The model assumes that changes in the human behavior will come from appropriate changes in the social environment. This can be seen from the five layers proposed by the model: (i) the individual self which among other things include one's beliefs, values, and education level; (ii) the interpersonal level, which is about relationships and interaction with other individuals including family members, friends and social networks; (iii) the institutional level, such as school and church in this research; (iv) the community level, which is about cultural or social norms; and (iv) the public policy level, which refers to policies and regulations at a national level.

It is evident from what the children said that health was perceived in terms of the individual self, the family, school as well as the community -social habits (norms). The framework developed here therefore touched on four of the five levels of the ecological model. The policy level might appropriately have not shaped perception of health among children and adolescents since they were still school going. This conceptual framework of health was developed from the data which was collected.

7.4.3.2 Data analysis and development of the conceptual framework

Data was analyzed using thematic coding employing the following three steps: i) open coding, ii) axial coding; and iii) selective coding to identify concepts and sub-concepts.

The first step was a line-by-line open coding of the transcripts from the first two FGDs (FGD1 and FGD2). The coding was done in Microsoft word document (Microsoft Word for Mac Version 16.55) using the comment section. For quality assurance, about 10% (two pages) of the coding went to an independent researcher from MLW Behavioral Group for review. These were discussed and through consensus some of the codes were revised. The analysis was not drawn arbitrary but was based on findings from the study. Key issues that came out were noted and these were linked to similar ones to form codes. Following this step, the full codes were reviewed by two supervisors on the supervisory panel. All discrepancies were discussed and resolved at this stage in readiness for the next step. The next step was to combine overarching open codes from these two transcripts into categories (axial coding). The last step in coding was the grouping together of categories to come up with concepts (selective coding). The concepts were generated by looking at certain words that supported that category. For example, every quotation that had 'healthy', 'good life', 'healthy life', 'good health' was placed into a concept called 'healthy'. Additionally, during each FGD, the participants as a group were also asked to define health in terms of what are the main concepts of health. These were discussed in detail during each FGD including what composed each of those concepts. A list of identified main emerging concepts were drawn, either on paper or classroom blackboard, and then individually ranked these in terms of what they felt was important first. Part of the analysis also involved comparing and merging categories from open coding with the ranked concepts.

The ranked concepts were considered as the principal concepts and formed the building blocks of the conceptual framework with health at the center. Any categories that were open coded but were not part of the ranked concepts were considered sub-concepts (one level lower than concepts). Further, items that did not make a sub-category were assigned one lower level than sub-concepts. Each concept, sub-concept and item was color coded as follows: concepts (orange), sub-concepts (green) and items (no color) in the framework. Any new emerging concepts, sub-concepts and items were added to the conceptual framework based on the category. Depending on the corresponding color coding that these belonged to on the conceptual framework, it was revised accordingly.

The first draft version of the conceptual framework was developed from the first two FGDs. This was then discussed with the supervisory panel and subsequently revised. Concepts, sub-concepts and items that emerged from FGD3 were added to the first draft of the conceptual framework to come up with the second draft version. Emerging concepts from FGD4 were added to the second draft of the conceptual framework to come up with a third version. The third version of the conceptual framework was further re-examined to determine if any concepts needed to be combined or linked, and also determine the concept levels. This was done by looking at the conceptual framework and going back to the codes. The process was done iteratively several times to come up with the fourth and final version of the conceptual framework.

7.4.3.3 Content comparison between content of EQ-5D-Y (and EQ-5D-Y-5L) and the new conceptual framework for Malawi

The conceptual framework concepts were compared with the dimensions of the EQ-5D-Y to determine the content validity of the latter. The comparison was in terms of the content i.e., the conceptual framework, the number of dimensions and the number of items included. A second

comparison was in terms of similarities and differences between the EQ-5D-Y and Malawi conceptual framework.

7.5 Results

7.5.1 Participant characteristics

The final sample had a total of 21 females and 16 males (mean age 10.4years). Except for one group which comprised girls only (n=7) as this sample was from a girl only school, the other three groups were made up on average of five each of both males and females. Each group had 7-11 children, with a mean of nine children per group. The mean age for females was 10 years and that for males was 15 years. Nine females against twelve males were from the high school, the rest being from primary school. Only four FGDs were conducted out of a planned minimum of six due to schools' closure (20th March 2020) by the government as a precautionary measure to prevent outbreak of the first COVID-19 wave in Malawi.

7.5.2 Developing the health conceptual framework for children in Malawi

7.5.2.1 Understanding and use of the word's 'life', 'health', and 'healthy'

The challenge to translate and disentangle between the term's "*health*", "*life*" and "*healthy*" was noted during the adaptation process reported in chapter 4 of this thesis. Throughout the analysis of the data and development of the conceptual framework, this was similarly found to be problematic. There are places in the text therefore where the differences between these terms remain ambiguous. The Chichewa words for these three terms can be used interchangeably even though the concepts mean slightly different things. The correct Chichewa word for '*health'* is '*umoyo'* which means general health outlook but has synonymously been used with i) '*moyo'* (life) – existence, and ii) '*thanzi'* (being healthy) - a nutritious or strength potential

capacity. As an example, the official English designation for Malawi "Ministry of Health" is rendered "Unduna wa za umoyo" in Chichewa. Clearly, the word '*health'* is translated '*umoyo'* in this official title. However, due to the ambiguity mentioned above it is not unusual to see '*health'* being translated as '*thanzi'* elsewhere. This is the case with an initiative between the Malawi Ministry of Health and University of York called 'thanzi la onse' which is translated 'health for all'. (75) Further, in everyday usage, '*life*' can be used to mean '*health'*. For example, the Chichewa greeting 'moyo uli bwanji' which when literally translated mean 'how is your life' goes beyond just mere existence.

Participants in one focus group discussion were specifically asked to draw a difference, if any, between "*life*", "*health*" and "*healthy*". The following quotes illustrate the complexity of distinguishing these three ideas:

"These two things at some point they are related, but at some point these things seem not to be related… Okay like how number 8 said that it's how you live your everyday life something like when a person has asked you how you are doing and you say I am okay that means you answer depending on how you are doing so there is a relationship between health and Health and when you say you are okay that means everything that has anything to do with your health is good too and like how it was said way back someone who is not healthy has no health" (Male 16yrs, FGD3 line243)

One respondent mentioned that there was a clear difference between '*life'* and '*health'*. The reasoning behind this was that one can be alive but not healthy which would imply that much more is required to be healthy:

"At times one can, be in coma but alive not being in good condition so health and health are two different things" (Female 12yrs, FGD3 line255)

"Because you are on Coma you can breathe and you are still alive but maybe you are sick but you are still alive" (Female 12yrs, FGD3 line263)

This view was not an isolated incident but was shared by most other participants who similarly expressed that there is a relationship between these concepts. The apparent difference is in that being healthy is part of having health but goes beyond just being alive.

"I agree that healthy is one part of health because someone who is sick in hospital and is on Coma they can't manage to do any type of work and everything is done for them and the sign of health is minimal it's like is on low battery so for the health to be said the health is good you are supposed to be healthy" (Male 16yrs, FGD3 line152)

"I think these two things are related because I might ask a friend how they are doing and if he says I am okay then that means in my initial thinking when I was asking her/ him I already had it in mind that I wanted him to tell me about his healthy or work or anything so if they are answering about health since we said healthy has to do with what you do on daily basis so it' does connect" (Male 17yrs, FGD3 line218)

Two other participants, however, expressed the idea that these two were different but from what they said it was clear that the apparent difference was the same as those that said there was a relationship. The thought expressed was that being healthy goes beyond just being alive.

"They are different because number 8 also said that a person can be alive but still be alive but when we say the person is fine then that means they are well and they are not sick" (Male 17yrs, FGD3 line195)

"I think there's a difference between them in that health can be divided into two parts, good health and bad health and for example when someone asks you how are you doing and you say you are okay that means your health is good but when it's bad then that means you are sick so there is a difference there" (Male 16yrs, FGD3 line202)

The difference between '*life'* and being '*healthy'* was evident in that in three of the four focus group discussions, these two appeared as two separate concepts. Further, when the participants were asked to rank these, life was consistently ranked ahead of healthy by most participants as illustrated by the following quotes:

"The first thing is breathing like my friend has said that if out are not breathing then you can't be alive seconded by healthy because that's what is required since if you are health then you are breathing" (Male 17yrs, FGD3 line653)

"The first thing is breathing for one to be alive then you are supposed to be able to breathe," (Female 14yrs, FGD3 line674)

"We must start with breathing because one can't say that you have stopped breathing and you are walking" (Female 13yrs, FGD3, line704)

"On number 1 we will have breathing, because breathing is what brings health and health is the root of everything it's like when the person is not able to breath they are put on oxygen and people say he/ she is in a critical state" (Female 12yrs, FGD3 line724)

"First of all, we must put breathing because one can't say that you are not breathing but you are walking" (Male 17yrs, FGD3 line740)

Ultimately, although the subtle distinction of 'life', 'health' and 'healthy' remained intertwined and did not affect the development of the conceptual framework, it provided some insight. The consensus was that life is central to having health or being healthy. However, the understanding was that health is the central piece with life and healthy protruding from it. Health therefore formed the nucleus in the development of the conceptual framework.

7.5.2.2 Development of conceptual framework for children in Malawi

The conceptual framework for health was developed in stages as discussed in methods section for this chapter. Analysis of FGD1 and FGD2 initially revealed eight major concepts: 1) 'everyday life'; 2) 'everyday health'; 3) 'healthy/good health', 4) 'inanimate'; 5) 'God-given'; 6) 'requires care to prevent illness/maintain health'; 7) 'looking after oneself/care', and 8) 'peace of mind'. After discussion with the supervisory team, 'everyday health' was combined with 'healthy/ good health'. This combination was necessary because both these concepts carry the thought of being healthy. In the end, conceptual framework draft version one had seven concepts that emerged from the first two FGDs. Each major concept in turn had several sub-concepts and items that contributed towards it. Figure 7.1 shows the first draft version of the conceptual framework of health that was developed from the first two FGDs.

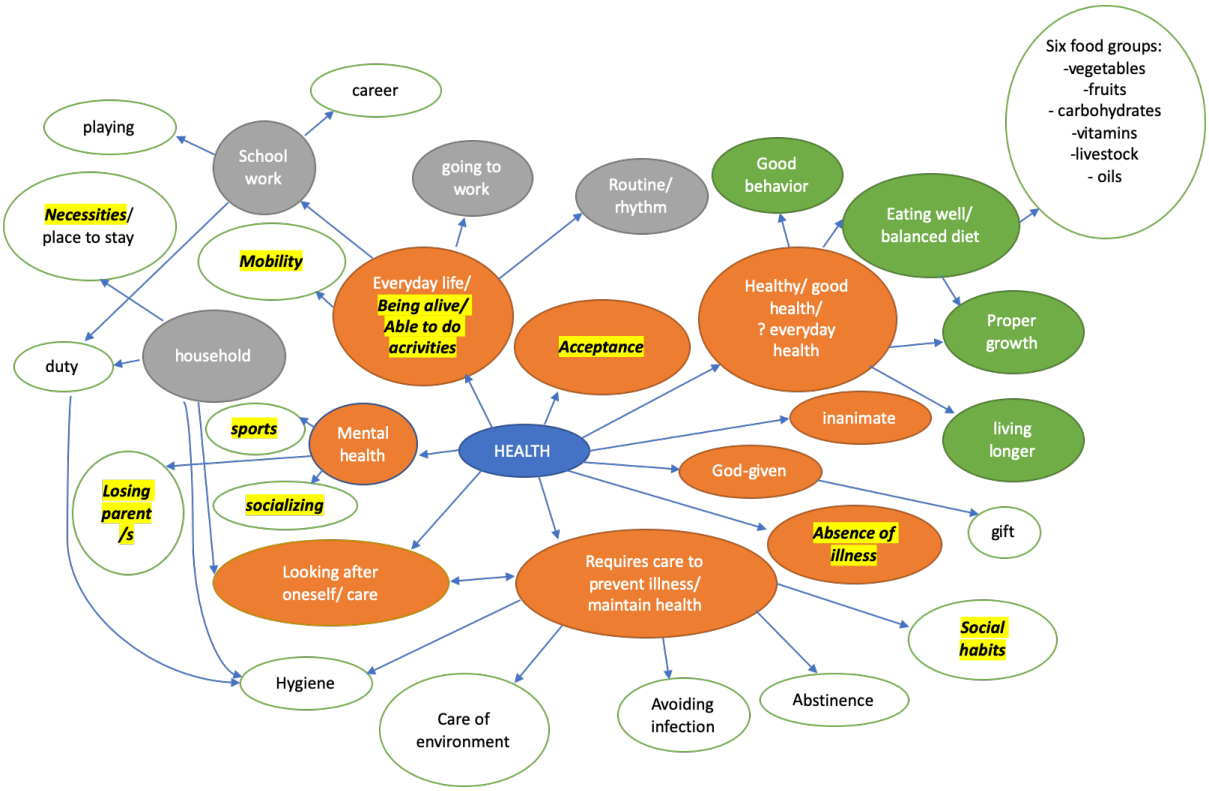
Figure 7.1 Conceptual framework draft version 1 developed for children in Malawi

color code: yellow represent concepts/sub-concepts/items that emerged from FGD 3 and 4; the rest emerged from FGD 1 and 2

FGD3 coding also resulted in seven concepts: 1) 'being alive'; 2) 'healthy'; 3) 'activities'; 4) 'necessities'; 5) 'social habits'; 6) 'looking after oneself'; and 7) 'living well'. These concepts were either the same or similar to those that had emerged from the first two FGDs. For example, 'being alive', 'healthy' and 'looking after oneself' were exactly as the ones that emerged from the first two FGDs. 'Necessities' was embedded within 'place to stay' item as some examples given for the latter were concerning the former. 'Social habits' was also placed under the 'prevention and maintain health' concept as social habits examples pertained to prevention. Similarly, thoughts that were expressed within the 'living well' were falling under 'healthy' concept, and so the former was embedded within this. At the end of FGD3, there were still seven concepts with additional sub-themes and items and so this became conceptual framework draft version two.

Finally, three new concepts emerged from coding FGD4: 1) 'balanced diet'; 2) 'acceptance of illness' and 3) 'absence of illness', which were added to the first conceptual framework. However, the 'balanced diet' was merged with eating well under 'healthy' concept since balanced diet was about eating from six food groups which was a sub-concept that had already been identified. Ultimately then, two new concepts: 'absence of illness', and 'acceptance of illness' emerged from FGD4. These two new concepts were added to the seven concepts in the conceptual framework draft two. In the end, the conceptual framework draft version three had nine concepts. Figure 7.2 shows the second draft version of the conceptual framework of health that was developed from draft 1 version and the last two FGDs (FGD3 and FDG4).

Figure 7.2 Conceptual framework draft version 2 developed for children in Malawi

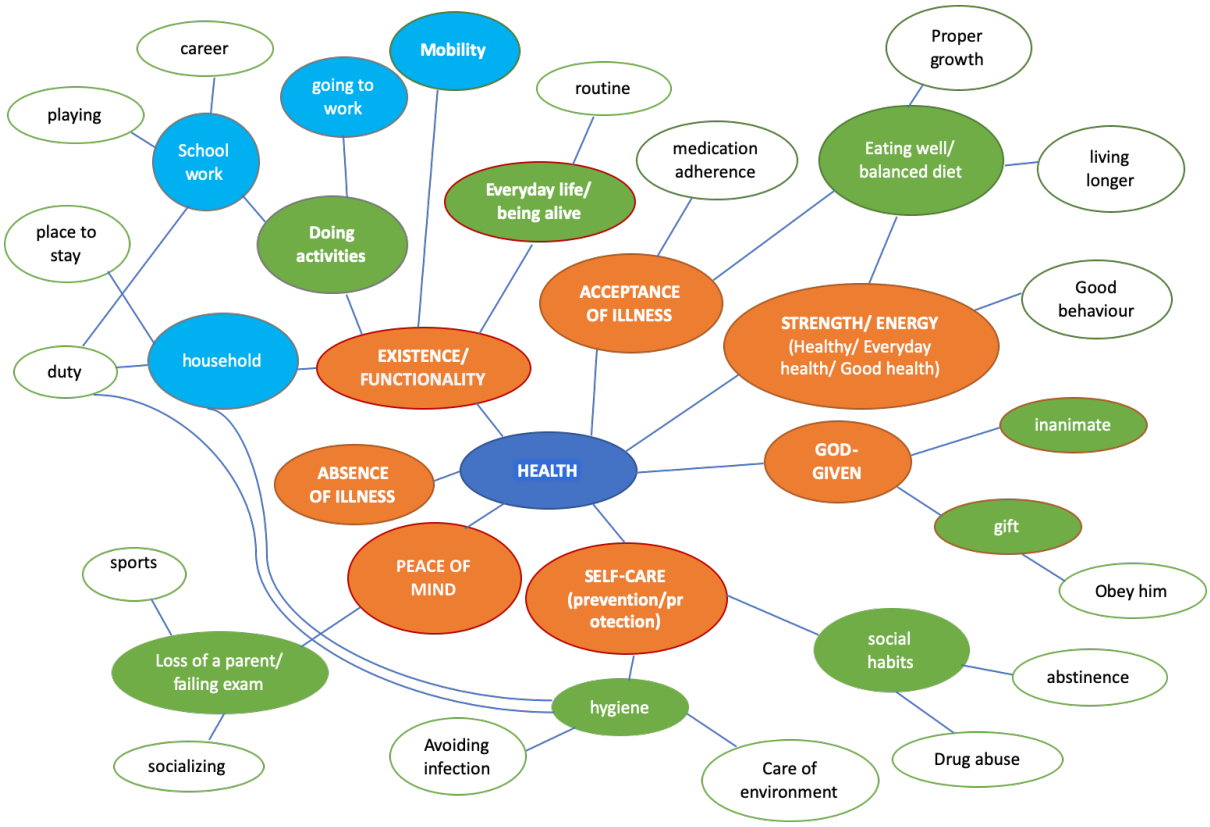


color code: yellow represent concepts/sub-concepts/items that emerged from FGD 3 and 4; the rest emerged from FGD 1 and 2

Further review and discussion of the emerging conceptual framework enabled fine tuning of the framework. For example, the 'looking after oneself/care' and 'requires care to prevent illness/maintain health' were merged into one concept. Both these, it was reasoned, pointed at taking care of oneself whether this was due to prevention or protection. So, 'protection' and 'prevention' actions were aimed at maintaining or taking care of one's health. In the same vein, 'God-given' and 'inanimate' concepts were merged into one concept. Initially it had been indicated that life was inanimate but that was the same that could be said about God. It was therefore reasoned that since life is inanimate it consequently had to do with an understanding about God. As God cannot be seen and therefore inanimate, the 'inanimate' concept was a branch of the fact that it was life that was God-given. Lastly, 'hygiene', 'care of environment', and 'infection control' for example were re-adjusted to the same level. This was at same level as items under the 'looking after oneself/care'. 'Care of environment' and 'avoiding infection' were also combined to be under hygiene. Similarly, 'abstinence' and 'drug abuse' were combined to fall under the social habits. At the end of this refining process, the concepts were reduced from nine to seven. This was now the pre-final version of the conceptual framework.

A final step in the development of the conceptual framework was refining of descriptor of the final seven concepts. The concept of 'acceptance' was refined to what had been said, 'acceptance of illness' The concept 'mental well-being' was changed to 'peace of mind'. This was done not to distort the thought that the participants had expressed. Additionally, the arrows linking health to the concepts to sub-concepts and items were removed as these implied a causal relationship which was not the case. After all necessary wording revisions and edits, the final health conceptual framework version developed for children in Malawi which is in Figure 7.3.

Figure 7.3 Conceptual framework developed for children in Malawi



Colour code: orange: main concepts; green: sub-concepts; no colour: items

The seven concepts in the final conceptual framework are described in detail below:

7.5.2.2.1 Existence & Functionality concept

This was cited right at the first focus group discussion and in all the other three focus group discussions as well. It was expressed as one of, if not, the main component of health. When asked to formally rank the emerging concepts, participants consistently ranked this concept either first or second ahead of all the other identified concepts. The concept of existence (moyo) was described as the life that participants felt preceded everything else – that is, that one can't start talking about health without being alive. This assertion is well illustrated by the following remarks:

"The first thing is breathing because if you are not breathing then that means you are dead." (Male 17yrs, FGD3 line630)

"Like living happily, or you are living a difficult health but as long as you are alive all that's health." (Male 16yrs, FGD3 line22)

The discussion around this concept, however, went beyond just existence. Participants in all discussions also mentioned functionality in terms of one's ability to do various activities including going to work, school, or carrying out some household activities.

"I think when a person says life, they mean that the person is living health good health, they are able to breathe and that they can do different things." (Male 13yrs, FGD3 line85)

"I just want to add on healthy and health because when we say the person has good health, they are supposed to be healthy and be able to perform duties well so that everything should move well in their everyday life." (Male 17yrs, FGD3 line143)

Things mentioned in going to school included listening to teachers, expectation to pass exams, playing, sweeping and mopping the classrooms as well as surroundings. The children also indicated some of the careers that schooling aspire them to become including being a teacher, nurse, lawyer or pilot. Under household, having a place to stay as well as being duty bound to listen to parents, going to the market, doing some gardening, going to the maize mill, cleaning the home (both inside and outside), and cleaning plates were some of the activities that participants mentioned. Interestingly, the idea of cleaning plates went beyond just the female gender as is customarily the case in this setting.

Additionally, participants went further to describe existence as life that is to be expected every day and therefore was something ordinary and normal. This understanding of life was about having life, enabling participants to wake up and go to bed, and see next day without necessarily attributing this to God. There seemed to be an understanding that this normality, or everyday life was physical, referred to as "body life". Children did not spontaneously include a spiritual dimension in this idea of everyday life.

"Body life." (Female 9yrs, FGD2 line 100)

"Having an everyday life" (Female 10yrs, FGD2 line204)

"Life of a person gives them health" (Female 8yrs, FGD2 line229)

"…Your everyday life something like when a person has asked you how you are doing and you say I am okay that means you answer depending on how you are doing…" (Male 16yrs, FGD3 line248)

In only one FGD, participants expressed the thought that existence has also a link to one's mobility. However, this was not a main concept that the participants listed as warranting ranking in that focus group discussion.

7.5.2.2.2 Strength/energy concept

Participants during all the four group discussions explicitly pointed out that strength/energy was a requirement to health. The concept of 'strength/energy' included ideas about 'healthy'

(thanzi), 'good health', and 'everyday life'. Generally, participants associated 'strength/energy' concept with ability to have strength or having an energetic body.

"Strong life" (Female 8yrs, FGD2 line127)

"A healthy life means having a strong body …" (Female 10yrs, FGD2 line462)

Further, 'strength/energy' manifested itself in the ability not only to do things as was the case with functionality domain, but also to do them well.

"When we say the person is healthy that means one is not getting sick often and everything is fine in their body, they are strong and they can do things very well" (Female 15yrs, FGD3 line94)

Additionally, 'strength/energy' concept was linked to one's ability to reach a physical growth potential irrespective of one's health status:

"Healthy like how the others have said it's like your health is progressing well you don't have any illness and also some people can be health while sick but on that we can say that they are healthy because they are able to look after themselves and the illness is not affecting them in any way and they are able to work whatever" (Female 14yrs, FGD3 113)

The idea of good health and thereby ability to reach one's full potential was likened to someone who is in a coma and therefore could only perform things sub optimally as if on 'low battery'.

"I agree that healthy is one part of health because someone who is sick in hospital and is on coma they can't manage to do any type of work and everything is done for them and the sign of health is minimal it's like is on low battery so for the health to be said the health is good you are supposed to be healthy" (Male 16yrs, FGD3 line152)

In all the group discussions, the participants linked strength/energy to nutritional health which was as a result of eating from 'different varieties of food' or eating a 'balanced diet'. The balance diet was said to be made up of six food groups.

"Life of a person means being healthy and eating from the food groups" (Female 10yrs, FGD2 line252)

"A healthy life is whereby one is eating different varieties of food so that their bodies should be health to do anything" (Female 10yrs, FGD2 line473)

"Life of eating from the various food groups gives us energy and healthy" (Female 9yrs, FGD2 line295)

In one of the group discussions, one participant also went further to indicate that physical growth was associated with living "long life". The connection between these two was in the sense that when one is eating from a balanced diet, they would have good body growth and strength, and thereby live longer. For some though, good behavior since it was connected to nourishing or maintenance of one's health, was an aspect that was included in being healthy.

7.5.2.2.3 Acceptance of illness concept

The participants described this concept in terms of the possibility of having health even when living with a condition such as HIV, for example. For some participants this was described in terms of medication enabling health despite the illness. The following quotes illustrates this point:

"If a person has admitted that they are HIV positive and they are taking in their drugs daily, it's possible it also helps them to have a healthy life, because as they ate taking the drugs it's like they are also boosting the immunity which had gone down." (Male 14yrs, FGD4 line 32)

"So, I think it's possible for someone who is HIV positive or any other illness if they are still continuing with their treatment to have a healthy life of excellent health" (Male 14yrs, FGD4 line 37)

The participants also felt that this went beyond just living with a certain illness as it extended to one's health status in general. That meant that if one accepted their health status they could continue with their life.

"The second thing you are supposed to accept how your life is because if you don't regardless of how much you are suffering or how much money you have but if don't accept your life's situations you can't live a better life" (Female 13yrs, FGD4 line546)

Participants also went further to express the negative consequences of not adhering to drug regimen which could lead to new infections and possible death as well.

"One can prevent a lot of illnesses I just want to add on the point that was said by number 2 that if one has HIV and AIDS is supposed to be taking the drugs in order without skipping days because to someone who has HIV and AIDS when they stop taking medication it's like they are inviting other infections as a result the person dies" (Female 15yrs, FGD4 line 250)

And further participants indicated that not accepting health status could result in stress or even in one committing suicide.

"On that point number 4, if a person doesn't accept how their life is that can cause them also to be stressed, and some can even commit suicide or do unnecessary things, be drinking beer and doing reckless things" (Male 15yrs, FGD4 line 638)

7.5.2.2.4 Absence of physical illness concept

This concept came out several times in most of the FGDs and might appear to be a direct contradiction to the concept of "acceptance of illness". In one group discussion, participants were specifically asked to make a distinction between absence and acceptance of illness which they found difficult to do. While some viewed the two as not being different, others were of a different view. The following quotes illustrate some of the reasoning.

"There is no difference because when a person says they are okay that means their health is good and they don't have any other problems like sickness" (Male 16yrs, FGD3 line 175)

"They are different because number 8 also said that a person can be alive but still be alive but when we say the person is fine the that means they are well and they are not sick" (Male 17yrs, FGD3 line 189)

An alternative view was that you can completely detach health from sickness as one can have health only if they are completely free from sickness.

"Their bodies are functioning well they are not sick they are okay they are not suffering from any illness" (Male 17yrs, FGD3 line 56)

Some were of a different view in that absence of physical illness on its own was not sufficient to describe health as this required functioning in other ways as well. This is reminiscent of the WHO (1947) definition of health as not merely the absence of illness. (77)

"Healthy like how the others have said it's like your health is progressing well you don't have any illness and also some people can be health while sick but on that we can say that they are healthy because they are able to look after themselves and the illness is not affecting them in any way and they are able to work whatever" (Female 14yrs, FGD3 line 113)

"When we say healthy, we mean the absence of illness in one's body, meaning that the person is able to do different types of duties/ chores and also to add on that healthy is when a person has good mental health since one can have good health physically but be disturbed mentally" (Male 17yrs, FGD3 line 100)

Still others expressed a quite sophisticated idea about this concept expressing the view that health was a spectrum and could be good or bad or anything in between.

"I think there's a difference between them in that health can be divided into two parts, good

health and bad health and for example when someone asks you how are you doing and you say you are okay that means your health is good but when it's bad then that means you are sick so there is a difference there". (Male 16yrs, FGD3 line 202)

"Staying well I think there is both internal and external. When we talk of staying well internally it has to do with the essence that your health and body is all good and you are not sick from anything". (Male 17yrs, FGD3 line 525)

7.5.2.2.5 Peace of mind concept

Mental health expressed as "peace of mind" was another important aspect of health. This was mainly discussed in one of the FGDs. The participants expressed that peace of mind or stress-free life was a positive state that contributed to health so much that even when one had enough to eat, if peace of mind was missing one could not be healthy.

"On number 1 we should have what one can do to have a healthy life because let's say one is lacking peace in their everyday life so even if you might be eating from all the food groups that can't have any positive impact on your life because you are lacking peace of mind" (Female 13yrs, FGD4 line528)

The importance of peace of mind was also reflected in terms of the negative impact that lack of this has on someone's health. The lack of peace of mind was consistently described in terms of disturbance to the mind (that was affected by the individual's behavior), as lack of peace of mind.

"I can agree with number 11 what she is saying that for example the person is a serial killer but himself too when he goes home, he lacks peace and that means he doesn't live a normal health so at times the mental state of that person is also what makes them to lack peace of mind if that person is doing wrong things they do get disturbed mentally." (Male 17yrs, FGD3 line 463)

While specific suggestions were mentioned in order to achieve peace of mind, one recurring idea was that of remaining 'happy by smiling'. Peace of mind was said could be attained through socialization or involvement in sports. The thought below was one of many the participants made to illustrate this point.

"You find yourself stresses and you just having negative random thoughts so if you associate with people, it helps you manage this well and also being happy not sad it also helps." (Female 14yrs, FGD4 line 621)

Regarding sources of stress or lack of peace of mind, there were various aspects which were said to trigger this including loss of parent/s in death, failing exams or due to some home environment.

"Stress can come about for example here at school you have failed the examinations, so maybe you had friends, best friends and all of them have passed and you are the only one in the group who has failed all the friends you interact with have passed, and you were not expecting to be on that particular position, so you can be stressed too wondering how that has happened" (Female 14yrs, FGD4 line 646)

"But you were with your very same friends and have themselves passed and you have failed, and also maybe you left your home and on your way back your parents have shouted at you or you have done something wrong so that also brings in stress and even your attention at school is affected, you don't pay attention to what the teacher say all that happens due to stress" (Female 14yrs, FGD4 line 654)

7.5.2.2.6 God-given concept

This was mentioned in several group discussions, though often it was not discussed in much detail. This concept was expressed as:

"The life that God gave us that we should have every day." (Female 11yrs, FGD1 line 1125)

In prompting the participants in one group discussion composed of young children aged 8-10years, a reference was made as to whether they would imagine health to be something tangible like an animal for example. The following quote was used to solicit more thoughts from participants when ideas were not coming forth.

"Okay number 3 can you tell me; you have nothing to say? You don't know… you have never heard of an animal called life? You have never heard of that animal it moves in the bush with the name of life?" (Interviewer)

To this end, one of the participants had the following to say:

"When you said life is an animal, can life be touched?" (Male 10yrs, FGD1 line261)

The participants questioned as to whether life or health can be touched meaning whether life was something tangible and could be touched. This point well illustrated the idea that life was thought to be abstract- inanimate. Additionally, there seemed to be this guarantee that since life was God-given, it was expected daily. This was in a way similar to the routine aspect which was highlighted in the 'Existence/Functionality' concept, however, in the 'God-given' concept health definition went beyond routine existence but was perceived as a gift from God.

"She is saying they wake up alive/ with life" (Female 12yrs, FGD1 line 925)

Further, since life was said to be God-given, another idea that was expressed was that it is sacred and so none could take it away.

"Don't' commit suicide" (Female 12yrs, FGD line 938)

Participants also expressed the thought that since life was God-given, the recipients were duty bound to listen to God by obeying what he says and that this would contribute to health.

"We should listen to what God says" (Male 12yrs, FGD1 line 951)

"He gave us life" (Female 11yrs, FGD1 line 967)

"We should study His words" (Male 10yrs, FGD1 line 970)

7.5.2.2.7 Self-care concept

In all the focus group discussions self-care was considered an important aspect of health which was discussed at length. There were two main items to this, protection and prevention. Sometimes both these were expressed with just one action:

"Some you are preventing an infection, while with the other you are just protecting yourself" (Male 11yrs, FGD1 line1091)

When the participants were asked to draw a line between protection and prevention, the following were the sentiments that were expressed.

"It's the same because what you can say about prevention of an illness and protecting oneself you could mean the same thing" (Female 12yrs, FGD1 1077)

"It's the same, its different because at times it can happen that while moving and unfortunately you step on something or cut yourself with something, prevention of an illness maybe you can share something with someone like razor blades and use it without cutting yourself" (Male 10yrs, FGD1 1085)

"…protecting oneself, you can have your own toothbrush and not sharing it with anyone, so with that you can protect yourself from getting infected" (Male 12yrs, FGD1 line1097)

From the above, however, it can be said that protection was described by participants as being more of an individual's effort or practices done in order to maintain good health. Such measures were mainly hygienic practices like bathing, washing hands before eating foods, covering

foods, and cleaning. The importance that participants put on self-care to maintain health can be seen from the following quotes:

"I think health is how you live your health and how you take care of yourself, whether you are hygienic or not" (Female 12yrs, FGD3 line61)

"That one also said that practicing hygiene so I just want to add on that it's like looking after your body so as to be healthy and prevent other illnesses, but also eating hygiene food not just eating anywhere or buying food from anywhere like fried Irish potatoes or mandazi (cakes) just buying from uncovered containers, but you should also make sure that before you eat you have washed your hands, whether you are coming from the toilet you have washed your hands before touching your food , that is also one way of being healthy" (FDG4, line217)

Self-care was also constantly expressed in terms of preventive measures like infection control and other hygiene practices that went beyond just one individual. Things like touching blood, or sharing razor blade, and taking care of how nappies are disposed of were mentioned. Additionally, they mentioned cleaning of surroundings as an important item.

"Okay for example with malaria one can prevent malaria by following the measures that are encouraged at the hospital, like using mosquito nets, taking care of the home like slashing bushes that has grown tall and draining stagnant water so as to have a better surrounding" (Female 14yrs, FGD line 187)

"Slashing bushes and draining stagnant water helps eliminate the mosquito that bleed in those areas, which when bite us can cause malaria, so here it means the slashing of bushes and

draining stagnant water is the same as eliminating mosquitoes at home" (Female 15yrs, FGD4 line199)

Additionally, social habits for example, abstinence from sex or drug abuse which can harm one's health were also included in the category of both protection and prevention. The following quote illustrates why such things could be avoided

"Activities are things that even if you don't do it's not like you will not have health no like if you don't smoke it doesn't mean you are going to break down into pieces no" (Female 12yrs, FGD3 line315)

7.5.3 Comparison of EQ-5D-Y and the new conceptual framework developed for children in Malawi

Firstly, the comparison between the EQ-5D-Y with the Malawi conceptual framework show mixed outcome. (Figure 7.4) The EQ-5D-Y 'looking after myself' and 'worry, sad or unhappy' dimensions can be linked to 'self-care' and 'peace of mind' concepts respectively in the new conceptual framework developed here. The activities listed under EQ-5D-Y 'looking after oneself' dimension like washing and dressing oneself, were the same activities that participants in this study included within the 'Self-care' concept. Participants also indicated additional activities like "washing hands", "cleaning plates" and other hygiene practices as impacting on health. While these examples pertain to hygiene practices as opposed to the EQ-5D-Y which is about functioning, it can be argued that it's practically impossible to do the former without the latter. Also, while "dressing" might not have been explicitly spelt out in the conceptual framework, the participants mentioned "putting on clean clothes" which might be similar.

The ‘peace of mind’ concept in the conceptual framework can be linked to the EQ-5D-Y ‘Worried, sad or unhappiness’ dimension. This was expressed in that ‘*even if one had enough food to eat but if they did not have peace of mind, they would not be described as being healthy’*. Additionally, participants did not only identify sources of stress but also provided possible solutions in order to attain peace of mind.

The other two EQ-5D-Y dimensions of ‘usual activities’ (doing activities) and ‘mobility’ can also be linked to the conceptual framework, though perhaps less directly and as sub-concepts and items rather than main concepts. The functionality concept includes all activities listed under ‘usual activities’ dimension like going to work, school or doing some household chores, and playing. However, doing things with family as indicated in the EQ-5D-Y was not explicitly mentioned by participants in this study. Additionally, the ‘existence/functionality’ concept in the conceptual framework included the existence aspect that is not in the EQ-5D-Y ‘usual activities’ dimension.

The EQ-5D-Y ‘pain or discomfort’ dimension did not explicitly emerge as a health concept at any stage in any of the focus group discussions.

Figure 7.4 Comparison of the EQ-5D-Y (in blue font) and conceptual framework developed for children in Malawi

EQ-5D-Y dimension are written in blue font

Conclusively, the conceptual framework has five concepts ('strength/energy', 'God-given', 'absence of illness' and 'acceptance of illness') that are not captured by any of the five EQ-5D-Y dimensions.

.

7.6 Discussion

This study is the first to develop a conceptual framework of child health for Malawi and is the first to compare the content of the EQ-5D-Y with new qualitative data from Malawi about how child health is defined. To the best knowledge of this researcher, this is the first study to evaluate the content validity of the EQ-5D-Y or any EQ-5D for that matter in a sSA population. This research therefore is an important first step in establishing the gaps that exist within the EQ-5D-Y and in understanding health state measurement in the sSA region in general. The new conceptual framework of health for children in Malawi includes seven concepts. The seven concepts support the WHO definition of not only absence of illness but also physical, mental and social well-being. However, the study findings show other concepts of health that are either new or rarely included in other health status measurement, these include: 'acceptance of illness', 'absence of illness' and 'God-given'.

Regarding 'acceptance of illness', this was predominantly discussed in context of HIV. This finding was not entirely surprising considering that ART has improved people's health and HIV perception. Despite this, however, there is still some stigma from HIV and Tuberculosis in this setting. (358) The general perception is sometimes that once someone has been diagnosed with HIV/AIDS they could as well give up in life. (359) The study found that the participants, though young, believe that health or healthy living is possible even with HIV. This is a significant finding in this setting where HIV prevalence is high, and stigma still exists. Future studies could explore this point further.

The concept of 'absence of illness' is an interesting one. At one end of the spectrum this is perhaps not surprising as other findings have shown the same among children and adolescents perceiving health as such. (108) However, of note is the fact that while participants equated

health to absence of illness, they at the same time felt acceptance of illness or particular chronic condition is important which would sound contradictory or possibly imply that there is a spectrum of health- from greater to lesser extent.

The study also found other concepts that are not regularly included in health existing instruments- whether preference-based or not. For example, the spiritual aspect was identified as a fundamental component of health in the early 1990s. (360) Despite this, the concept has rarely been included in health measures with the exception of the WHOQOL. (82) It may be possible that since most of these HRQoL or health measures have originated in HICs which might loosely be labelled 'secular', this could be the reason. Due to this the 'God-given' concept or spiritual aspect may not be given similar importance between the two societies. A distinction, however, needs to be drawn between God, religion and spirituality concepts. While western cultures might be labelled 'secular' as to the belief in God, this does not necessarily equate to spirituality. Spirituality transcends across cultures so much that even in Euro-American cultures as in African, Latin and Asian cultures, this has been found to impact on healing process (i.e., of health). For example, in a few studies, spirituality was found to be an important item to coping with sickness like cancer. (360-362) Similarly, in another study carried out in Malawi regarding coping mechanism for parents of children with intellectual disabilities, spirituality was found to be an important factor. (363) However, the difference is that while in the African, Latin and Asian culture's spirituality has looked to God, the Euro-American has predominantly been in medical staff. (360) The conclusion nonetheless is that spirituality (in whatever form) has been found to be an important factor in describing health and related constructs. (237, 364) Despite this, spirituality, religion or God has rarely been included as a domain in health measurements. The concept of God as a giver of life is fundamental and something that is expected in some of these cultures (270) as this study found.

There is, however, need to strike balance between what came out from the conceptual framework, the effect these have on HRQoL and subsequent need for new instruments. While it is important to consider all concepts that emerged in the conceptual framework, these need to be weighed against their effect on HRQoL measurement. Most of the concepts in the conceptual framework for childhood health in Malawi can also be found in existing HRQoL with the exception of God-given and acceptance of illness. These two concepts may be important to children and adolescents in a Malawian context but may raise questions on how they affect one's HRQoL. For one thing it may be difficult to construct items that would fall under the God-given and their effect on HRQoL. This concept would ideally fall under the spiritual aspect which as pointed above is mostly missing in HRQoL measurements. The spirituality domain includes asking how satisfied one is with their spiritual life. The concept of acceptance of illness, on the other hand, bordered around acceptance of a chronic condition and sticking to a drug regimen. This concept directly affects HRQoL measurement both mental and physical functioning. The domain could, however, be revised/changed to an adherence to a medical condition. Items under this would include asking if one has a chronic condition, and if they stick to drug regimen and subsequently how their HRQoL is affected. The absence of illness is another concept that emerged in the conceptual framework developed for children and adolescents in Malawi. The WHO definition of health does include absence of illness but is rarely included in HRQoL instruments as a standalone domain. This could be due to the fact that the focus is on how certain domains affect HRQoL. It may follow therefore that absence of illness should not necessarily be added as a separate domain in the conceptual framework.

It is important to highlight here that the WHO health definition is an important way of describing health, but it is just a mere guide at that. For example, whereas pain and self-care

had commonalities and these concepts are part of the physical in an applied way, role could be part of physical, mental or social concept depending on the function. In this model, such activities like eating, sleep and rest, activity and exercise, personal grooming could not be linked directly to the physical, mental or social domains. This raises the need to perhaps explicitly include list of activities which the EQ-5D-Y does, and some role functioning in health instruments which other health measures have done (for example CHQ-PF50 and PedsQL ™4.0). In addition, differences were evident in terms of the length of the instruments as well.

The comparison of the conceptual framework model and the EQ-5D-Y, and whether the latter is a comprehensive measure of health in this setting is significant. The EQ-5D-Y claim to measure of *health status*, but its content was found lacking in this setting in comparison to the conceptual framework. There are quite a number of concepts that emerged but are missing on the EQ-5D-Y. The 'having pain or discomfort' concept, did not come up at any point during any of the focus group discussions. Additionally, 'having pain or discomfort' proved somewhat problematic in the adaptation work outlined in chapter 4. This has similarly been noted in other parts of the sSA region. (274) While this could be surprising since having pain has been found to be a biggest decrement in valuation of health state in HICs, (112) this was not given any attention in describing health among this population. It may be proper to argue that due to the lack of functional health systems in most part of developing countries, including Malawi which ranks among the bottom 10 poorest countries in the world, people have become resilient. Adults including children and adolescents have therefore come to accept pain or discomfort as part and parcel of life and therefore it is not given much prominence when describing health. For example, acceptance of illness was found important and therefore can be concluded that it encompasses the pain or discomfort item. It should be understood, however, that in this study

the acceptance of illness was more about acceptance of a chronic condition and not necessarily enduring body pain associated with an illness. Future studies could investigate this more.

Further, compared to the activities listed within EQ-5D-Y dimensions of 'looking after oneself' and 'usual activities', there were a lot more items that came out in the conceptual framework. While this might indicate that children and adolescents irrespective of the setting have similar understanding of what constitutes daily or usual activities, there is an indication that more is involved in a poor setting. For example, in the conceptual framework there are other items that children included in 'activities' such as going to the maize mill, going to the market, sweeping and mopping both at home and at school. These activities could be perceived differently in a western cultural context, for example, but they are considered normal in LMICs settings which is consistent with other findings. For example, in one study that discussed enjoyment among Egyptian women living in or close to poverty, enjoyment was found to be difficult to define. (365) This is because the women's day was filled up with doing chores, looking for food, or just subsisting anyway. While it could be argued that the Egyptian findings had a much older population (women) from that of this study (children and adolescents), there are some clear similarities. The list of household and school chores mentioned in this present study are considered part and parcel of life and thereby constituting 'usual activities.' Extending this point further is what was observed in one study in a South American country when considering concept of abuse. (237) Things that constituted abuse in North America setting were considered as normal in that this target culture, a clear indication of cultural difference in defining concepts. The point being that cultural context must be taken into consideration when defining constructs.

Extending this argument further is seen in the fact that doing household chores was not just limited to those of the female gender but male ones too. Their involvement in household chores was evident even among participants in the youngest FGD that involved children aged 8-10years. The involvement of both genders in household chores is of particular interest in that it is different from what has traditionally been the case. It may be possible that this signals a shift in cultural norms. However, there is a possibility also that the sample in this research consisted of an urban population which may be different from the rural settings. Whatever the case, this may go in a small or large way to illustrate the changes in cultural barriers to the way things have traditionally been handled.

The mismatch between the content of the EQ-5D-Y and the conceptual framework developed here is generally similar to what has been found elsewhere. Several studies have qualitatively assessed the content validity of the EQ-5D among the among the adult general population. (270, 330, 331) Most of these studies have found the EQ-5D to be inadequate and that it lacks certain health dimensions. As mentioned in the introduction of this chapter, this has been found to the case in UK, (330) New Zealand, (331) and Singapore. (270)

This disparity between the contents of the Malawi conceptual framework for children and that of the EQ-5D-Y as a health status measure raises some serious reflections. Admittedly health as a construct is a complicated concept to define, but the fundamental point remain that health measurements should encompass all the 'possible' relevant concepts. This is especially the case when a health measure is frequently used in a health technology assessment and thereby impacting directly on policy which can be said of the EQ-5D (and by extension the EQ-5D-Y). However, it should be pointed out that the EQ-5D-Y was streamlined to only contain 'core' health components. (349, 350) This was purposely done to make it simple for participants to

complete as well as making it easy for utility valuation purposes. The EQ-5D-Y was therefore by design intended to be used alongside other measures and not as a standalone instrument to measure health status. (366) Over the years it has widely been used to measure health status. The findings from this study show that using the EQ-5D-Y alone as a generic health status measure in this setting has some risk. Other studies have similarly found mixed findings regarding the content validity of the adult EQ-5D version. (270, 330, 331) The EQ-5D-Y is a utility instrument that is suited for use in economic evaluations but has some risk when used as a health status measurement. As such, the EQ-5D-Y needs content that is relevant across a wide range of conditions and potentially cultures. It may be argued that the EQ-5D contains the "lowest common denominator" content, and this was by design to make it easy for valuation studies, but the consequence is that may be at a risk in a Malawian context. This trade-off is what health economist must contend with when using the EQ-5D-Y and other EQ-5D instruments.

This study has some limitations that are worth highlighting. While four focus group discussions are sufficient to highlight aspects of health regarded important in this population, there were other points that would have benefited from further exploration. For example, finding out more about 'Acceptance of illness', 'God-given' and 'peace of mind' concepts would have been more informative. How much an impact 'God-given' concept has on health could not be explored further due to COVID-19 pandemic restrictions on data collection. Also, the HIV/AIDS link to health definition would have benefited from further exploration but could not be due to the same restriction problems.

Additionally, the content validity of the EQ-5D-Y in this population should have tested by introducing the instrument to the participants. The direct comparison of the EQ-5D-Y

dimensions vis-a-vis the newly developed conceptual framework in another round of focus group discussions would probably have elicited more views. This approach has been used in some EQ-5D studies with very positive results. (270, 330, 331)

Another limitation may be that the sample in this study was from an urban school going population only. It is possible that urbanization exposes people to material and other things that are not in the rural communities. This might influence their understanding of health which could be different from a child from the rural part of Malawi. It is also possible that children that are school going have health education that enables them to get a perspective on health based on lessons taught in school. Interviewing children not in school would perhaps have identified items specific to that group. Also, the sample did not consist of sick children and therefore unable to establish if the dimensions that emerged would be as important to them.

7.7 Conclusion

A new conceptual framework of health in children and adolescents in Malawi has been developed. The framework identified seven concepts that children and adolescents said were important in defining health. Only two concepts were similar to those of the EQ-5D-Y and another two on lower conceptual levels. The fact that four out of seven concepts in the conceptual framework represented ideas that were not included in the EQ-5D-Y indicate that the latter is not comprehensive health status measurement in this setting. The EQ-5D-Y is a valid utility instrument for use in economic evaluation, but caution is needed whether using it as a utility instrument or to measure heath status. Alternatively, the EQ-5D-Y should be used alongside other HRQoL instruments like the PedsQL™4.0 in order to make it more comprehensive in this setting. Another possible option would be to develop a health measurement from the local setting but the process to do such takes long. While this study

identified health concepts that children and adolescents consider important some gaps exist. Replication in further studies would be helpful.

In conclusion, the limited content of the EQ-5D-Y has implications for economic evaluation in a Malawian setting. In absence of alternatives, it may be suitable for use as a utility instrument, but with limitations. In this setting and in this population, the EQ-5D-Y could be used alongside the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report, to measure HRQoL.

7.8 Summary of Chapter 7

Firstly, the chapter gave an overview of development of early health instruments including conceptual model and item length. The concepts largely built on the WHO definition of health and therefore found in most health instruments but there were some exceptions too.

This qualitative investigation produced a conceptual framework with seven domains that children and adolescents said are important. The development of the framework enabled evaluation of the content validity of the EQ-5D-Y in Malawi. Only two of the EQ-5D-Y dimensions were identified at the same domain level as the conceptual model, though two further EQ-5D dimensions were included as part of sub-concepts. There were other domains that emerged like 'God-given', which borders around religion, but are rarely included in instruments. The chapter highlighted the need for inclusion of children and adolescents in defining concepts relevant to them especially in view of cultural differences. This is particularly important since most instruments are used cross-culturally.

The following concluding chapter will discuss the findings of this research and the importance of cross-cultural validation of existing instruments in light of economic evaluation. Further, the

chapter reflects on the relevance of the EQ-5D-Y in a Malawian context, challenges and opportunities for future research.

8 Chapter 8-Discussion and Conclusion

8.1 Introduction

This doctorate sought to explore existing childhood preference-based HRQoL instruments and their use in sSA settings. This started with a systematic review where ten existing childhood preference-based HRQoL instruments were identified. Only two instruments (EQ-5D-Y and HUI3) had been used in sSA. The EQ-5D-Y and CHU-9D (replaced by PedsQL™ 4.0 self-report) were selected for cross-cultural adaptation into the Chichewa language for Malawi. The cross-cultural adaptation process into Chichewa involved a standardized translation protocol that included forward translation, backward translation, and cognitive interviews. The cross-culturally adapted Chichewa versions of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report were then psychometrically evaluated (reliability and validity) in a Malawian setting. Finally, the doctorate employed a comprehensive approach to delve into the meaning of health among children and adolescents in Malawi and to establish if the existing instruments align themselves to their perceptions.

To achieve the aim of the doctorate, an exploratory mixed methods design was employed: qualitatively, to establish linguistic and conceptual equivalence of the existing EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report into Chichewa; and to get an understanding of health and establish whether the existing instruments align themselves to the perception of health in a Malawian setting. Quantitatively, the thesis aimed to psychometrically evaluate the Chichewa adapted versions of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report. This chapter 8 provides an integrated discussion of the research findings as well as a report on its strengths and limitations, policy implications and avenues for future research.

8.2 Main findings

The research undertaken in the thesis highlights issues surrounding the definition and conceptualisation of health. Health amongst Malawian children and adolescents was conceptualized in terms of seven concepts some of which are not commonly included in existing health instruments, including EQ-5D-Y. The cross-cultural linguistic and conceptual processes resulted in acceptable Chichewa versions of the EQ-5D-Y and EQ-5D-Y-5L equivalent to the English versions. There were some caveats to this linguistic and conceptual process that were mentioned at the end of chapter 4. The psychometric evaluation established acceptable levels of internal consistency reliability and mixed validity evidence. Self-completion of instruments was found to be particularly problematic among children aged <12 years. Two psychometric properties of test-retest reliability and responsiveness were not assessed due to the COVID-19 restrictions, which puts limitations on the use of the validated instruments.

These main findings are organised along four inter-related broad themes, starting with the qualitative study on the conceptualisation of children's perceptions of health. These themes allow for a conceptual synthesis of the objectives and topic areas that are reported in the previous chapters that constitutes this doctorate research. The following are the four selected themes:

I. Understanding of health among children and adolescents in Malawi and establishing if the existing instruments align themselves to these perceptions.
II. Cross-cultural adaptation of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report into Chichewa.

III. Implication of use of the psychometrically evaluated Chichewa versions of the EQ-5D-Y and EQ-5D-Y-5L in a Malawian setting.

IV. Use of existing childhood preference-based HRQoL instruments in sub-Saharan African settings.

THEME I

8.2.1 Conceptualization of health perception among children and adolescents in Malawi

The research established how difficult it is to conceptualize health in this setting, not only from childrens' perspective but also in general. The fundamental conceptual differences were imperative to establish what concepts '*health'*, '*life'* and '*healthy'* include or involve. The findings tend to confirm that the concept of health is a complex one to define, in this context, as was noted during the translation process. There was no single Chichewa word that translators agreed on except through a consensus. The complexity of health was also noted during the cognitive interviews that sought to establish conceptual equivalence of the translated versions. To a large extent, the interplay between '*health'*, '*life'* and '*healthy'* reflects the confusion around the usage of these terms. This lack of agreement in the terminology to use seems not to matter as much if the intended message is put across. After all, on a global level health takes many definitions. (80)

Across the sSA region, the confluence of '*health'* and '*life'* concepts seems to be common to other Bantu languages. (269) The challenge has been observed to be problematic in the Zulu language of South Africa, in the Kinyarwanda language of Rwanda, and in Kenya. (194) In this thesis research conducted in the Chichewa language of Malawi, this was similarly found to be the case. All this goes to highlight that although '*health'* and '*life'* are used

interchangeably in Malawi and across some of the sSA settings, there are fundamental differences between these concepts. These findings are important in health measurement and intervention evaluation perspective since what is measured determines the end. Inaccurate conception of '*health'* could be misleading and end up with misallocation of the much need limited resources.

The conceptual framework developed in chapter 7 provides evidence against 'imposing' of concepts from the existing instruments in a setting other than where they were developed. Children and adolescents in Malawi identified other aspects of '*health'* that are important to them, lying beyond those defined by the EQ-5D-Y. This has implications for the use of health instruments in general but also specifically that of the validated Chichewa EQ-5D-Y version in Malawi (discussed in detail in section 8.3.3 below). In general, the findings of other concepts of '*health'* that are either new or rarely included in health status measurements point to two things. Either the domains or concepts relevance of existing health instruments should be established first before use and new content added where necessary or that the cross-cultural adaptation should be replaced with development of new instruments. If the former, it is noteworthy that the cross-cultural adaptation of an existing HRQoL instrument, the process must be thorough. This is a well-known fact in psychometrics and is discussed in detail in the next section. Suffice to say the findings from the literature review outlined in chapter 2 show that while the use of generic childhood preference-based HRQoL in sSA is rare, the adaptation process is sub-optimal. If the option is to develop new instruments on the other hand, the process is long. This trade-off must be weighed accordingly.

The fact that the conceptual framework has more concepts than existing instrument point to the second option (relativist approach): development of new instruments. This might be the

best approach to ascertain the concepts that matter to the target population. The concepts identified by this thesis framework seem to broader definition of health. However, the downside to developing new instrument is the time and costs involved. A possible option could be to use a combination of other childhood preference-based HRQoL instruments identified in chapter 2 or already cross-culturally validated HRQoL instruments to complement for the missing concepts. In Malawi, no other childhood preference-based HRQoL instrument has been validated other than the EQ-5D-Y and EQ-5D-Y-5L. The alternative is to use HRQoL measures such as the PROMIS-25 and WHOQOL BREF which have been validated among children with different health conditions in Malawi, (195, 367) although these are not preference-based HRQoL instruments. Until such a time when alternative childhood preference-based HRQoL instruments become available, this will be limited to the EQ-5D-Y and EQ-5D-Y-5L in carrying out economic evaluation in this population. The conceptual framework highlights the need for greater involvement of children and adolescents themselves in the development of childhood health measures. (107) As highlighted above and shown in chapter 7, development of new instruments can be long and expensive. The psychometric evaluation in chapters 5 and 6 further goes to show how involving this whole process can be.

However, these steps are necessary if childrens' 'voices' are to be taken into account cross-culturally. While adult domains/dimensions relate to children and adolescents, (159) it is also a known fact that children have different perspectives and goals. (246) This implies that concepts and subsequent domains or dimension structures should ideally be developed directly from children themselves within specific contexts. (142) This is consistent with previous authors (189) that have emphasized the need for involvement of children within the relevant cultural context. (191, 193) Generally, children have been fairly involved in the construction of dimensions of health instruments. In development of childhood preference-based HRQoL

instruments, however, childrens' involvement has been sparse. The CHU-9D is the only preference-based HRQoL measure identified in the literature review to have been generated directly from children. (107) The findings from this thesis research confirm the need for childrens' involvement as more concepts emerged from the conceptual framework of health in comparison to dimensions covered by existing instruments. Developing new instruments not only takes time but requires considerable resources to generate items, and to conduct the pilot testing and psychometric evaluation. The research in this thesis has shown that although development of childhood preference-based HRQoL instruments has lagged worldwide, this has particularly been the case in the sSA region. As such, cross-cultural adaptation of existing instruments will for the time being considered a preferable 'quick' solution in the sSA region.

There is another element of the conceptualisation framework of health that needs to be highlighted. The results from the psychometric validation of the PedsQL™ 4.0 and EQ-5D-Y instruments in chapters 5 and 6 respectively showed that gender had no influence on HRQoL scores. This provides evidence that gender does not influence the understanding of health in a Malawian setting among children and adolescents. The conceptual framework further showed no difference in health perception between boys and girls. For example, there were no concepts that stood out as being distinguishable to girls only or that were more strongly felt by one gender compared to the other. However, there were concepts such as absence of illness and acceptance of illness that emerged from among adolescents only. This may well reflect developmental phase and what may be considered important among this age group. Having emerged from one FGD only there was no room to further test this but might benefit from future research. Suffice to say generic instruments generally seem to capture domains for the whole childhood, i.e., age 8-17years. It may be possible that much as there may be some disparity between children and adolescents, this may not be as significant. In addition, the lack

of differences between health perception by boys and girls shows that there is no need to develop gender specific or sensitive instruments. As is the case the existing childhood HRQoL instruments including those that were adapted and validated in this thesis target both genders and age groups (children and adolescents). It can therefore be concluded that the adapted instruments and any future work at developing instruments will have concepts applicable to both genders and age groups.

THEME II

8.2.2 Cross-cultural adaptation of existing instruments

As indicated above, the lack of locally available childhood preference-based HRQoL instruments will mean two things: a continued adaptation of existing instruments cross-culturally or developing new ones. The challenges of devoting energy, time and resources to developing robust locally relevant instruments has been highlighted above. Hypothetically, there could be a case for a class of expanded generic instruments that are relevant to both HICs and LMICs. This also might be constrained with time as data from all settings needs to be collated and compared on the same scale. Alternatively, perhaps there is a case for new content within existing generic instruments to enhance their usefulness in LMICs, although the trade-off is that it cannot be compared between the two settings. Due to factors such as time and resource constraints, adapting existing instruments may be a quick and cheaper alternative. Additionally, the need to compare health outcomes internationally necessitates the need to cross-culturally adapt the same instruments. This research established that while it is possible to produce capable versions from a source version to a target culture there are associated limitations. The Chichewa language versions of the EQ-5D-Y, EQ-5D-Y-5L, PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report established linguistic and acceptable levels of conceptual equivalence with the original English versions. However, adapting a licensed

instrument such as the EQ-5D-Y is difficult due to political as well as logistical challenges of working with EuoQol Group to help them understand the need for localised content. In doing so there would be a loss of the generic nature as the content would not be the same across all uses of the instrument. So, although content might be shown to be missing, there is some limitation on what can be done to enhance content from the local perspective.

The psychometric literature has well established the importance of not relying on translation or cognitive processes of existing instruments alone. (145) To this end, several guidelines have been developed to guide the cross-cultural adaptation process of existing instruments. In as far as childhood preference-based HRQoL instruments are concerned, this doctorate found this has been a challenge in sub-Saharan Africa. Where this attempt was made, it was done at a sub-optimal level. The problem with applying only a sub-optimal adaptation process is that it could result in risks of good data collected from such instruments. To come up with a thorough cross-cultural adaptation process might be time consuming and has considerable challenges as this research found, but it is necessary. The following highlights the need and challenges of carrying out thorough cross-culturally adaptation processes of existing instruments.

Translation costs

Potentially high translation costs can be a barrier to cross-culturally adapting existing childhood preference-based HRQoL and likely other instruments as well in research. The extent to which an existing instrument is adapted and used locally depends, to a large degree, on whether the adapted version exists. This research could not adapt the CHU-9D into Chichewa because of the translation costs. Even at the request of involvement of this researcher, the cost was still high for young and other researchers. The positive thing is that once the instrument has been adapted into the local language, subsequent users are only asked to pay a relatively small user fee. On the other hand, the EQ-5D-Y, EQ-5D-Y-5L, PedsQL™

4.0 child self-report and PedsQL™ 4.0 were adapted into Chichewa by this researcher. There were no costs associated with the adaptation of these measures as this was part of this research but could have been the case otherwise. The EQ-5D-Y and EQ-5D-Y-5L adaptation process included the VMC reviewer who was financed by the EuroQol Group. This is very encouraging for research. Future users will likely benefit from use of these cross-culturally adapted Chichewa versions plus they will only be required to pay a small user fee.

Translation

Studies or users often settle for keeping the same concepts/items underpinning measures (absolutist approach) without going through the whole adaptation process (forward translation, backward translation, and cognitive interviews) as was the case with the HUI3 in Kenya. This 'shortened' approach (translation only) has the advantage that it is less costly and time saving, but risks compromising research findings. Translation does not necessarily ensure concepts will be understood by the target population and potentially renders the work worthless. This amplifies the need for cognitive interviews to establish conceptual equivalence of a translated version.

Cognitive interviews

Cognitive interviews identify 'hidden' problems with translation and may reveal major conceptual issues that need to be addressed. For example, and as highlighted in chapter 4, the PedsQL™ 4.0 child self-report "It is hard for me to take a ***bath or shower*** by myself" item was translated "Zikumandivuta ***kusamba***". However, the term '*kusamba'* in this context has a cultural connotation to 'menstrual cycle' for adolescent girls. When "m'thupi" (body) was added, to read "Zikumandivuta ***kusamba m'thupi***", the confusion was resolved. This shows how instrumental cognitive interviews were in identifying and resolving this problem. The

cognitive interviews were also fundamental in identifying problems that children had in completing the questionnaires. This helps inform the type of questionnaire to use in different populations (i.e., self-completion, interviewer assisted or proxy). The cognitive interviews identified that children aged 8-10years had problems with self-completion of questionnaires.

Further, the findings from the psychometric validation confirmed this, missing data was common amongst this younger age group. This goes on to confirm that while comprehensibility might generally be quantitatively assessed through missing responses, for example, this can be picked up earlier on during the adaptation stage. This has many comparative advantages including ensuring that a best instrument is used. In addition, it may cut down on time and resources to adapt an instrument that will later have serious issues when administered.

The establishment of a correct understanding of severity in HRQoL levels

The thesis highlights the cognitive interviews alone might not be adequate for some generic childhood preference-based HRQoL instruments. For example, the cognitive interview process was thorough for the EQ-5D-Y-5L. However, severity level inversion existed even after this thorough process that involved self-completion of questionnaire and in-depth interviews. The EQ-5D-Y-5L ranking exercise was instrumental in identifying problems in the translation of the five different severity levels. The PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report versions similarly have five severity levels, but the ranking exercise is not part of the translation protocol. This research could therefore not establish whether the severity level inversion exists within the adapted PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report versions. Arguably, the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report are not utility instruments, but this transcends to all kinds of instruments. Both the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report have a scoring algorithm

and severity level inversion if it exists could affect the health profile scores. It is very likely that had it not been for the card ranking exercise, coming up with accurate qualifiers for the EQ-5D-Y-5L correct hierarchical ranking would not have been adequately established.

The effect of severity inversion on utility valuation exercise has been occasionally reported as a problem. (275) It is highly recommended to use the ranking exercise when adapting the EQ-5D-Y-5L cross-culturally. The findings on the EQ-5D-Y-5L scale would support the need to review the severity inversion in instruments that don't have this feature in their translation protocol. The recommendation would be to consider doing the same in the cross-cultural adaptation of the PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report in other settings.

THEME III

8.2.3 Implication of using the EQ-5D-Y and EQ-5D-Y-5L in Malawi

The work presented in this thesis identified both psychometric and conceptual flaws in both the adapted EQ-5D-Y and EQ-5D-Y-5L. Table 8.1 below show the psychometric properties that were evaluated in this thesis.

Table 8.1 EQ-5D-Y and EQ-5D-Y-5L psychometric properties investigated in this thesis

Main Psychometric property	Extent investigated: Full/Partial/No	What was investigated: Yes/No (chapter)	Reason
1. Conceptual framework	Full	*Yes (Chapter 7)*	
2. Cross-cultural and language adaptation	Full		
Linguistic equivalence		*Yes (Chapter 4)*	
Conceptual equivalence		*Yes (Chapter 4)*	
3. Item analysis	Full		
Missing data		*Yes (Chapter 6)*	
Maximum endorsement frequency		*Yes (Chapter 6)*	
Aggregate adjustment endorsement frequency		*Yes (Chapter 6)*	
Item redundancy		*Yes (Chapter 6)*	
4. Reliability	Partial		
Internal consistency		*Yes (Chapter 6)*	
Test re-test		*No*	*COVID-19 restriction*
Inter-rater		*No*	*Data collected from children only*
5. Validity	Partial		
Convergent		*Yes (Chapter 6)*	
Discriminant		*Yes (Chapter 6)*	
Known-group		*Yes (Chapter 6)*	
Criterion		*No*	*Lack of gold standard and if this was available there would be no need for adapting/developing new instruments*
Empirical		*Yes (Chapter 6)*	
Content		*Yes (Chapter 7)*	
6. Responsiveness	No	*No*	*COVID-19 restriction*
7. Practicality/burden	Partial		
Missing data		*Yes (Chapter 6)*	
comprehension		*Yes (Chapter 4)*	
Completion time and rate		*Partly (Chapter 4)*	

Source for psychometric properties: FDA, MOT, ISOQOL, COSMIN (145-148)

The psychometric properties evaluated in this thesis generally established reliability (internal consistency, but not test-retest) and mixed validity of both the EQ-5D-Y and EQ-5D-Y-5L Chichewa (Malawi) for use in children and adolescents in Malawi. There were no differences between the EQ-5D-Y and EQ-5D-Y-5L in terms of missing data, item redundancy, reliability and known-group validity. The EQ-5D-Y outperformed the EQ-5D-Y-5L in utilization of response options as well as empirical validity, and the latter had better discriminant validity.

Missing responses were relatively high in this study in comparison to other studies consisting of a general population. (160, 368) The particularly high missing values among children (8-12 years) would indicate difficulty in providing good quality self-reported HRQoL assessment amongst the younger population. (284, 316) The findings suggest younger children may benefit more from an interviewer assisted version. (369)

The evidence for convergent validity shows that criteria were met at scale but not at dimension level. This might imply that the EQ-5D-Y-3L and EQ-5D-Y-5L are best suited to assess physical functioning as opposed to other aspects of HRQoL. While the adult EQ-5D-5L has been found to be associated with high correlation with other health measures compared to the EQ-5D-3L, (318-320) this was not the case with the two youth versions. These correlations were low to moderate (due to differences in recall period, dimension composition and the scoring systems) which is consistent with other findings. (181, 204, 208)

Both the EQ-5D-Y-3L and EQ-5D-Y-5L showed evidence of known-group validity which has been observed elsewhere. (160, 279, 370, 371) While the EQ-5D-Y-3L had the largest effect size in children, the EQ-5D-Y-5L was able to discriminate better in adolescents. The adult EQ-5D-5L has shown evidence of ability to distinguish between known-groups based on age. (322)

The findings from this study confirms that the EQ-5D-Y-5L may be best suited for adolescents since they have better distinguishing ability of the responses.

Regarding content validity, four EQ-5D-Y dimensions could be linked to the conceptual framework for health although only two of the EQ-5D-Y dimensions were on the same conceptual level as the framework. These findings would suggest limitations of using the EQ-5D-Y as a HRQoL measurement in this setting. This is a finding that has been observed elsewhere. (270, 330, 331) Additionally, test-retest reliability and responsiveness were not evaluated in this thesis. Given these limitations, and in light of whether to cross-culturally adapt or develop a new instrument, the EQ-5D-Y has potential for use in this setting although it would be useful to verify the other properties. Further, the EQ-5D-Y should be used in this setting with limitations and weighed against the following points.

What the EQ-5D-Y intended to measure

The EQ-5D-Y and EQ-5D-Y-5L, and by extension all EQ-5D instruments, were developed for use in economic evaluation and should be used as such. The findings from the conceptual framework of health in this thesis are similar to what has been found elsewhere regarding the adult EQ-5D's lack of certain health dimensions. (270, 330, 331) The PROMIS (367), WHOQOL-BREF (195) and, through this doctorate research, the PedsQL ™ 4.0 self-report have been validated in Malawi. These are possible options for use alongside the EQ-5D-Y if the aim is to measure health status and HRQoL in children and adolescents. However, the aim of this thesis was to evaluate the EQ-5D-Y (and EQ-5D-Y-5L) as instruments for economic evaluation. The findings show that the EQ-5D-Y can be used with limitations in this setting.

On the other hand, the lack of a pain concept in the conceptual framework might indicate the lack of importance of pain in this setting. This is similar to what has been observed in other LMICs (372, 373) and cultural settings where pain may be taken as part and parcel of life (374) but is in sharp contrast to importance of the same in valuation studies in HICs. (112) The omission of pain in the conceptual framework supports findings in the EQ-5D-Y psychometric validation in this research, where 'having pain or discomfort' is the only dimension that had missing values among children and adolescents. Further, 'having pain or discomfort' failed on the adjacent aggregate endorsement frequency.

Additionally, some items that might be viewed or treated with cynicism in HICs are just normal in resource limited settings. This was evident in the items that were listed under the 'activities' concept in the conceptual framework. Such things as going to the maize mill, going to the market, sweeping, and mopping could be perceived differently in a western cultural context. Both these points underscore the important fact that the makeup of concepts has to be defined and understood within the local cultural context.(237, 365) This finding must be treated with caution, however, as not only did the sample include children who were relatively healthy but is also against recent findings in the sSA region. The EQ-5D-5L utility valuation work in Uganda identified pain as the main driver for the utility scores. (375) Therefore, while pain did not come up in the conceptual framework it may still be a factor that needs to be looked into when assessing HRQoL.

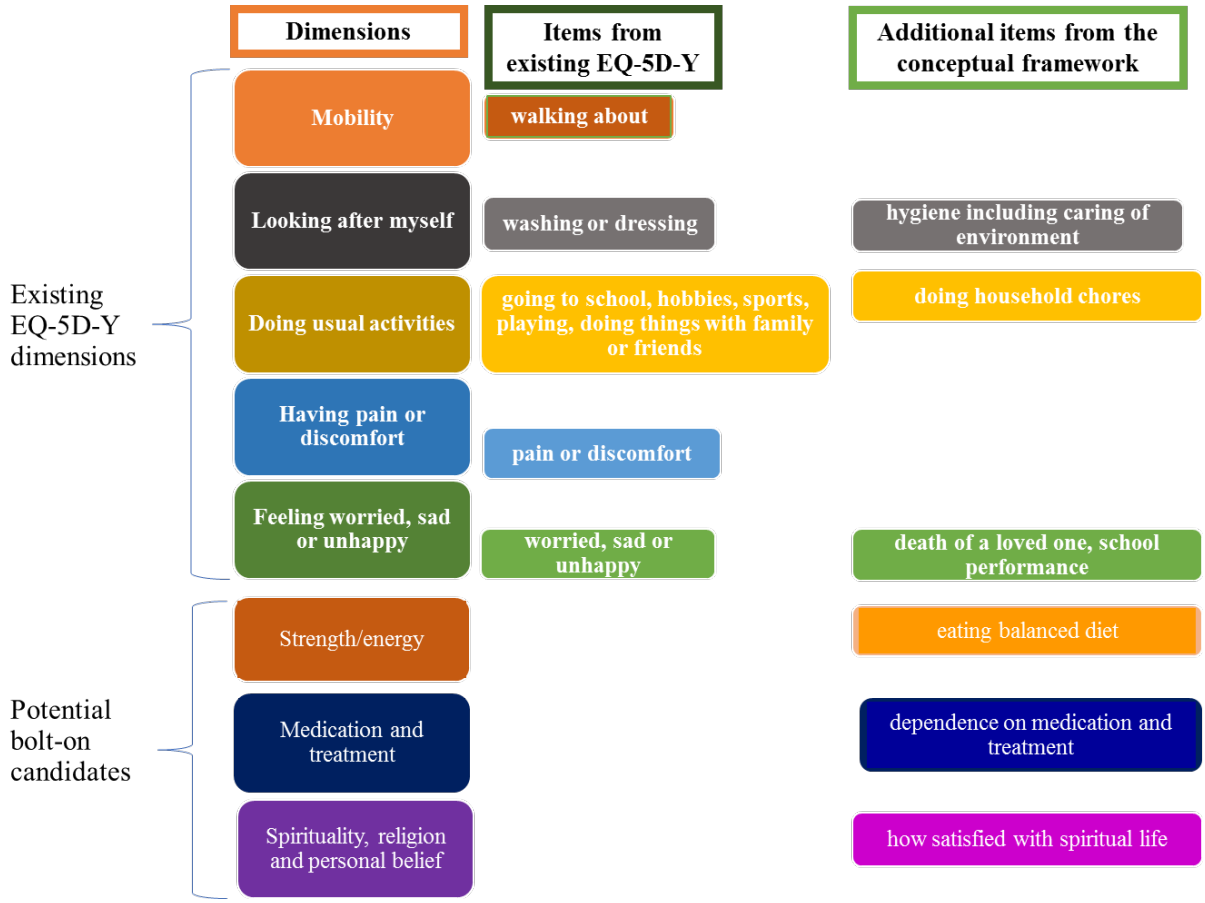
Inclusion of additional dimensions to existing EQ-5D-Y through bolt-ons

In order to improve the content validity, various attempts including adding extra dimensions (known as bolt-on) have been attempted to the existing EQ-5D instrument. (376) Several issues have been highlighted with regards to such attempts. Firstly, the EQ-5D has utility values

derived from existing dimensions and any attempt to add extra dimensions would impact on the utility values. (330, 377) Secondly, the EQ-5D is a generic instrument and therefore the extra dimension/s to be added would depend on the disease area of interest, although these could be contextualized instead of focussing on a disease.

It has been argued that as a generic instrument, the dimensions for bolt-on have to come from the general public after all the utility values are generated from the general population. (270) In UK and New Zealand, the sensory functioning has been identified as a potential candidate for inclusion in the EQ-5D dimensions. (330-332) A recent study in Germany has specifically looked at a bolt-on for the EQ-5D-Y through qualitative work with children and adolescents. (332) Cognitive functioning was singled out as a potential candidate for bolt-on to the EQ-5D-Y. The most recent literature review looking at EQ-5D bolt-ons (376) found that cognitive functioning is the most common dimension that has been suggested for bolt-on in most studies. However, cognitive functioning did not emerge in this thesis's conceptual framework. The potential bolt-on candidates for the EQ-5D-Y in Malawi include strength/energy, acceptance of illness, absence of illness and God-given as shown in Figure 8.1 below.

Figure 8.1 showing how the EQ-5D-Y bolt-on version for Malawi would look like



Source: based on own synthesis

Any adapted EQ-5D-Y bolt-on for Malawi could contain the five dimensions and the potential candidates from the conceptual framework. Currently there are no standardized methods for establishing bolt-on (376) as to whether to include some or all the concepts from the framework. There is work under way looking at standardizing methods for inclusion of bolt-ons, (378) but until then this remains to be seen. Suffice to say sources for potential bolt-on could either flow from a top-down approach, e.g., literature reviews, or from a bottom-up qualitative work with a target population like was case in this thesis. The former approach would support inclusion of the cognitive functioning, which has frequently been flagged up in most work but not in this thesis. On the other hand, the latter approach has shown that cognitive functioning is not an important dimension in LMIC settings, at least from one study that has

evaluated the content validity of the EQ-5D. (270) This mirrors the finding from this doctorate research.

In the LMIC study referred to above, although based on adults, social relationships, medical conditions and health promotion knowledge and behaviour were dimensions that participants considered important. This current doctorate research resonates well with findings from that study. Developers of the EQ-5D-Y themselves suggested that future studies should consider other aspects of HRQoL that are important to children and adolescents for inclusion in the measure. (197) As shown by this thesis research, consideration of potential candidates for bolt-on to the EQ-5D-Y from most studies is a matter of context and need to come from the target population. In the interim, due to lack of standards for bolt-ons, all the identified concepts from the framework could be added to the EQ-5D-Y. The importance of this approach is that it would render the EQ-5D-Y comprehensive to measure health or HRQoL in this setting. This would however make the EQ-5D-Y non generic across settings and therefore not comparable internationally.

Irrespective of whether to only add some or all concepts from the framework, one challenge that remains, other than methodology, is the utility values to be applied. The current adult EQ-5D versions have utility scores which at times are used for evaluating childhood health interventions. This is discussed in detail below, but just to highlight that extending the dimensions beyond the existing ones would pose serious challenges for how existing value sets can be mapped to added dimensions. Additionally, some dimensions might not be relevant to the local setting as was the case with 'pain or discomfort' in this thesis. So, while adding extra dimensions to an existing instrument might solve some problems, it does not eliminate all problems. To this effect, some have advocated for development of a new but extended version

of the EQ-5D to capture health and well-being.

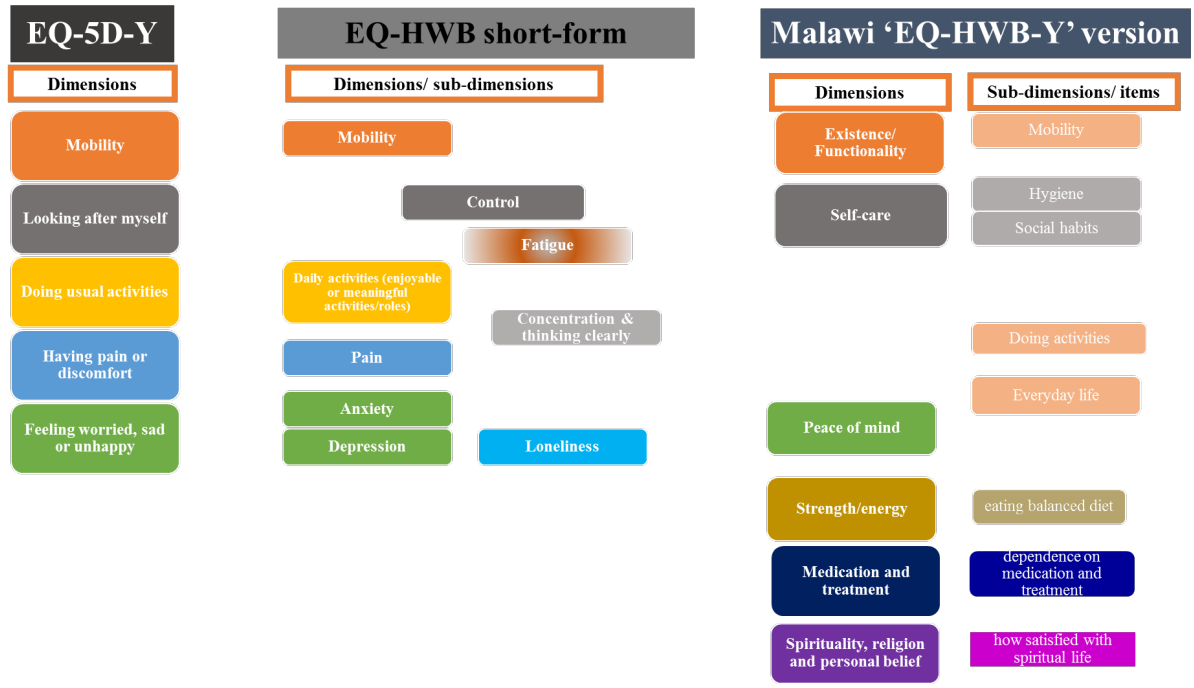
Extension of the EQ-5D-Y using the EQ-Health and Well Being approach

The European EQ-Health and Well Being (EQ-HWB) is a new approach for developing a specific instrument that captures aspects not covered within the EQ-5D-Y. (379) The EQ-HWB project, which took just over three years, was born from the need to have an instrument that captures the impact of interventions on patients, carers and social workers but at the same time being used for economic evaluation. It is important to note that a generic instrument will not have all the dimensions required to measure a construct. The trade-off will always be whether and to what extent it covers all relevant dimensions that measure the construct. Generally, preference-based instruments may have one item to capture a concept or dimension which might be different from other health instruments e.g., a profile index that may have several items to capture a dimension. To overcome this, the EQ-HWB has 25 items over nine dimensions of: mobility, daily activities, anxiety, depression, pain, control, concentration and thinking clearly, loneliness, and fatigue. The combined anxiety or depression in the current adult EQ-5D version has been split into distinct dimensions. Similarly, the combined pain or discomfort has been left as representing the latter only and discomfort as one of the items. Additional dimensions have been included, which were developed from a conceptual framework. Another important feature of the EQ-HWB is that it is being developed to have associated utility values.

The performance of the EQ-HWB is yet to be assessed, however, it is hoped to address dimension limitations of the adult EQ-5D. The EQ-HWB project is therefore an important step in developing a comprehensive generic preference-based instrument that is equivalent to the adult EQ-5D. This kind of approach can be extended to a Malawian context. The question

would perhaps be what the EQ-HWB would look like in Malawi. The conceptual framework developed from this thesis has provided insight into dimensions not captured within the existing EQ-5D-Y. Blending the existing dimensions and new concepts from the model would form an 'extended' childhood version of the 'EQ-HWB-Y' for Malawi. Figure 8.2 below presents a comparison of the EQ-5D-Y, EQ-HWB and the hypothetical Malawi 'EQ-HWB-Y' version.

Figure 8.2 showing comparison of the EQ-5D-Y, EQ-HWB and a hypothetical 'EQ-HWB-Y' for Malawi.



Adapted by the author from Brazier et al: (380)
Colours correspond to equivalent dimensions (or sub-dimension) across the different versions

In total, the hypothetical 'EQ-HWB-Y' for Malawi would have at least six dimensions spanning across 19 items. It must be pointed out that there is a sizable gap in the middle of the EQ-HWB that is still not mirrored in concepts reported to be important in Malawi. This includes such dimensions as pain, fatigue and concentration among others. The EQ-HWB has similarly been developed in HICs and reflects the same issues observed with the EQ-5D-Y.

(379) This evidence may point to the need to avoid adapting but develop instruments that meets context needs. Looking into the future, the work presented in this thesis might be a first step to developing a generic childhood preference-based HRQoL instrument for Malawi. Future studies will likely extend this framework to have a comprehensive childhood health and HRQoL instrument. The second step would be to develop utility values so that the measure can also be used in economic evaluation.

The points raised above have highlighted challenges and opportunities to extend or make the EQ-5D-Y more comprehensive for use in Malawi. While these point to some future adaptations and areas that might need to be addressed, there is need for an interim instrument that can be used to inform resource allocation decisions. Conclusively, the EQ-5D-Y will still be a useful tool in the interim but bearing in mind the above limitations as well as the points discussed below.

Interviewer administered versus self-completion questionnaires

The thesis research findings show that interviewer administered is especially necessary for children aged 8-10 years as has been established elsewhere. This is a known fact from use of self-reported health instruments in HICs but carries extra weight in a Malawian context due to literacy and competency issues, especially among children. As discussed in chapter 1, there is a high proportion of the population with low literacy. An economic evaluation need to include the health information from this section of the population, so that decision making is based on a representation of the whole population. Tendency to tick every box instead of choosing one option unless interviewer assisted implies difficulty in getting accurate HRQoL assessments in children. Additionally, both the EQ-5D-Y and EQ-5D-Y-5L, as well as PedsQL™ 4.0 child self-report and PedsQL™ 4.0 failed on the discriminant validity criterion by school grade

among adolescents. While this could be due to a small number of respondents in the lower school grade among adolescents, the findings were significant for combined age. This finding suggests that self-completion of the questionnaire is dependent on education level. As such the use of the self-completed questionnaire is only as good as the population it is administered to. Both these findings point to the option for use of interviewer administered questionnaire among children to yield good HRQoL data and there is recent evidence to that effect. (369)

Another option would be changing the format of the instrument with addition of pictures to assist with comprehension. Use of pictures has been seen to improve understanding of health knowledge and have been recommended for use in healthcare implementation. (381) The development of the disability questionnaire in Malawi, with pictorial representations for children aged <7 years, has proved successful. (382) While there are advantages, there are some known disadvantages too in using pictures from a psychometric point of view. Generally, it may be hard to know that everyone is interpreting the pictures the same way and therefore that the instrument is standardised. (383) Future users of the self-completed health instruments should be mindful of these issues.

Comprehensiveness of relevant dimensions

The findings show that most participants gave specific examples pertaining to all the EQ-5D-Y dimensions except one. The examples given for 'looking after oneself' that related to lack of basic provision such as soap highlighted contextual differences in terms of how people understood and interpreted this dimension of the instrument. In HICs, washing is simply the action of washing the body. However, in LMICs, the availability of water and soap may be considered by the participants rather than simply the *ability* to wash and dress themselves. This could be a specific Malawian issue but the fact that this has similarly been observed elsewhere

(in Sierra Leone during translation of the EQ-5D into Krio) lead to the conclusion that this is a general problem. The inclusion of needing to get water, i.e., from a well or river in the activity of self-care, was highlighted.

The EuroQol Group should specifically investigate this to ensure that accuracy in the concept is retained cross-culturally. One option could be to provide a clear direction in the translation protocol. The implication of this goes beyond just getting the correct translation as this likely is due to conceptual differences. Using the EQ-5D-Y assumes that the target population fully understand the concept captured by each dimension which was found not to be the case in this research. The lack of clear direction in the translation protocol could likely distort the concept and lead to wrong conclusions. This would subsequently affect future utility valuation studies as well as evaluation of interventions assessed using the EQ-5D-Y.

Combined concepts within one EQ-5D-Y dimension

Related to comprehensibility of relevant dimensions discussed above, is the issue of combined dimensions, i.e., 'having pain or discomfort', and 'feeling worried, sad or unhappy'. As discussed, children might not have related to the 'having pain or discomfort' dimension due to cultural differences. In addition, discomfort was not regarded as a sensation separate from pain. (274) This doctorate research established that while it is possible to retain translation from existing instruments, it is important that these must be given consideration on a case-by-case basis.

The combined dimensions are problematic even with adults, (384) which is a flaw in the design of the EQ-5D instruments. The use of the EQ-5D-Y which combines two or more dimensions into one, might similarly be problematic for children. (270) This is further compounded by the

fact that self-administering any of the EQ-5D instruments among a population that are less educated or children is already a problem. (331) As discussed in chapter 1, there is a high literacy rate in Malawi at 70% but that conversely means another 30% might be illiterate. In addition, although primary education is compulsory, there is low school attendance especially among the poorest. This means that administering self-complete questionnaires could disadvantage this group. This doctorate also established that there is an association between school grade and HRQoL scores. It is important that future users of the Chichewa EQ-5D-Y and EQ-5D-Y-5L consider using the interviewer administered in order to maintain the meaning of discomfort intended here, i.e., physical sensation. This is a preferable approach to ensure that the dimensions are retained as per the existing EQ-5D-Y instrument but also for ease of comparisons across studies. If the dimensions are not fully understood by the target population, there is a risk of using the EQ-5D-Y in this context. This has further consequence on HRQoL data, utility valuation studies and in evaluating interventions for which the values are applied to. The EQ-HWB has shown that it is possible to separate dimensions although development of utility values is another major step that still lies ahead. The conceptual framework in this research has two concepts that are combined: existence/functionality, and strength/energy. Although this might reflect the existing problem within the two EQ-5D-Y dimensions, the combined concepts from the framework can be refined to become one concept to make it easier for comprehension.

Lack of availability of appropriate tariff sets

The tariff sets for the EQ-5D-Y (and subsequently for the EQ-5D-Y-5L in future) are still in development. Recently there has been progress in developing the childhood specific EQ-5D-Y tariffs in Japan, Slovenia and Spain. (201, 385, 386) These utility values have been generated using the EQ-5D-Y valuation protocol which asks adults to imagine themselves as a 10 year old child. (387) This means that these utility values are not necessarily the preferences of

children (388) which is problematic since adults attach different values when valuing for children and for themselves. (389) There is ongoing debate whether to use a hybrid protocol that combines both children/adolescent preferences and those of adults but this raises issues comparibility of utility values from different protocols. (388) Additionally, there are outstanding issues regarding the appropriate valuation technique especially for children/adolescents i.e., TTO versus DCEs or BWS. (390) In comparison to adults therefore, there is still much work to generate utility values available for children and adolescents even though the recent research developments are positive steps. In the meantime, reliance will still be on using adult utility values. However, applying adult utilities in children is not ideal. (141, 391, 392) In Malawi, the problem is made worse by lack of country-specific adult utility values as is the case in most of the sSA region. For a long time, reliance has been on the only available Zimbabwe EQ-5D-3L utility values developed back in 2000. (202) Recently, the EQ-5D-5L utilities have been developed in Ethiopia, Egypt and Uganda. (73, 74, 375) This is good progress. Still, there are questions regarding which utility values should be developed first, those for the EQ-5D-3L (EQ-5D-Y) or EQ-5D-5L (EQ-5D-Y-5L).

The adult EQ-5D-5L was designed to be more sensitive and have a better response system and subsequent utility values than the EQ-5D-3L. However, a systematic review that looked at studies evaluating head-to-head performance (descriptive system and utility values) of the EQ-5D-5L and EQ-5D-3L showed not much difference between the two instruments. (315) While the EQ-5D-5L performed better in terms of ceiling effects and distribution of descriptive system responses, its responsiveness (generated from utility values) was found to be mixed. The latter is of particular interest to countries like Malawi since by design the utility values are intended for resource allocation decisions. However, deciding which instrument to go for will likely go beyond methodological issues and will include purpose for use, setting, availability

of instrument in the local language and utility values among others. In Malawi, the adult EQ-5D-3L has been validated in the Chichewa language (264) but not the EQ-5D-5L. Further, utility values for both instruments have not yet been developed locally and researchers will need to use those from elsewhere. The users of the adult EQ-5D in Malawi will have to decide whether the EQ-5D-5L has the advantage over the EQ-5D-3L bearing in mind these points.

This same argument goes for the youth EQ-5D-Y and EQ-5D-Y-5L versions. Through this thesis research, both the EQ-5D-Y and EQ-5D-Y-5L have been validated in the Chichewa language for Malawi among the general population and sick, children and adolescents. The ideal will be to develop the tariffs for the EQ-5D-Y first since the EQ-5D-Y has been validated in most countries compared to the EQ-5D-Y-5L which is currently under development. Judging by the systematic review above for the adult EQ-5D-3L and EQ-5D-5L, there are mixed results when it comes to utility values for the two instruments. It therefore remains to be seen whether the EQ-5D-Y-5L will have comparative advantage over the EQ-5D-Y for utility values in patients and the general population. The findings from tests of empirical validity support the argument to develop the EQ-5D-Y utility values first since it performed better at predicting stated preferences compared to the EQ-5D-Y-5L.

Use of DALYs for cost-utility analysis in absence of availability of appropriate tariff sets

The global burden of disease by the Institute for Health Metrics and Evaluation (IHME) provides an alternative measure for use in economic evaluation and resource allocation assessment for HTAs. As discussed in chapter 6 of this thesis, there is a surge of HTAs on the African continent. Malawi will likely benefit from the same through the effort of the Malawi MoH, CoM (now KUHeS) and University of York in establishing Health Economics Policy Unit (HEPU). It is anticipated that economic evaluation will be more centralized for informing

policy makers with evidence for resource allocation decisions. The current lack of utility for childhood preference-based HRQoL instruments like the EQ-5D-Y versions means that DALYs will be continued to be relied on for CUA. However, even with the availability of utility values for quantifying QALYs, DALYs can and will still be used as an alternative outcome measure and vice versa. All this will equip HTAs like HEPU with different options for performing economic evaluation. A further advantage of DALYs is that these were generated globally which makes it to compare interventions within the sSA (or African) region and around the globe making generalizability somewhat easy.

THEME IV

8.2.4 Use of generic childhood preference-based health-related quality of life instruments in sub-Saharan Africa

Cost-effectiveness analysis has rarely been used to inform health policy in the sSA region, but its potential has been seen in many HICs. A major barrier has been the availability of appropriate and culturally sensitive instruments. The development of new instruments is constrained by time and resources, implying that the use of childhood preference-based HRQoL instruments will still lag in the sSA region. This problem is amplified by the lack of well-established health technology assessment agencies (HTAs) in the sSA region. A deliberate shift in policy to use cost-utility analysis in resource allocation could be one step to see a step-change in the use of childhood preference-based HRQoL instruments in the sSA region. Although donor partners have set standards for childhood outcomes as formulated in the SDGs, these have primarily focused on reducing child mortality. (192) Most countries in the sSA region including Malawi have succeeded in reducing childhood mortality considering SDGs. There will be a need to measure the HRQoL gained through these SDG interventions using the existing preference-based HRQoL measures.

Advances have been made and some of these challenges have been met elsewhere leading to establishment of HTAs in several countries. This shows that there is potential to advance use of childhood preference-based HRQoL instruments although this will require more work. The work could target developing methods for eliciting utility values from children. A starting point would be the utility valuation work with an adult population to establish if these valuation techniques are understood before moving to children. It is worth noting that a few sSA countries are moving towards setting up HTAs and therefore developing country specific utility values. This can be seen from the development of EQ-5D-3L or EQ-5D-5L utility values in Ethiopia and Uganda, and in a few other African countries. (73, 74, 393) This will likely provide platforms to use utility values across most countries in the sSA region. Additionally, this will also encourage other countries in the sSA region to do the same, leading to a subsequent increase in use of childhood preference-based HRQoL measures in the coming years.

Up to the time of conducting this research, the childhood preference-based HRQoL instruments have only been applied in a small number of sSA countries, raising questions about their applicability across the sSA region. Apart from Uganda, where the HUI3 was applied, the use of the childhood preference-based HRQoL measures was limited to South Africa alone. However, South Africa is not representative of most economies in the sSA region as it is ranked as an upper middle-income country according to the World Bank. (159, 160) The EQ-5D-Y has the potential for use across other sSA settings outside South Africa, albeit with some limitations. This was explored in Malawi through this thesis research with interesting, relevant, and practical results. Researchers in other countries in the sSA region and indeed on the entire African continent will likely build on the successes of this research. The findings from this doctorate are already being put in practice in conjuction with two big projects: POPQOL and

EQ-DAPHNIE.

Population Quality of Life African QOL proposal

The POPulation Quality of Life (POPQOL) study is a proposal that is being prepared for submission to the EuroQol Research Foundation for funding in 2022. (394) The proposal has the following three objectives: i) to assess HRQoL of the general population in countries from the African continent; ii) to investigate characteristics associated with lower HRQoL among different population groups; and iii) to explore health inequalities within and across countries. As evidenced through this doctorate research, LMICs have lagged behind in developing childhood preference-based HRQoL instruments for economic evaluation. Specifically, the use of existing childhood instruments in sSA region has, up to the time of doing this research, been limited to two countries. This goes to highlight the work that lies ahead in as far as evidence based approach for economic evaluation and resource allocation by different HTAs is concerned. POPQOL, for which this researcher is part of, if funded will go a long way in addressing these gaps.

EQ-DAPHNIE Project

The EuroQol- Data for Assessment of Population Health Needs and Instrument Evaluation (EQ-DAPHNIE) is a project that has been funded by the EuroQol Research Foundation. (395) The EQ-DAPHNIE, for which this researcher is also a part of, aim to collect data from the general adult populations globally.The idea is to have population norms and have this data readily available to answer different questions relating to population health. This is another big step to developing an instrument that is representative of the different cultural settings. Of course, these are only first steps to developing universal health concepts. Developing

associated utility values that can be used in economic evaluation may be another goal.

8.3 Strengths and limitations of this doctorate research

The strengths of this research are, firstly, that it is the first such work to develop a conceptual framework of health among children and adolescents in Malawi. The framework identified some of the concepts that are not included in generic health and HRQoL instruments. This is an important finding especially since most instruments have been developed in HICs and are cross-culturally adapted for use in other settings. Rarely has their content validity been evaluated in the sSA settings to ascertain if the instrument being adapted contains adequate content (items, dimensions or domains). This thesis research provides validated tools for use in Malawi in the interim but also highlights some of the gaps existing within health and HRQoL instruments. Over the longer term, research will be needed to refine the conceptual framework developed through this doctorate and generate associated utility values for use in Malawi.

To the best of my knowledge, this is also the first research to compare the EQ-5D-Y and EQ-5D-Y-5L developed using a ranking exercise. As has been documented elsewhere, (283) the importance of rigorous translation and cross-cultural adaptation of HRQoL cannot be overemphasized. The use of a poorly translated instrument may well invalidate otherwise carefully planned research. The card ranking exercise was instrumental in identifying and resolving severity level inversion through a series of interviews. This work has contributed to the evidence of using the ranking exercise as part of the EQ-5D-Y-5L translation protocol. (265) For the psychometric tests that both the EQ-5D-Y and EQ-5D-Y-5L were evaluated for, the instruments generally demonstrated acceptability, reliability (internal consistency) and validity (but not responsiveness) for their use in a Malawian setting with some limitation. Further psychometric tests including test re-test reliability and responsiveness need to be

evaluated in this setting. Conclusively, these validated instruments have potential for use, although being mindful of some limitations of other psychometric properties that were not evaluated.

This study has some limitations that deserve highlighting. The qualitative work in chapter 7 was curtailed and consequently some elements were not investigated as thoroughly as they might have been. For example, finding out more about items under ‘God-given’ and ‘peace of mind’ would have been more informative. How much of an impact the ‘God-given’ concept has on health could not be explored further due to COVID-19 pandemic restrictions on data collection. Additionally, it would have been worthwhile directly comparing the EQ-5D-Y dimensions vis-a-vis the newly developed conceptual framework. Other psychometric quantitative tests like test-retest reliability and responsiveness could similarly not be carried out due to the same restrictions.

Another limitation is that both the quantitative and qualitative samples in this research were drawn from an urban school going population. It is possible that urban and rural needs and comprehension vary even among children and adolescents. This might have an influence in their understanding of health and comprehension of the questionnaires. Related to this is the fact that the children that were school going likely had a better perspective on health due to health education taught in school. It would have been important to interview non-school going children about their own needs and concerns. The sick children involved in the development of the conceptual framework would have established the relevance of concepts in the model among this group.

8.4 Potential use of childhood health-related quality of life instruments and recommendations for research and policy

8.4.1 Potential use of childhood preference-based HRQoL instruments in Malawi

In absence of childhood preference-based HRQoL instruments in Malawi this doctorate has adapted and validated the two EQ-5D-Y versions, and another two generic HRQoL instruments (PedsQL™ 4.0 child and teen self-report). While the EQ-5D-Y can be used with limitation as utility instrument in economic evaluation in this setting, its use as a health status instrument was found to be greatly insufficient. The EQ-5D-Y and EQ-5D-Y-5L have demonstrated some linguistic problems, evidence to lack content validity, and limited psychometric validity. The findings highlight the need to adapt another utility instrument in Chichewa for use alongside the EQ-5D-Y in economic evaluation. For use to measure health or HRQoL in this setting, the validated PedsQL™ 4.0 child self-report and PedsQL™ 4.0 teen self-report are possible options to be used alongside the EQ-5D-Y. Alternatively, using the already validated Chichewa versions of PROMIS-25 (367) and WHOQOL-BREF (195) could be an option.

In addition, this doctorate has developed a conceptual framework that shows the lack of content validity of existing instruments in relation to what children and adolescents consider important in a Malawian setting. It would be important to build on this work and come up with an instrument that considers the local perception of health. However, as discussed above it is important to weigh need of local instruments against generalizability. The work that is being planned ahead, POPQOL and EQ-DAPHNIE, will look at addressing these issues not only at a local (Malawian) level but across other settings as well. The multi country studies will look at health definition and generating an instrument on the African continent (POPQOL) and global (EQ-DAPHNIE).

8.4.2 Application of ranking exercise in the cross-cultural adaptation of the EQ-5D-Y-5L

The ranking exercise was instrumental in identifying translation problems in understanding the qualifiers for the five different severity levels. It is highly likely that had it not been for the card ranking exercise, coming up with accurate qualifiers to establish linguistic equivalence and hierarchical ranking would not have been adequately established. It is therefore highly recommended to use the ranking exercise in adapting the EQ-5D-Y-5L cross-culturally.

Additionally, the struggle by some children to provide fitting examples for the EQ-5D-Y 'looking after oneself' dimension points to the need to give attention when cross-culturally translating this concept. One recommendation would be that the EuroQol Group should provide further guidance on environmental factors in the translation protocol. It is believed that omission of such clarification as water availability and 'mobility' issues might affect item equivalence particularly in some LMICs. (273) These findings endorse this recommendation.

8.5 Further research

8.5.1 Further understanding of health and development of childhood health instruments

This research developed a conceptual framework for health in Malawi. The findings show that there are concepts that are not usually included in generic health instruments. Locally, there has been effort to develop role functioning tool from among children. (382) The PROMIS-25 and WHOQOL BREF have been validated among children with different health conditions in Malawi. (195, 367) Both these efforts have either adapted existed instruments or developed a tool for a specific function and therefore do not have unique concepts identified by this research. Concepts such as 'God-given', 'acceptance of illness', and 'peace of mind' would have benefited from more probing. Also, the HIV/AIDS link to health definition would have

benefited from further exploration but could not due to the same COVID-19 restriction.

8.5.2 Development of EuroQol tariff sets.

Currently, there are no Malawi utility values for the adult EQ-5D-3L, EQ-5D-5L and let alone for the EQ-5D-Y. The empirical validity carried out in this research was based on adult utility values and from a different geographical location. Ideally, the sSA regional utility values should have been preferred but none are available for both the EQ-5D-3L and EQ-5D-5L. The EuroQol Group advised use of adult USA utility values instead for easy comparison as these are available for both the three and five response levels. The implication of using the USA utility values is twofold. Firstly, even though these are available, they were derived using two different utility valuation protocols which could translate to different outcomes. Secondly, the USA utility values might reflect preferences that are not valued in the same way in the sSA setting. This is an area that need further research in most sSA setting and Malawi in particular.

8.5.3 Further evaluation of the EQ-5D-Y-5L

The study found not much difference in most psychometric properties of the EQ-5D-Y and EQ-5D-Y-5L in Malawi. To a large extent both the EQ-5D-Y and EQ-5D-Y-5L showed acceptability, reliability and limited validity, and can be used with limitation in this setting. However, future studies should evaluate other psychometric properties like test-retest and responsiveness which could not be evaluated due to COVID-19 restriction as indicated above. Additionally, further psychometric testing might be required hitherto empirical validity to support the EQ-5D-Y-5L as a generic preference based HRQoL instrument.

8.6 Policy relevance and implications

Based on the findings of this research, the following are recommendations and policy implications as summarised in Table 8.2 below.

Table 8.2 Guide for childhood HRQoL use in Malawi

Choice of childhood instrument to use in Malawi	Age (years)	Mode of Administration	PBM (Yes/ No)	Validated in Chichewa for Malawi		Utilities availability			If utilities unavailable, alternative	Future work for Malawi
				Yes/No (version)	Source	In sSA (Yes/No) (source)	Around the globe (Yes/No) (source)	Perspective		
i) Childhood PBM Instrument										
EQ-5D-Y-3L	8-15yrs	Self and proxy	Yes	Yes (self-report)	This work	No	Yes (Japan, (201); Slovenia,(385); Spain (295))	Adult imagining they are 10yr old child	N/A	Develop child specific utilities, bolt-on
EQ-5D-Y-5L	8-15yrs	Self and proxy	Yes	Yes (self-report)	This work	No	No	N/A, EQ-5D-Y-5L still in development	Adult EQ-5D-5L (although not recommended)	Develop child specific utilities and bolt-on
EQ-5D-3L	16-17yrs; 18yrs +	Self and proxy	Yes	Yes (self-report)	Chokotho et al (264)	Yes (Zimbabwe (396))	Yes	Adults	N/A	Develop utilities
EQ-5D-5L	16-17yrs; 18yrs +	Self and proxy	Yes	No	N/A	Yes (Ethiopia (73); Uganda (375))	Yes	Adults	N/A	Develop utilities

ii) Non PBM HRQoL instruments										
PedsQL child self-report	8-12	Self and proxy	No	Yes (self-report)	This work	N/A	N/A	N/A	N/A	N/A
PedsQL child self-report	13-18	Self and proxy	No	Yes (self-report)	This work	N/A	N/A	N/A	N/A	N/A
PROMIS-25	8-17	Self	No	Yes	Westmore-land et al (367)	N/A	N/A	N/A	N/A	N/A
WHOQOL BREF	11-18	Self	No	Yes	Colbourn et al (195)	N/A	N/A	N/A	N/A	N/A
iii) Need for contextual instrument										
Conceptual framework for health*	8-17yrs	Self (11-17yrs); Proxy (8-10yrs)	No	No	This work	N/A	N/A	N/A	N/A	N/A

PBM: preference-based HRQoL instrument

*Conceptual framework developed here provides groundwork for a localized instrument

On a local (Malawi) level, promotion of the adaptation of health measurement will contribute to providing tools necessary for evaluating the effectiveness of health interventions aimed at children and adolescents. Further, the adapted and any future scales will contribute to cost-utility analysis of different children and adolescent health interventions. The cost-utility analysis will assist in efficient resource allocation decisions in the design of EHP in pursuit of UHC as well as SDGs. The observed low utilization of health measurements in Malawi suggests the need for more effort to improve usage of available instruments and developing new ones. The newly validated Chichewa versions of the EQ-5D-Y and EQ-5D-Y-5L, as well as heath status measurement using PedsQL™ 4.0 child self-report and PedsQL™ 4.0 have the potential for such use among children and adolescents in Malawi. Further, the conceptual framework provides evidence that there is need to either develop local instruments or use of bolt-ons to existing dimensions.

At sSA level, the measurement of HRQoL considering SDGs should be considered. The validation of the Chichewa language versions of EQ-5D-Y and EQ-5D-Y-5L, as well as heath status measurement using PedsQL™ 4.0 child self-report and PedsQL™ 4.0 show that this is possible. Cross-cultural adaptation is possible but there is need for close adherence to translation protocol to achieve linguistic and conceptual equivalence. This research highlighted the need for thorough adaptation processes to adequately resolve translation issues. There is need to utilize other ways to have adapted versions that maintain same meaning and become suitable for use in the sSA region. Alternatively, developing new instruments will ensure that cultural expectation in measuring what is important is met in this setting. However, this process can be time consuming and expensive. The decision to opt for adaptation or construct new instrument will depend on the purpose of the study. For example, if the study is intended to evaluate the cost-effectiveness between interventions in children and adolescents, then the

current validated EQ-5D-Y and EQ-5D-Y-5L would be adequate. However, if the study is looking at measuring health improvement, then these instruments would not be adequate and would need to be complemented with other instruments like the PedsQL™ 4.0. Further, if a study is looking at a specific objective or disease area then a new instrument would need to be developed altogether. This is the case with the Chichewa MDAT specifically developed for assessment of neurological disabilities in children. (382)

On the global level, development of generic childhood preference-based HRQoL instruments has been limited to HICs. Several reasons have been responsible for this including well established HTAs which are a driver for the resource allocation agenda in these countries. Despite the prominence of childhood preference-based HRQoL instruments in HICs, developers have only involved children to some extent in scale development. This is an area that need strengthening. There are also few childhood preference-based HRQoL instruments that have available utility values developed from children perspective. While there are several arguments for this, it is apparent that there has been progress towards development of valuation techniques that children can relate to for a valuation exercise. Recently, utility values have been generated directly from children and adolescents using DCE (children and adolescent preferences) for EQ-5D-Y (397), BWS for EQ-5D-Y (398) and BWS (adolescent perspective) for CHU-9D (139, 142, 399). This is at an early stage but is an interesting and challenging area that need further strengthening on a global level through in-country collaborative multi-country research and adaptation projects. (138, 400)

8.7 Dissemination of research findings

This doctorate research identified areas that would benefit from further research. Some of these findings have been published and others are undergoing development for peer review. Findings

from the literature review of development, adaptation and use of existing childhood preference-based HRQoL as well as the ranking exercise development (as reported in chapter 2) have been published in peer reviewed journals. (Appendix 26) The cross-cultural adaptation of the EQ-5D-Y-5L in chapter 4 has similarly been published in a peer reviewed journal. (Appendix 10.26.1) The rest of the findings (development of conceptual framework and psychometric performance of the EQ-5D-Y and EQ-5D-Y-5L, as well as PedsQL™ 4.0 child self-report and PedsQL™ 4.0) are to be submitted for publication. Additionally, some of these findings have been shared at international conferences and most importantly at the local policy meetings (see development portfolio). The full findings will be disseminated through policy briefs with the Malawi's Health Economics Policy Unit, MoH and Health Economics Association, and ISOQOL and EuroQol conferences. Several policy recommendations have been identified to promote health-measurement and economic evaluation in Malawi. This is both at the Malawi government level and research context, through direct interactions and through the professional health economics association.

8.8 Summary conclusions

In conclusion, this doctorate research has demonstrated that it is possible to get an in-depth understanding of health from children and adolescents in Malawi and how to measure this. Further, the research has also established that it is possible to cross-culturally validate an adapted version from existing instruments. However, the existing instruments have not aligned themselves completely to perceptions of perceived health concepts in this setting among children and adolescents, nor have they met all the psychometric requirements. While it is possible to use existing self-completion instruments by carefully going through the adaptive process, children especially below the age of 12 years might benefit more from interviewer administered instruments or other formats. Additionally, while these instruments perform

better in several psychometric properties but not discriminant validity by grade. Education likely plays an important role in understanding / comprehensibility of the instruments. The self-completion of these instruments might be limited to children and adolescents who are in school.

In all, this doctorate has contributed to better understanding of health through the development of a conceptual framework, provided evidence for the need for elaborate steps required for cross-cultural validation of existing instruments. The work has further validated into Chichewa language, four HRQoL instruments and contributed evidence for the use of ranking exercise as part of the EQ-5D-Y-5L translation protocol. The validated Chichewa EQ-5D-Y and EQ-5D-Y-5L will be instrumental to assessing health interventions intended for children and adolescents. This will be an important step in relevant economic evaluation and contribute to future design of an EHP in Malawi. The inability to perform the test-retest reliability and responsiveness will mean that the instrument needs to be evaluated how they perform overtime. Additionally, this work sets the groundwork to develop utility values for the EQ-5D-Y and EQ-5D-Y-5L in Malawi.

9 References

1. Malawi Government. https://www.malawi.gov.mw/. accessed last on 3rd March 2021.
2. National Statistical Office. 2018 Malawi Population and Housing Census. National Statistics Office. Zomba, Malawi.
3. The World Bank WD, USA. http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=CountryProfile&Id=b450fd57&tbar=y&dd=y&inf=n&zm=n&country=ZAF. last assessed 14th September 2017.
4. World Bank U. https://www.worldbank.org/en/country/malawi/overview. last accessed 3rd March 2021.
5. World Bank U. https://data.worldbank.org/indicator/SH.DYN.MORT?end=2019&start=1960. last accessed 3rd March 2021.
6. National Statistical Office (NSO) [Malawi] and ICF. Malawi Demographic and Health Survey 2015-2016. Zomba, Malawi, and Rockville, Maryland, USA: NSO and ICF, 2017.
7. Internal IDEA, Elections Commissions Forum of SADC countries. Lessons from Malawi's Fresh Presidential Elections of 23 June 2020. Int IDEA and Elections Commissions Forum of SADC countries. Webinar, 31 August 2020.
8. Phiri I, Masanjala W. Willingness to pay for micro health insurance in Malawi. In: Rösner H-J, Leppert G, Degens P, et al., eds., Hand book of Micro Health Insurance in African. 1st edition ed. Berlin: Lit Verlag, 2012.
9. Kayambazinthu E. The Language Planning Situation in Malawi. Journal of Multilingual and Multicultural Development. 1998; 19: 369-439.
10. National Statistical Office. Malawi Demographic and Health Survey 2015/16.

11. Government M. Health Sector Strategic Plan II 2017-2022. Towards Universal Health Coverage. 2017.
12. Malawi Ministry of Health and Population. Malawi Facilitty Registry http://zipatala.health.gov.mw/. last accessed 6th April 2021.
13. King M. Medical care in developing countries, A primer on the medicine of poverty and a symposium from Makerere. London: Oxford University Press, 1966.
14. Levey R, Ilana, Gitonga N, et al. Malawi Private Health Sector Assessment. Strenghening Health Outcomes through the Private Sector Project. Bethseda, MD: Abt Associates Inc., May 2011.
15. Chirwa ML, Kazanga I, Faedo G, et al. Promoting universal financial protection: contracting faith-based health facilities to expand access--lessons learned from Malawi. Health Res Policy Syst. 2013; 11: 27.
16. Ochalek J, Manthalu G, Nkhoma D, et al. Supporting the Development of Health Benefits Packages (HBPs): Principles and Initial Assessment for Malawi. In: Revill P, Suhrcke M, Moreno-Serra R, et al., eds., Global Health Economics Shaping Health Policy in Low- and Middle- Income Countries. Singapore: World Scientific Publishing Co. Pte. Ltd, 2020.
17. Mwase T. Health Financing Policy Reforms in Eastern Central and Southern Africa (ECSA)- Health Community Region. CHE Research Paper. Centre for Health Economics: University of York, 2021.
18. Akin JS, Birdsall N, De Ferranti DM. Financing Health Services in Developing Countries: An Agenda for Reform. Washington DC: The World Bank, 1987.
19. WHO. Macroeconomics and Health: Investing in Health for Economic Development. 2001.

20. Zere E, Walker O, Kirigia J, et al. Health financing in Malawi: Evidence from National Health Accounts. BMC Int Health Hum Rights. 2010; 10: 27.

21. Pearson M. Impact Evaluation of the Sector Wide Approach (SWAp), Malawi. Human Development Resource Center, UKAID: DFID, 2010.

22. Carlson C, Boivin M, Chirwa A, et al. Malawi Health Swap Mid-Term Review. Norad Collected Reviews. Oslo: NORAD, 2008.

23. Phoya A, Araru T, Kachala R, et al. Setting Strategic Health Sector Prirorities in Malawi. . Disease Control Priorities in Developing Countries. 3rd ed, 2010.

24. WHO and Ministry of Health. Malawi National STEPS Survey for Chronic Non-Communicale Diseases and their Risk Factors Final Report. Lilongwe, Malawi: Malawi Ministry of Health, 2010.

25. World Health Organisation. International statistical classification for dieases and related health problems, tenth revision. Geneva, Switzerland: WHO, 2010.

26. McGuire F, Revill P, Twea P, et al. Allocating resources to support universal health coverage: development of a geographical funding formula in Malawi. BMJ Glob Health. 2020; 5.

27. Malawi Government. Malawi Growth and Development Strategy. Lilongwe: Ministry of Finance and Ministry of Economic Planning and Development, 2005.

28. Malawi Government. Malawi Growth and Development Strategy II 2011-2016. Lilongwe: Ministry of Economic Planning and Development, 2010.

29. Malawi Government. Malawi Growth and Development Strategy MGDS III 2017-2022 Lilongwe: Ministry of Finance, Economic Planning and Development, 2016.

30. Ministry of Health. Malawi Health Sector Strategic Plan 2011-2016. Lilongwe, Malawi: Ministry of Health, 2012.

31. Abiiro GA, Mbera GB, De Allegri M. Gaps in universal health coverage in Malawi: a qualitative study in rural communities. BMC Health Serv Res. 2014; 14: 234.

32. Ministry of Health. Draft Joint Annual Report for 2018/19 fiscal year., 2019.

33. Ministry of Health. Malawi Health Financing Situation Analysis. Lilongwe: Ministry of Health, 2011.

34. Ministry of Health. Health Financing Strategy in Malawi- draft 2. 2020.

35. Chansa C, Mwase T, Matsebula TC, et al. Fresh Money for Health? The (False?) Promise of "Innovative Financing" for Health in Malawi. Health Systems & Reform. 2018; 4: 324-35.

36. Bailey N, Mandeville KL, Rhodes T, et al. Postgraduate career intentions of medical students and recent graduates in Malawi: a qualitative interview study. BMC Medical Education. 2012; 12: 87.

37. Muula AS. Is there any solution to the "brain drain" of health professionals and knowledge from Africa? Croatian Medical Journal. 2005; 46: 21-29.

38. Borghi J, Munthali S, Million LB, et al. Health financing at district level in Malawi: an analysis of the distribution of funds at two points in time. Health Policy Plan. 2018; 33: 59-69.

39. Education Encyclopedia. Stages of growth in child development. Available from: http://education.stateuniversity.com/pages/1826/Child-Development-Stages-Growth.html#ixzz0j0jMHgRBExternal assessed on 14th April 2022.

40. Shonkoff JP, Boyce WT, McEwen BS. Neuroscience, Molecular Biology, and the Childhood Roots of Health Disparities: Building a New Framework for Health Promotion and Disease Prevention. JAMA. 2009; 301: 2252-59.

41. Munthali AC, Mvula PM, Silo L. Early childhood development: the role of community based childcare centres in Malawi. SpringerPlus 2014; 3: 305.

42. UNICEF. Country Profile: Malawi. Available from https://data.unicef.org/country/mwi/. last accessed 2nd April 2021.

43. WHO. https://www.who.int/immunization/monitoring_surveillance/data/mwi.pdf?ua=1. WHO and UNICEF estimates of national immunization coverage. . last accessed 7th April 2021.

44. National AIDS Commission. Annual report for July 2019 to June 2020 financial year. Lilongwe, Malawi: National AIDS Commission, 2020.

45. van Lettow M, Landes M, van Oosterhout JJ, et al. Prevention of mother-to-child transmission of HIV: a cross-sectional study in Malawi. Bull World Health Organ. 2018; 96: 256-65.

46. Ministry of Health. The Mid-Term Review of the Malawi Health Sector Strategic Plan II, March 2021 draft. Lilongwe, Malawi: Department of Planning and Policy Development, 2021.

47. AVERT. HIV and AIDS in Malawi https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/malawi date accessed 13th April 2022. last review date 01st March 2022.

48. WHO. Global tuberculosis report 2020. Geneva, Switzerland: World Health Organization, 2020.

49. da Silva ICM, Franca GV, Barros AJD, et al. Socioeconomic Inequalities Persist Despite Declining Stunting Prevalence in Low- and Middle-Income Countries. J Nutr. 2018; 148: 254-58.

50. Malawi Malaria National Control Programme. Progress of key indicators and level of M&E activities implementation. Lilongwe: Ministry of Health, 2021.

51. Nshimyiryo A, Hedt-Gauthier B, Mutaganzwa C, et al. Risk factors for stunting among children under five years: a cross-sectional population-based study in Rwanda using the 2015 Demographic and Health Survey. BMC Public Health. 2019; 19.

52. Martorell R, Khan LK, Schroeder DG. Reversibility of stunting: epidemiological findings in children from developing countries. Eur J Clin Nutr. 1994; 48: S45-57.

53. Ciancio A, Kampfen F, Kohler HP, et al. Health screening for emerging non-communicable disease burdens among the global poor: Evidence from sub-Saharan Africa. J Health Econ. 2021; 75: 102388.

54. WHO. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. Geneva, Switzerland: WHO, last accessed 10th April 2021.

55. International Diabetes Federation. IDF Diabetes Atlas. 8th ed. Brussels, Belgium: International Diabetes Federation, 2017.

56. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006; 3: e442.

57. Institute for Health Metrics and Evaluation. http://www.healthdata.org/malawi. last accessed 7th April 2021.

58. Price AJ, Crampin AC, Amberbir A, et al. Prevalence of obesity, hypertension, and diabetes, and cascade of care in sub-Saharan Africa: a cross-sectional, population-based study in rural and urban Malawi. The Lancet Diabetes & Endocrinology. 2018; 6: 208-22.

59. Lacroix P, Houehanou C, Preux PM, et al. NCD risk factors in Malawi: population characteristics matter. The Lancet Diabetes & Endocrinology. 2018; 6: 163-64.

60. Ochalek J, Revill P, Manthalu G, et al. Supporting the development of a health benefits package in Malawi. BMJ Global Health. 2018; 3.

61. Twea P, Manthalu G, Mohan S. Allocating resources to support universal health coverage: policy processes and implementation in Malawi. BMJ Glob Health. 2020; 5.

62. Drummond M, Sculpher M, Torrance G, et al. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford: Oxford University Press, 2005.

63. Ungar WJ, Santos MT. Paediatric Economic Database Evaluation: Trends in Economic Evaluation. http://pede.ccb.sickkids.ca/pede/trends.jsp. Last assessed 25th February 2022.

64. Kwon J, Kim SW, Ungar WJ, et al. Patterns, trends and methodological associations in the measurement and valuation of childhood health utilities. Qual Life Res. 2019.

65. Kromm SK, Bethell J, Kraglund F, et al. Characteristics and quality of pediatric cost-utility analyses. Qual Life Res. 2012; 21: 1315-25.

66. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health. 2003; 6: 9-17.

67. Ishaque S, Karnon J, Chen G, et al. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). Qual Life Res. 2018.

68. Devlin N. Preference-based measure' is misleading – can we agree on something better? In: Sampson C, ed.: THE ACADEMIC HEALTH ECONOMISTS' BLOG, 2020.

69. Maheswaran H, Petrou S, MacPherson P, et al. Cost and quality of life analysis of HIV self-testing and facility-based HIV testing and counselling in Blantyre, Malawi. BMC Med. 2016; 14: 34.

70. Zwerling AA, Sahu M, Ngwira LG, et al. Screening for Tuberculosis Among Adults Newly Diagnosed With HIV in Sub-Saharan Africa: A Cost-Effectivnes Analysis. J Acquir Immune Defic Syndr. 2015; 70: 83-90.

71. Maheswaran H, Petrou S, Cohen D, et al. Economic costs and health-related quality of life outcomes of hospitalised patients with high HIV prevalence: A prospective hospital cohort study in Malawi. PLoS One. 2018; 13: e0192991.

72. Zeng W, Sun D, Mphwanthe H, et al. The impact and cost-effectiveness of user fee exemption by contracting out essential health package services in Malawi. BMJ Glob Health. 2019; 4: e001286.

73. Welie AG, Gebretekle GB, Stolk E, et al. Valuing Health State: An EQ-5D-5L Value Set for Ethiopians. Value Health Reg Issues. 2019; 22: 7-14.

74. Al Shabasy SA, Abbassi MM, Finch AP, et al. The EQ-5D-5L Valuation Study in Egypt. PharmacoEconomics. 2021.

75. University of York. Thanzi La Onse. https://www.york.ac.uk/igdc/research/thanzi-la-onse/. accessed on 16 April 2022.

76. Shah K. A brief review of concepts: health, quality of life, health-related quality of life and well-being. EuroQol Working Paper Series. Rotterdam, Netherlands: EuroQol Research Foundation & Office of Health Economics, UK, 2017.

77. WHO. The constitution of the World Health Organization. WHO Chron. 1947; 1: 29.

78. Eiser C, Cotter I, Oades P, et al. Health-related quality of life measures for children. Int J Cancer. 1999; Supplement 12: 87-90.

79. Eiser C, Morse R. The measurement of quality of life in children: past and future perspectives. Developmental and behavioral pediatrics. 2001; 22: 248-56.

80. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics. 2016; 34: 645-9.

81. Matza LS, Swensen AR, Flood EM, et al. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. Value Health. 2004; 7: 79-92.

82. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychological Medicine. 1998; 28: 551-8.

83. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. 1989; 27: S217-32.

84. Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics. 1995; 7: 490-502.

85. Romero M, Vivas-Consuelo D, Alvis-Guzman N. Is Health Related Quality of Life (HRQoL) a valid indicator for health systems evaluation? Springer Plus. 2013; 2: 664.

86. Krabbe PF. The Measurement of Health and Health Status Concepts, Methods and Applications from a Multidisciplinary Perspective. 1st ed.: Academic Press, 2017.

87. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technology Assessment. 2014; 18.

88. Wallander J, Schmitt M, Koot H. Quality of Life Measurement in Children and Adolescents: Issues, Instruments, and Applications. Journal of Clinical Psychology. 2001; 57: 571-85.

89. Earnshaw J, Lewis G. NICE Guide to the methods of technology appraisal. Pharmacoeconomics. 2008; 26: 725-27.

90. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princetown: Princetown University Press, 1953.

91. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. Journal of Health Services Research and Policy. 1999; 4: 174-84.

92. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. Annu Rev Public Health. 2002; 23: 115-34.

93. Bill and Melinda Gates Foundation. Methods for Economic Evaluation Project (MEEP) - Final Report. A partnership between Bill and Melinda Gates Foundation, NICE International, the Health Intervention and Technology Assessment Program (Thailand), and the University of York, Centre for Health Economics. 2014.

94. World Bank U. World Development Report 1993. Investing in Health. New York: Oxford University Press, 1993.

95. Fox-Rushby J. Disability Adjusted Life Years (DALYs) For Decision-Making? An overview of the literature. Office of Health Economics: BSC Print Ltd., London, 2002.

96. Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. The Lancet. 2012; 380: 2129-43.

97. Loomes G, McKenzie L. The use of QALYs in health care decision making. Social Science & Medicine. 1989; 28: 299-308.

98. Neumann PJ, Goldie SJ, Weinstein MC. Preference-Based Measures in Economic Evaluation in Health Care. Annu Rev Public Health. 1986; 21: 587-611.

99. Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998; 2: 1-74.

100. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. PLoS One. 2015; 10: e0135672.

101. 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: 1-114.

102. Lorgelly PK, Doble B, Rowen D, et al. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. Qual Life Res. 2017; 26: 1163-76.

103. Rowen D, Brazier J, Ara R, et al. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. PharmacoEconomics. 2017; 35: 33-41.

104. Brazier J. Valuing health states for use in cost-effectiveness analysis. Pharmacoeconomics. 2008; 26: 769-79.

105. Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, ed., Enclyopedia of health economics. San Diego: Elsevier, 2014.

106. Brazier J, Deverill M, Green C, et al. A review of the use of health status measures in economic evaluation. Health Technol Assess. 1999; 3.

107. Stevens KJ. How Well Do the Generic Multi-attribute Utility Instruments Incorporate Patient and Public Views Into Their Descriptive Systems? Patient. 2016; 9: 5-13.

108. Bevans KB, Riley AW, Moon J, et al. Conceptual and methodological advances in child-reported outcomes measurement. Expert Rev Pharmacoecon Outcomes Res. 2010; 10: 385-96.

109. Stevens K, Palfreyman S. The use of qualitative methods in developing the descriptive systems of preference-based measures of health-related quality of life for use in economic evaluation. Value Health. 2012; 15: 991-8.

110. Froberg DG, Kane RL. Methodology for measuring health-state preferences—I: Measurement strategies. Journal of Clinical Epidemiology. 1989; 42: 345-54.

111. Brazier J, Akehurst R, Brennan A, et al. Should patients have a greater role in valuing health states? Appl Health Econ Health Policy. 2005; 4: 201-8.

112. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35: 1095-108.

113. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care. 1989; 5: 559-75.

114. Brazier J, Ratcliffe J, Salomon JA, et al. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press, 2007.

115. Paravastu S, Michaels J. Outcomes and health economic issues in surgery. In: Paterson-Brown S, ed., Core topics in general and emergency surgery- Companion to specialist surgical practice. Fifth ed. Edinburgh: Saunders/Elsevier 2014.

116. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. Socio-Economic Planning Sciences. 1976; 10: 129-36.

117. Dolan P, Gudex C, Kind P, et al. Valuing health states: A comparison of methods. Journal of Health Economics. 1996; 15: 209-31.

118. Petrou S, Henderson J. Preference-based approaches to measuring the benefits of perinatal care. Birth. 2003; 30: 217-26.

119. Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012; 31: 306-18.

120. Luce RD, Turkey JW. Simultaneous conjoint measurement: a new type of fundamental measurement. J Math Psychol. 1964; 1: 1-27.

121. Parvin S, Wang P, Uddin J, et al. Using best-worst scaling method to examine consumers' value preferences: A multidimensional perspective. Cogent Business & Management. 2016; 3.

122. Chen G, Ratcliffe J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. Pharmacoeconomics. 2015; 33: 1013-28.

123. Rowen D, Brazier J, Van Hout B. A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. Medical Decision Making. 2015; 35.

124. Buckingham K. A note on HYE (healthy years equivalent). Journal of Health Economics. 1993; 12: 301-09.

125. Ubel P, Loewenstein G, Jepson C. Whose qulity of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res. 2003; 12: 599-607.

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

127. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. Journal of Chronic Diseases. 1978; 31: 697-704.

128. de Wit G, Busschbach J, de Charro F. Sensitivity and perspective in the valuation of health status: Whose values count? Health Econ. 2000; 9: 109-26.

129. Dolan P, Gudex C, Kind P, et al. The time trade-off method: results from a general population study. Health Econ. 1996; 5: 141-54.

130. Ungar W. Challenges in health sate valuation in paediatric economic evaluation. Pharmacoeconomics. 2011; 29: 641-52.

131. Piaget J. The theory of stages in cognitive development. In: Green DR, Flamer GB, eds., Measurement and Piaget: McGraw-Hill, 1971.

132. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: an analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes. 2007; 5: 1.

133. Binger CM, Ablin AR, Feuerstein RC, et al. Childhood Leukemia. Emotional impact on patient and family. New England Journal of Medicine. 1969; 280: 414-18.

134. Rand S, Caiels J. Using Proxies to assess Quality of Life: A Review of the Issues and Challenges. Canterbury: University of Kent, 2015.

135. Khadka J, Kwon J, Petrou S, et al. 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. Social Science & Medicine. 2019; 240.

136. Eiser C, Morse R. A review of measures of qulaity of life for children with chronic illness. Arch Dis Child. 2001; 84: 205-11.

137. Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics. 2012; 30: 729-47.

138. Rowen D, Rivero-Arias O, Devlin N, et al. 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.

139. Ratcliffe J, Couzner L, Flynn T, et al. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. Appl Health Econ Health Policy. 2011; 9: 15-27.

140. Moodie M, Richardson J, Rankin B, et al. 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: 1014-27.

141. Petrou S. Should health gains by children be give the same value as health gains by adults in an economic evaluation framework. In: Ungar W, ed., Economic Evaluation in Child Health. Oxford: Oxford Unviersity Press, 2010.

142. Ratcliffe J, Huynh E, Stevens K, et al. 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: 486-96.

143. Richardson J, Day N, Peacock S, et al. Measurement of the Quality of Life for Eonomic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. The Australian Economic Review. 2004; 37: 62-88.

144. Brazier J, Ara R, Rowen D, et al. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. Pharmacoeconomics. 2017; 35: 21-31.

145. (FDA) UDoHaHSFaDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. In: FDA U, ed., Federal Register. Rockville, MD: FDA, 2009.

146. Trust MO. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002; 11: 193-205.

147. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013; 22: 1889-905.

148. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checlist for evaluating the methdolodgical quality of studies on measurement properties: A clarification of its content. BMC Mecical Research Methodology. 2010; 10.

149. Health USDo, Human Services FDACfDE, Research, et al. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. last accessed 1 December 2017.

150. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

151. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016; 20: 105-13.

152. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009; 62: 1006-12.

153. Canaway AG, Frew EJ. Measuring preference-based quality of life in children aged 6-7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y--the WAVES pilot study. Qual Life Res. 2013; 22: 173-83.

154. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. Health Econ. 1999; 8: 41-51.

155. Critical Appraisal Skills Programme. CASP Systematic Review Checklist. Available from: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Case-Control-Study-Checklist-2018_fillable_form.pdf accessed on 16 April 2022. 2018.

156. Kang E. Validity of Child Health-6 Dimension(CH-6D) for Adolescents. Value in Health. 2016; 19.

157. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res. 2009; 18: 1105-13.

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

159. 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: 875-86.

160. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res. 2010; 19: 887-97.

161. The EuroQol Group. https://euroqol.org/eq-5d-instruments/. last accessed 30th March 2021.

162. Wolstenholme JL, Bargo D, Wang K, et al. Preference-based measures to obtain health state utility values for use in economic evaluations with child-based populations: a review and UK-based focus group assessment of patient and parent choices. Quality of Life Research. 2018.

163. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. Value Health. 2012; 15: 1092-9.

164. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. Qual Health Res. 2010; 20: 340-51.

165. Stevens K. Assessing the Performance of a New Generic Measure of Health-Related Quality of Life for Children and Refining it for Use in Health State Valuation. Appl Health Econ Health Policy. 2011; 9: 157-69.

166. Richardson JR, Peacock SJ, Hawthorne G, et al. Construction of the descriptive system for the Assessment of Quality of Life AQoL-6D utility instrument. Health Qual Life Outcomes. 2012; 10: 38.

167. Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolesence: A sixteen-dimensional health-related measure (16D). Quality of Life Research. 1996; 5: 205-11.

168. 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: 532-8.

169. Kaplan R, Sieber W, Ganiats T. The quality of well-bing scale: comparison of the interviewer-administered version with a self-administered questionnaire. Psychol Health. 1997; 12: 783-91.

170. Seiber WJ, Groessl EJ, David KM, et al. Quality of Well Being Self-Administered (QWB-SA) Scale user's manual.: San Diego: Health Services Research Center, University of California, 2008.

171. Kaplan R, Ganiats T, Sieber W, et al. The Quality of Well-Being Scale: critical similarities and differences with SF-36. International Journal for Quality of Health Care. 1998; 10: 509-20.

172. Boyle MH, Furlong W, Feeny D, et al. Reliability of the Health Utilities Index-Mark Ill used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. Quality of Life Research. 1995; 4: 249-57.

173. Feeny D. The Health Utilities Index: A tool for assessing health benefits. PRO Newsletter. 2005; 34.

174. Torrance G, Furlong W, Feeny D, et al. Multi-attribue preference functions. Health utilities index. Pharmacoeconomics. 1995; 7: 503-20.

175. Torrance G, Feeny D, Furlong W, et al. Multriattribute utility function for a comprehensive health status classification system: Health Utlities Index Mark 2. Med Care. 1996; 34: 702-22.

176. Richardson J, Peacock S, Iezzi A, et al. Construction and Validation of the Assessment of Quality of Life (AQoL) Mark II Instrument. Research Paper 24. Melbourne: Monash University, 2007.

177. Rowen D, Mulhern B, Stevens K, et al. Estimating a Dutch Value Set for the Pediatric Preference-Based CHU9D Using a Discrete Choice Experiment with Duration. Value Health. 2018; 21: 1234-42.

178. The EuroQol Group. EQ-5D-3L valuation. Available from https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/. Accessed on 16 April 2022.

179. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60: 34-42.

180. Sims-Williams HJ, Sims-Williams HP, Mbabazi Kabachelor E, et al. Quality of life among children with spina bifida in Uganda. Arch Dis Child. 2017; 102: 1057-61.

181. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness - a longitudinal, analytical study. Health Qual Life Outcomes. 2017; 15: 12.

182. Jelsma J, Ramma L. How do children at special schools and their parents perceive their HRQoL compared to children at open schools? Health and Qual Life Outcomes. 2010; 8: 72.

183. Jelsma J. A comparison of the performance of the EQ-5D and the EQ-5D-Y health-related quality of life instruments in South African children. Int J Rehabil Res. 2010; 33: 172-7.

184. 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: 391-6.

185. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychol Bull. 1987; 101: 213-32.

186. Ware Jr. JE, Snow K, Kosinski M, et al. SF-36 Health Survey Manual and Interpretation Guide. Boston, Massachusetts: The Health Institute, New England Medical Center, 1993.

187. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res. 1976; 11: 478-507.

188. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res. 1998; 7: 323-35.

189. Paltzer J, Barker E, Witt WP. Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures. Qual Life Res. 2013; 22: 1177-87.

190. Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. Soc Sci Med. 2003; 57: 1289-306.

191. Herdman M, Fox-Rushby J, Badia X. 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. Quality of Life Research. 1997; 6: 237-47.

192. WHO G, Switzerland. Millenium Development Goals and Beyond 2015.

193. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. 2015; 68: 435-41.

194. Fox-Rushby J. Operationalising conceptions of 'Health' Amongst the Wakamba and Maragoli of Kenya: the basis of the KENQOL instrument. Qual Life Res. 2000; 9: 316e16.

195. Colbourn T, Masache G, Skordis-Worrall J. Development, reliability and validity of the Chichewa WHOQOL-BREF in adults in Lilongwe, Malawi. BMC Res Notes. 2012; 5: 346.

196. Mabugu T, Revill P, van den Berg B. The Methodological Challenges for the Estimation of Quality of Life in Children for Use in Economic Evaluation in Low-Income Countries. Value Health Reg Issues. 2013; 2: 231-39.

197. Connolly MA, Johnson JA. Measuring quality of life in paediatric patients. Pharmacoeconomics. 1999; 16: 605-25.

198. 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.

199. Ngwira LG, Khan K, Maheswaran H, et al. A Systematic Literature Review of Preference-Based Health-Related Quality-of-Life Measures Applied and Validated for Use in Childhood and Adolescent Populations in Sub-Saharan Africa. Value Health Reg Issues. 2021; 25: 37–47.

200. Ratcliffe J, Stevens K, Flynn T, et al. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. Qual Life Res. 2012; 21: 717-25.

201. Shiroiwa T, Ikeda S, Noto S, et al. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. Medical Decision Making. 2021.

202. Jelsma J, Hansen K, de Weerdt W, et al. How do Zimbabweans value health states? Population Health Metrics. 2003; 1: 11.

203. Schor E. Children's health and the assessment of health-related quality of life. In: Drotar D, editor. Measuring health-related quality of life in children and adolescents. Mahwah (NJ): Lawrence Hillbaum. 1998: 25-39.

204. Scalone L, Tomasetto C, Matteucci M, et al. Assessing quality of life in children and adolescents: development and validation of the Italian version of the EQ-5D-Y. Italian Journal of Public Health. 2011; 8: 331-41.

205. Åström M. Measurement and valuation of health among children and adolescents using the EQ-5D-Y– Methodological considerations and implications. THE DEPARTMENT OF LEARNING, INFORMATICS, MANAGEMENT AND ETHICS. Stockholm, Sweden: Karolinska Institutet, 2021.

206. Verstraete J, Marthinus Z, Dix-Peek S, et al. Measurement properties and responsiveness of the EQ-5D-Y-5L compared to the EQ-5D-Y-3L in children and adolescents receiving acute orthopaedic care. Health Qual Life Outcomes. 2022; 20: 28.

207. Verstraete J, Amien R, Scott D. Comparing Measurement Properties of the English EQ-5D-Y Three-Level Version with the Five-Level Version in South Africa. Preprints (wwwpreprintsorg) 2022.

208. Fitriana TS, Purba FD, Rahmatika R, et al. Comparing measurement properties of EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients. Health and Quality of Life Outcomes. 2021; 19: 256.

209. The University of Melbourne. Professor Kim Dalziel was awarded a 2021 NHMRC investigator grant, 2020. Available: https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-policy/research-group/health-economics/news-and-

events/associate-professor-kim-dalziel-was-awarded-a-2021-nhmrc-investigator-grant. Accessed, 15 April 2022.

210. Shah KK, Bennett B, Lenny A, et al. Adapting preference-based utility measures to capture the impact of cancer treatment-related symptoms. The European Journal of Health Economics. 2021.

211. Berg B. Qualitative Research Methods for the Social Sciences. 6th ed. Boston: Pearson Education, 2007.

212. Green J, Thorogood N. Qualitative methods for health research. SAGE Publications, 2004.

213. Kielmann K, Cataldo F, Seeley J. Introduction to Qualitative Research Methodology: A Training Manual. DFID, UK, 2012.

214. Krueger RA. Focus Groups: A Practical Guide for Applied Research 3rd ed. Thousand Oak, California: Sage, 2000.

215. Heary C-M, Hennessy E. The use of focus group interviews in pediatric health care research. Journal of Pediatric Psychology. 2002; 27: 47-57.

216. Hoban E. Creative Methodologies to Stimulate Children's Participation during Focus Group Discussions in Rural Cambodia. Journal of Healthcare Communications. 2017; 2 No. S1:71.

217. Collins D. Pretesting survey instruments: An overview of cognitive methods. Quality of Life Research. 2003; 12: 229-38.

218. Walsh TR, Irwin DE, Meier A, et al. The use of focus groups in the development of the PROMIS pediatric item bank. Qual Life Res. 2008; 17: 725-35.

219. Heary CM, Hennessy E. Focus groups versus individual interviews with children : A comparison of data. The Irish Journal of Psychology. 2006; 27: 58-68.

220. Vaughn S, Schumm JS, Sinagub J. Focus group interviews in education and psychology. London: Sage, 1996.

221. Michell L. Telling it how it is; telling it how it feels. In: Barbour RS, Kitzinger J, eds., Developing focus group research (pp 36-46) London: Sage. London: Sage, 1999.

222. Kitzinger J. The methodology of focus groups: The importance of interaction between research participants. . Sociology of Health & Illness. 1994; 16: 103-21.

223. Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educational Researcher. 2004; 33: 14-26.

224. Lowhorn G. Qualitative and Quantitative Research How to Choose the Best Design Academic Business World International Conference. Nashville, Tennessee: Regent University, 2007.

225. Onwuegbuzie AJ, Dickinson WB, Leech NL, et al. A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. International Journal of Qualitative Methods. 2009; 8.

226. Morgan DL. Paradigms Lost and Pragmatism Regained: Methological Implications of Combining Qualtitative and Quantitative Methods. Journal of Mixed Methods Research. 2016; 1: 48-76.

227. Onwuegbuzie AJ, Leech NL. On becoming a pragmatic researcher: the importance of combining quntitative and qualitative methodologies International Journal of Social Research Methodology. 2005; 8: 375-87.

228. Howe KR. Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard. Educational Researcher. 1988; 17.

229. Sieber SD. The integration of fieldwork and survey methods. American Journal of Sociology. 1973; 73: 1335-59.

230. Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw-Hill, 1994.

231. Smith SC, Lamping DL, Banerjee S, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. Health Technol Assess, 2005.

232. Mislevy RJ, Wilson MR, Ercikan K, et al. Psychometric Principles in Student Assessment. International Handbook of Educational Evaluation, 2003.

233. Messick S. The interplay of evidence and consequences in the validation of performance assessments. Education Researcher. 1994; 32: 13-23.

234. Petrillo J, Cano SJ, McLeod LD, et al. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value Health. 2015; 18: 25-34.

235. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technology Assessment. 2009; 13.

236. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. . Educ Measurement Issues and Practice. 1993: 38–47.

237. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use 5th ed. Oxford: Oxford University Press, 2015.

238. Lord FM. A theory of test scores. Psychometric monographs. 1952; 7.

239. Tractenberg RE. Classical and modern measurement theories, patient reports, and clinical outcomes. Contemp Clin Trials. 2010; 31: 1-3.

240. US FDA. U.S. Food and Drug Administration: What We Do. Available from https://www.fda.gov/about-fda/what-we-do. last accessed 26th March 2021.

241. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status

measurement instruments: an international Delphi study. Qual Life Res. 2010; 19: 539-49.

242. Neill SJ. Research with children: a critical review of the guidelines. J Child Health Care. 2005; 9: 46-58.

243. Darbyshire P, MacDougall C, Schiller W. Multiple methods in qualitative research with children: more insight or just more? Qualitative Research. 2016; 5: 417-36.

244. Petrou S. Methodological challenges surrounding QALY estimation for paediatric economic evaluation. Cost Eff Resour Alloc. 2022; 20: 10.

245. Eiser C, Morse R. Can parents rate their child's health-related qulaity of life? Results of a systematic review. Quality of Life Research. 2001; 10: 347-57.

246. National Research Council and Institute of Medicine. Children's Health, the Nation's Wealth: Assessing and Improving Child Health Washington, DC: The National Academies Press, 2004.

247. United Nations General Assembly. Convention on the Rights of the Child. United Nations General Assembly, Treaty Series, vol 1577. Available from: https://downloads.unicef.org.uk/wpcontent/uploads/2010/05/UNCRC_.

248. Matza LS, Patrick DL, Riley AW, et al. Pediatric Patient-Reported Outcome Instruments for Research to Support Medical Product Labeling: Report of the ISPOR PRO Good Research Practices for the Assessment of Children and Adolescents Task Force. Value in Health. 2013; 16: 461-79.

249. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA: The Journal of the American Medical Association. 1995; 273: 59-65.

250. World Health Organization. International Classification of Functioning, Disability and Health: Children and Youth Version: ICF-CY. Switzerland: WHO Press: World Health Organization, 2007.

251. WHO G, Switzerland. Sustainable development goals. http://wwwunorg/sustainabledevelopment/sustainable-development-goals/. last assessed 15th September 2017.

252. WHOQOL Group. The World Health Organisation quality of life assessment (WHOQOL): development and general psychometric properties. . Soc Sci Med 1998; 46: 1569–85.

253. Group TW. The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties. Soc Sci Med. 1998; 46: 1569-85.

254. Campbell D, Fiske D. Convergent and discriminant validiation by the multitrait-multimed matrix. Psychol Bulletin. 1959; 56: 81-105.

255. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science. 2014; 43: 115-35.

256. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011; 20: 1727-36.

257. IBM Corp. IBM SPSS statistics for mac. Version 26.0. In: Corp I, ed. Armonk, NY, 2018.

258. 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: 315-33.

259. Kreimeier S, Astrom M, Burstrom K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res. 2019; 28: 1951-61.

260. Wong CKH, Cheung PWH, Luo N, et al. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. Eur J Health Econ. 2019; 20: 647-56.

261. Kreimeier S, Greiner W. EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. Value Health. 2019; 22: 31-37.

262. Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. Med Care. 2009; 47: 634-41.

263. The EuroQol Group. Full translation guidelines for the EuroQol EQ-5D-3L,5L and Y versions. Rotterdam, Netherlands: EuroQol Research Foundation, 2018.

264. Chokotho L, Mkandawire N, Conway D, et al. Validation and reliability of the Chichewa translation of the EQ-5D quality of life questionnaire in adults with orthopaedic injuries in Malawi. Malawi Medical Journal. 2017; 29: 84-88.

265. Derrett S, Herdman M, Ngwira LG, et al. A New Approach to Assessing Children's Interpretation of Severity Qualifiers in a Multi-Attribute Utility Instrument–The EQ-5D-Y-5L: Development and Testing. The Patient - Patient-Centered Outcomes Research. 2021; 14: 591-600.

266. Trust MR, Varni JW. Scaling and Scoring of the Pediatric Quality of Life Inventory ™ PedsQL™. In: Trust MR, ed., 2017.

267. Varni JW, Burwinkle TM, Seid M, et al. The PedsQLy 4.0 as a Pediatric Population Health Measure:Feasibility, Reliability, and Validity. Ambulatory Pediatrics. 2003; 3: 329-41.

268. The EuroQol Group. https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/self-complete-on-paper/. Last assessed 22nd September 2021.

269. Chavula C, Suleman H. Assessing the Impact of Vocabulary Similarity on Multilingual Information Retrieval for Bantu Languages. the 8th annual meeting of the Forum, 2016.

270. Lee GL, Tan RL, Herdman M, et al. Assessing the Content Validity of the EQ-5D Questionnaire Among Asians in Singapore: A Qualitative Study. Annals Academy of Medicine Singapore. 2020; 49: 294-305.

271. McDonald R, Mullett TL, Tsuchiya A. Understanding the composite dimensions of the EQ-5D: An experimental approach. Soc Sci Med. 2020; 265: 113323.

272. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Appl Health Econ Health Policy. 2017; 15: 127-37.

273. Youkee D. EQ-5D-3L equivalence in Sierra Leone. 1st EuroQol African Regional Meeting. Cape Town, South Africa, 13-14 February 2020.

274. Jelsma J, Chivaura V, de Weerdt W, et al. A bridge between cultures: A report on the process of translating the EQ-5D instrument into Shona. South African Journal of Physiotherapy. 2000; 56: 3-9.

275. Craig BM, Monteiro AL, Herdman M, et al. Further evidence on EQ-5D-5L preference inversion: a Brazil/U.S. collaboration. Qual Life Res. 2017; 26: 2489-96.

276. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D data Netherlands: Springer, 2020.

277. Feng Y, Parkin D, Devlin NJ. Assessing the performance of the EQ-VAS in the NHS PROMs programme. Qual Life Res. 2014; 23: 977-89.

278. Miller GA. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review. 1956; 63: 81-97.

279. Astrom M, Krig S, Ryding S, et al. EQ-5D-Y-5L as a patient-reported outcome measure in psychiatric inpatient care for children and adolescents - a cross-sectional study. Health Qual Life Outcomes. 2020; 18: 164.

280. Wong CKH, Cheung PWH, Luo N, et al. Responsiveness of the EQ-5D youth version 5-level (EQ-5D-5L-Y) and 3-level (EQ-5D-3L-Y) in patients with idiopathic scoliosis. Spine. 2019; 44: 1507-14.

281. Ngwira LG, Jelsma J, Smith SC, et al. Development of the Chichewa (Malawi) EQ-5D-Y and EQ-5D-Y-5L using ranking exercises. 1st EuroQol African Regional Meeting. Cape Town, South Africa, 13-14 February 2020.

282. Ngwira LG, Jelsma J, Smith SC, et al. Development of the new EQ-5D-Y-5L using the ranking order exercises: : preliminary results. 1st EuroQol Early Career Researchers. Prague, Czech Republic, 1 March 2020.

283. Mkoka S, Vaughan J, Wylie T, et al. The pitfalls of translation--a case study based on the translation of the EQ-5D into Xhosa. S Afr Med J. 2003; 93: 265-6.

284. Ngwira LG, Jelsma J, Maheswaran H, et al. Cross-Cultural Adaptation of the Beta EQ-5D-Y-5L Into Chichewa (Malawi). Value Health Reg Issues. 2022; 29: 36-44.

285. 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: 800-12.

286. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. Expert Rev Pharmacoecon Outcomes Res. 2005; 5: 705-19.

287. Varni JW. Pediatric health-related quality of life measurement technology: a guide for health care decision makers. J Clin Outcomes Manag. 1999; 6: 33-40.

288. Kaplan RM, Bush JW. Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis. Health Psychology. 1982; 1: 61-80.

289. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg. 2018; 126: 1763-68.

290. Upton P, Eiser C, Cheung I, et al. Measurement properties of the UK-English version of the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. Health and Quality of Life Outcomes. 2005; 3.

291. Weng L-J. Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. Educational and Psychological Measurement. 2016; 64: 956-72.

292. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. Soc Sci Med. 2016; 165: 66-74.

293. Crump RT, Beverung LM, Lau R, et al. Reliability, Validity, and Feasibility of Direct Elicitation of Children's Preferences for Health States. Med Decis Making. 2017; 37: 314-26.

294. Ratcliffe J, Stevens K, Flynn T, et al. Whose values in health? An empirical comparison of the application of adolescent and adult values for the CHU-9D and AQOL-6D in the Australian adolescent general population. Value Health. 2012; 15: 730-6.

295. Prevolnik Rupel V, Ramos-Goni JM, Ogorevc M, et al. Comparison of Adult and Adolescent Preferences Toward EQ-5D-Y-3L Health States. Value Health. 2021; 24: 1350-59.

296. Choi Y, Kim JH, Park EC. The effect of subjective and objective social class on health-related quality of life: new paradigm using longitudinal analysis. Health Qual Life Outcomes. 2015; 13: 121.

297. Cox MJ, Paley B. Families as systems. Annu Rev Psychol Assess. 1997; 48: 243–67.

298. Janssen M, Birnie E, Haagsma JA, et al. Comparing the Standard EQ-5D Three-Level System with a Five-Level Version. Value Health. 2008; 11: 275-84.

299. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. Pharmacoeconomics. 2018; 36: 675-97.

300. Janssen MF, Szende A, Cabases J, et al. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. Eur J Health Econ. 2019; 20: 205-16.

301. Burstrom K, Bartonek A, Brostrom E, et al. EQ-5D-Y as a health-related quality of life measure in children and adolescents with functional disability in Sweden: testing feasibility and validity. Acta Paediatrica. 2014; 103: 426-35.

302. Shiroiwa T, Fukuda T, Shimozuma K. Psychometric properties of the Japanese version of the EQ-5D-Y by self-repot and proxy-version: reliability and construct validity. Quality of Life Research. 2019; Epub ahead of print.

303. Scalone L, Ciampichini R, Fagiuoli S, et al. Comparing the performance of the standard EQ-5D 3L with the new version EQ-5D 5L in patients with chronic hepatic diseases. Qual Life Res. 2013; 22: 1707-16.

304. Noyes J, Edwards RT. EQ-5D for the assessment of health-related quality of life and resource allocation in children: a systematic methodological review. Value Health. 2011; 14: 1117-29.

305. Zhou W, Shen A, Yang Z, et al. Patient-caregiver agreement and test-retest reliability of the EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients with haematological malignancies. Eur J Health Econ. 2021.

306. The EuroQol Group. EQ-5D-Y User Guide Basic information on how to use the EQ-5D-Y instrument. 2014.

307. Ravens-Sieberer U, Erhart M, Wille N, et al. Generic Health-Related Quality-of-Life Assessment in Children and Adolescents: Methodological Considerations. Pharmacoeconomics. 2006; 24: 1199-220.

308. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. Psychological medicine 1988; 18: 1007-19.

309. Bowling A. Just one question: If one question works, why ask several? J Epidemiol Community Health. 2005; 59: 342-45.

310. Parkin D, Rice N, Devlin N. Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? Med Decis Making. 2010; 30: 556-65.

311. The EuroQol Group. https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/choosing-a-value-set/. Last assessed 15th September 2021.

312. Shaw JW, Johnson JA, Coon SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 2005; 43: 203-20.

313. Pickard AS, Law EH, Jiang R, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. Value in Health. 2019; 22: 931-41.

314. Petrou S, Morrell J, Spiby H. Assessing the empirical validity of alternative multi-attribute utility measures in the maternity context. Health Qual Life Outcomes. 2009; 7: 40.

315. Buchholz I, Janssen MF, Kohlmann T, et al. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. PharmacoEconomics. 2018; 36: 645-61.

316. Pan CW, Zhong H, Li J, et al. Measuring health-related quality of life in elementary and secondary school students using the Chinese version of the EQ-5D-Y in rural China. BMC Public Health. 2020; 20: 982.

317. Feng Y, Devlin N, Herdman M. Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? Health Qual Life Outcomes. 2015; 13: 171.

318. Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. Qual Life Res. 2015; 24: 1775-84.

319. Kim SH, Kim HJ, Lee SI, et al. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. Qual Life Res. 2012; 21: 1065-73.

320. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013; 22: 1717-27.

321. Ferreira LN, Ferreira PL, Ribeiro FP, et al. Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. Health Qual Life Outcomes. 2016; 14: 89.

322. Rencz F, Lakatos PL, Gulacsi L, et al. Validity of the EQ-5D-5L and EQ-5D-3L in patients with Crohn's disease. Qual Life Res. 2019; 28: 141-52.

323. Wang P, Luo N, Tai ES, et al. The EQ-5D-5L is More Discriminative Than the EQ-5D-3L in Patients with Diabetes in Singapore. Value Health Reg Issues. 2016; 9: 57-62.

324. Pickard AS, De Leon MC, Kohlmann T, et al. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care. 2007; 45: 259-63.

325. Pan CW, Sun HP, Wang X, et al. The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Qual Life Res. 2015; 24: 1767-74.

326. Ngwira LG, Maheswaran H, Petrou S, et al. Head to head performance of the Chichewa language (Malawi) versions fo the EQ-5D-Y and EQ-5D-Y-5L. 37th EuroQol Plenary Meeting. Virtual 16-19 September 2020.

327. Verstraete J, Ngwira LG, Scott D, et al. Comparing the performance of the EQ-5D-Y-3L and EQ-5D-Y-5L across age groups draft manuscript. 2021.

328. Keeley T, Al-Janabi H, Lorgelly P, et al. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. PLoS One. 2013; 8: e85287.

329. van Leeuwen KM, Jansen AP, Muntinga ME, et al. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. BMC Health Serv Res. 2015; 15: 201.

330. Shah KK, Mulhern B, Longworth L, et al. Views of the UK General Public on Important Aspects of Health Not Captured by EQ-5D. Patient. 2017; 10: 701-09.

331. Devlin NJ, Hansen P, Selai C. Understanding health state valuations: a qualitative analysis of respondents' comments. Qual Life Res. 2004; 13: 1265-77.

332. Ludwig K, Surmann B, Racker E, et al. Developing and testing a cognitive bolt-on for the EQ-5D-Y (Youth). Qual Life Res. 2021.

333. Torrance GW. Utility approach to measuring health-related quality of life. J Chronic Dis. 1987; 40: 593-603.

334. Ware Jr. JE. The status of health assessment 1994. Ann Rev Public Halth. 1995; 16: 327-54.

335. Patrick DL, Bergner M. Measurement of Health Status in the 1990s. Ann Rev Public Halth. 1990; 11: 165-83.

336. Fanshel S, Bush J. A Health Status Index and its Application to Health Service Outcomes. Oper Res. 1970; 18: 1021-66.

337. Rosser RM. Recent Studies Using a Global Approach to Measuring Illness. Med Care. 1976; XIV: 138.

338. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J R Coll Gen Pract. 1985; 35: 185-8.

339. Bergner M, Bobbitt RA, Kressel S, et al. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. Int J Health Serv. 1976; 6: 393-415.

340. Kind P, Luo N. EQ-5D: What does it measure - and how would we (do we) know? , 37th EuroQol Plenary Meeting, September 2020. Virtual Conference, 2020.

341. Brazier J, Ratcliffe J, Salomon JA, et al. Measuring and Valuing Helth Benefits for Economic Evaluation. 2nd ed. Oxford, UK: Oxford University Press, 2017.

342. Bergner M, Rothman M. Health Status Measures: An Overview and Guide for Selection. Ann Rev Public Halth. 1987; 8: 191-210.

343. Drotar D. Measuring Health-Related Quality of Life in Children and Adolescents: Implications fo Research and Practice. New York: Psychology Press, Taylor and Francis Group, 2009.

344. Gilson BS, Gilson JS, Bergner M, et al. The sickness impact profile. Development of an outcome measure of health care. Am J Public Health. 1975; 65: 1304-10.

345. McDowell I, Jenkinson C. Development standards for health measures. J Health Serv Res Policy. 1996; 1: 238-46.

346. Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess. 2004; 16: 231-43.

347. Ware JE, Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field. Arch Phys Med Rehabil. 2003; 84: S43-51.

348. Keenaghan C, Kilroe J. A Study on the Quality of Life Tool KIDSCREEN for Children and Adolescents in Ireland: Results of the KIDSCREEN National Survey 2005. Health Service Executive and The KIDSCREEN Group Europe. Dublin: Office of the Minister for Children and Young People/Department of Health and Children, 2008.

349. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990; 16: 199-208.

350. Gudex C. The descriptive system of the EuroQol instrument. In: Kind P, Brooks R, Rabin R, eds., EQ-5D concepts and methods: a developmental history. Dodrecht: Springer, 2005.

351. Kind P, Gudex C. Meauring health status in the community: a comparison of methods. Journal of Epidemiology and Community Health. 1994; 48: 86-91.

352. Gudex C. The development of the EuroQol descriptive system: a review. EuroQol Plenary Meeting. Barcelona, 1995.

353. The EuroQol Group. https://euroqol.org/euroqol/. last accessed 25th February 2021.

354. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993; 46: 1417-32.

355. Gladstone M, Lancaster G, Umar E, et al. Perspectives of normal child development in rural Malawi - a qualitative analysis to create a more culturally appropriate

developmental assessment tool. Child: Care, Health and Development. 2010; 36: 346-53.

356. Guest G, Namey E, McKenna K. How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. Field Methods. 2016; 29: 3-22.

357. McLeroy KR, Bibeau D, Steckler A, et al. An Ecological Perspective on Health Promotion Programs. Health Education Quarterly. 1988; 15: 351-77.

358. Chikovore J, Hart G, Kumwenda M, et al. TB and HIV stigma compounded by threatened masculinity: implications for TB health-care seeking in Malawi. The International Journal of Tuberculosis and Lung Disease. 2017; 21: 26-33.

359. Kaler A, Watkins S. Asking God about the date you will die: HIV testing as a zone of uncertainty in rural Malawi. Demogr Res. 2010; 23: 905-32.

360. Kagawa-Singer M, Padilla GV, Ashing-Giwa K. Health-related quality of life and culture. Semin Oncol Nurs. 2010; 26: 59-67.

361. Ferrans CE. Development of a quality of life index for patients with cancer. Oncol Nurs Forum. 1990: 15-21.

362. Ashing-Giwa KT, Kagawa-Singer M, Padilla GV, et al. The impact of cervical cancer and dysplasia: a qualitative, multiethnic study. Psychooncology. 2004; 13: 709-28.

363. Masulani-Mwale C, Mathanga D, Silungwe D, et al. Parenting children with intellectual disabilities in Malawi: the impact that reaches beyond coping? Child Care Health Dev. 2016; 42: 871-80.

364. Kashem T, Al Sayah F, Tawiah A, et al. The relationship between individual-level deprivation and health-related quality of life. Health and Quality of Life Outcomes. 2019; 17.

365. Hunt SM. Cross-cultural issues in the use of socio-medical indicators. . Health Policy. 1986; 6: 149–58.

366. Williams A. The measurement and Valuation of Health: A Chronicle. Discussion Paper 136. Centre for Health Economics: University of York, 1995.

367. Westmoreland K, Reeve BB, Amuquandoh A, et al. Translation, psychometric validation, and baseline results of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric measures to assess health-related quality of life of patients with pediatric lymphoma in Malawi. Pediatr Blood Cancer. 2018: e27353.

368. Pérez-Sousa MÁ, Olivares PR, Ramírez-Vélez R, et al. Comparison of the Psychometric Properties of the EQ-5D-3L-Y and EQ-5D-5L-Y Instruments in Spanish Children and Adolescents. Value in Health. 2021.

369. Amien R, Scott D, Verstraete J. Performance of the EQ-5D-Y Interviewer Administered Version in Young Children. Children (Basel). 2022; 9.

370. Doeleman MJH, de Roock S, Buijsse N, et al. Monitoring patients with juvenile idiopathic arthritis using health-related quality of life. Pediatr Rheumatol Online J. 2021; 19: 40.

371. Pei W, Yue S, Zhi-Hao Y, et al. Testing measurement properties of two EQ-5D youth versions and KIDSCREEN-10 in China. Eur J Health Econ. 2021; 22: 1083-93.

372. Malik M, Gu NY, Hussain A, et al. A Pilot Study of Valuation Methods of the EQ-5D and the Impact of Literacy, Cultural and Religious Factors on Preferences. Value Health Reg Issues. 2022; 30: 48-58.

373. Purba FD, Hunfeld JAM, Iskandarsyah A, et al. The Indonesian EQ-5D-5L Value Set. Pharmacoeconomics. 2017; 35: 1153-65.

374. Chen LM, Miaskowski C, Dodd M, et al. Concepts within the Chinese culture that influence the cancer pain experience. Cancer Nurs. 2008; 31: 103-8.

375. Yang F, Katumba KR, Roudijk B, et al. Developing the EQ-5D-5L Value Set for Uganda Using the 'Lite' Protocol. Pharmacoeconomics. 2021.

376. Geraerds AJLM, Bonsel GJ, Janssen MF, et al. Methods Used to Identify, Test, and Assess Impact on Preferences of Bolt-Ons: A Systematic Review. Value in Health. 2021; 24: 901-16.

377. Brazier J, Rowen D, Tsuchiya A, et al. The impact of adding an extra dimension to a preference-based measure. Soc Sci Med. 2011; 73: 245-53.

378. Haywood P, Addo R, Reardon O, et al. Generating criteria for the development and selection of candidate bolt-on dimensions. Paper presented at: EuroQol Early Career Researcher Meeting March 2020; . Prague, Czech Republic.

379. Peasgood T, Mukuria C, Carlton J, et al. Criteria for item selection for a preference-based measure for use in economic evaluation. Quality of Life Research. 2020.

380. Brazier J. https://www.sheffield.ac.uk/media/20045/download. ScHARR, University of Sheffield, Last accessed 2nd August 2021.

381. Schubbe D, Scalia P, Yen RW, et al. Using pictures to convey health information: A systematic review and meta-analysis of the effects on patient and consumer health behaviors and outcomes. Patient Educ Couns. 2020; 103: 1935-60.

382. Gladstone M, Lancaster GA, Umar E, et al. The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings. PLoS Med. 2010; 7: e1000273.

383. Bandstra NF, Chambers CT. Pain Assessment in Children. In: Breivik H, Campbell WI, Nichols M, eds., Clinical Pain Management: Practice and Procedures. 2nd ed. Florida: CRC Press, Taylor & Francis Group, 2008.

384. Yang F, Jiang S, He X, et al. Do Rural Residents in China Understand EQ-5D-5L as Intended? Evidence From a Qualitative Study. PharmacoEconomics - Open. 2020.

385. Prevolnik Rupel V, Ogorevc M. EQ-5D-Y Value Set for Slovenia. PharmacoEconomics. 2021.

386. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, et al. Accounting for Unobservable Preference Heterogeneity and Evaluating Alternative Anchoring Approaches to Estimate Country-Speci!c EQ-5D-Y Value Sets: A Case Study Using Spanish Preference Data. Value Health. 2021; In press.

387. Ramos-Goni JM, Oppe M, Stolk E, et al. International Valuation Protocol for the EQ-5D-Y-3L. Pharmacoeconomics. 2020.

388. Lipman SA, Reckers-Droog VT, Kreimeier S. Think of the Children: A Discussion of the Rationale for and Implications of the Perspective Used for EQ-5D-Y Health State Valuation. Value in Health. 2021.

389. Reckers-Droog V, Karimi M, Lipman S, et al. Why Do Adults Value EQ-5D-Y-3L Health States Differently for Themselves Than for Children and Adolescents: A Think-Aloud Study. Value Health. 2022.

390. Rowen D, Mukuria C, Powell P, et al. Valuing child health: exploring the use of a mixed sample of adolescents and adults to value child and adolescent health states. NICE DSU Report, 2021.

391. Kreimeier S, Oppe M, Ramos-Goni 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: 1291-98.

392. Kind P, Klose K, Gusi N, et al. 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: 2519-39.

393. Chemli J, Drira C, Felfel H, et al. Valuing health-related quality of life using a hybrid approach: Tunisian value set for the EQ-5D-3L. Qual Life Res. 2021.

394. Haagsma JA, Derrett S. Population Quality of Life (POPQOL) African study. 2nd African EuroQol Conference 9-10 February 2022. Cairo (virtual), 2022.

395. APERSU. https://apersu.ca/wp-content/uploads/2021/10/APERSU-Annual-Report-2021.pdf. last accessed 25th March 2022.

396. Jelsma J, Shumba D, Hansen K, et al. Preferences of urban Zimbabweans for health and life lived at different ages. Bulletin of the World Health Organization. 2002; 80: 204-09.

397. Mott DJ, Shah KK, Ramos-Goni JM, et al. Valuing EQ-5D-Y-3L Health States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ? Med Decis Making. 2021: 272989X21999607.

398. Dalziel K, Catchpool M, Garcia-Lorenzo B, et al. 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; 38: 499-513.

399. 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.

400. Lipstein EA, Brinkman WB, Fiks AG, et al. An emerging field of research: challenges in pediatric decision making. Med Decis Making. 2015; 35: 403-8.

401. Grano N, Kieseppa T, Karjalainen M, et al. Exploratory factor analysis of a 16D Health-Related Quality of Life instrument with adolescents seeking help for early psychiatric symptoms. Nord J Psychiatry. 2016; 70: 81-7.

402. Furber G, Segal L. The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. Health Qual Life Outcomes. 2015; 13: 22.

403. Chen G, Flynn T, Stevens K, et al. Assessing the Health-Related Quality of Life of Australian Adolescents: An Empirical Comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. Value Health. 2015; 18: 432-8.

10 Appendices

10.1 Appendix 1-Detailed generic preference-based health-related quality of life

Table 10.1 Detailed psychometric properties as evaluated by developers of the identified generic childhood preference-based health-related quality of life

Measure *	Conceptual framework[a,b,c,d]	Reliability (dependability/ consistency of an instrument)[a,b,c,d]			Validity* (accuracy/ represent exactly what is on the ground)[a,b,c,d]					Interpretation of scores[a,b,c,d]	Practicality/ Burden [a,b,c,d]	Cross-cultural & language adaptation[a,b,c,d]
		Internal consistency [a,b,c,d]	test re-test (reproducibility) [a,b,c,d]	inter-rater [a,b, d]	Content [a,b,c,d]	Construct [a,b,c,d]			Responsiveness [a,b,c,d]			
						Convergent[a]	Discriminant[a]	Known group[a]				
HUI2 & HUI3*		Low correlation among attributes compared to other measures (173)	ICC≥0.70 (173)					MID ≥0.03 for clinically different groups (173)	Demonstrated responsiveness (173)	Few ceiling & floor effect in clinical studies but in health surveys(173)		
QWB-SA					*QWB has content validity similar to SF-36 (171)*	*QWB-SA and QWB were found to highly correlated (169)*					*Found to be practical (171)*	

16D		Repeatability coefficient by individual level dimensions ranged from 86-100%, and majority (11 out of 16 dimensions) had repeatability coefficient ≥*97% (only sleeping, and discomfort and symptoms* <94%). *(167);* Cronbach's α= 0.831(401)	Good. Overall reliability coefficient r=0.91 (lying 2 standard deviation (SD) from mean difference) and between 0.86-1.00 by dimension being over and above the accepted standards *(167)*		Developed by a team of children and adolescent experts and domains were mostly modified from an existing 15D instrument (167)			Able to differentiate between sick and healthy children HRQoL states- *(167)*			Easy to complete taking between five to ten minutes *(167) quite quick to complete among psychiatric symptoms (401)*	
17D	A clear methodology of questionnaire developmen	11 out of 17 dimensions had reliability coefficient	Reliability coefficient overall score, r = 0.95 with cases lying	No evidence of inter-rater	Developed from 16D by a team of experts in paediatrics			Children with more health problems not only			Children did not just complete filling the form but	

	t. Questionnaire revised with input from children (168)	r =0.97 with only two (sleeping and breathing) r<0.93. (168)	within 2 SD of the mean difference and by dimensions r= 0.91-1.0 (168)	correlation between childrenand proxies especially on speech (168)	and adolescent (168)			reported more problems but also scored lower (0.895+/-0.072) than those without health problem (0.937+/-0.053) on particular dimensions. (168)			also found it enjoyable. (168)	
AQoL-6D	Four broad steps involving interviews with target population. *(176)*				Five of the six dimensions had a correlation coefficient of ≥0.73 (lambda loading gamma coefficients are equivalent of correlation coefficient) except for	Five of the six dimensions had a correlation coefficient of ≥0.73 (lambda loading gamma coefficients are equivalent of correlation coefficient) except for						Demonstrated good cross-cultural validity when adapted for adolescents in four pacific countries (140, 176)

					vision (=0.51) *(166)*	vision (=0.51) *(166)*						
CHU-9D	*Excellent. Qualitative work, cognitive interviews then constructed descriptive system (107, 157)*	Cronbach's α = 0.781*(165)*	***NR***	***NR***	Excellent Only child preference-based instrument with domains constructed directly from children (7-11yrs) using qualitative research *(157, 164)* Good content validity as only 11% of children felt something missing in content (165)	Child-rated level of health correlated with descriptive system demonstrating construct validity *(165)*	***NR***	CHU9D able to discriminate between clinically known different groups based on indicators of long-standing illness and disability and general health as self-reported (153, 163, 200, 402, 403)	***NR***	***NR***	Response rate, completion rate for each question and mean completion time that ranged from 64% - 98%, 98.7%-100% and 3.8 - 5.3 minutes respectively, many finding it quick and easy to administer.(165)	***NR***
EQ-5D-Y	The modification process included revision of domain definitions,	***NR=0***	For dimensions, agreement ranged from 69.8 to 99.7% in the EQ-5D-	*Good. Self-report and proxy version*	*Poor, as it was developed from adult EQ-5D version, domains not*	Kappa coefficients were up to 67%, but no correlation with other self-rated	***NR=0***	Detect differences between known clinically ill	***NR=0***	*Between 91 to 100% provided valid scores (160)*	*Found to be feasible in a number of countries with*	There is also evidence of cross-cultural validity from

	revision of wording and layout, translation, cognitive interviews, integration and decision-making on harmonization, and comparing the two versions. *Cognitive interviews involved target population (160)*		Y; Kappa coefficients up to 0.67 (160)	*agreement ranged from 72.9 % to 97.1 % (160)*	*generated specifically for children. There was also specifically lack of in-depth examination of domain conceptualisation (159)*	health measures, correlation coefficient (r=-0.56) *(160)*		groups *(160)*			*either nothing missing or very few inappropriate responses both for DS and VAS ranging from 0-2%, and time taken to complete <5 minutes (160)*	multiple trials showing EQ-5D-Y is well understood and practical across several countries *(160)*
AHUM					*Contents derived from EQ-5D and SF-6D (158)*							
CH-6D						Spearman correlation coefficient with HUI2 was 0.658 (156)		Able to differentiate between those with known different				

								status (156)				

*Criterion validity (concurrent and/or predicted) not reported by any study

[a] US Department of Health and Human Services Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: FDA; 2009;

[b] Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205;

[c] Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889-905;

[d] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.

*Psychometric evaluation for both measures was extracted from one article provided by the developer;

+ indicate meeting minimum review criteria; - indicate not meeting minimum review criteria; NR indicate not reported.

HUI2 Health Utility Mark Index 2; HUI3 Health Utility Mark Index 3; QWB-SA Quality of Well-Being Self-Administered; 16D 16Dimension; 17D 17Dimension; AQoL-6D Assessment of Quality-of-Life 6Dimension adolescent; CHU-9D Child Health Utility 9 Dimension; EQ-5D-Y EuroQol 5Dimension Youth; AHUM Assessment of Health Utility Measurement; CH-6D Child Health 6Dimension;

ICC Intra-class correlation coefficient; MID-meaningfully important difference; + indicate meeting minimum review criteria; - indicate not meeting minimum review criteria; NR indicate not reported

10.2 Appendix 2- Detailed cross-cultural validation of generic childhood preference-based HRQoL measures

Table 10.2 Detailed cross-cultural validation of generic childhood preference-based HRQoL measures in sSA (n=6)

Psychometric property/ Measure (author/ country)	Conceptual framework	Reliability* (dependability/ consistency of an instrument) in sSA		Validity (accuracy/ represent exactly what is on the ground) in sSA						Practicality/ Burden in sSA	Cross-cultural & language adaptation in sSA
		test re-test	inter-rater	Content	Criterion		Construct			Responsiveness		
					concurrent	predictive	Convergent	Discriminant	Known group			
HUI3 (Sims-Williams et al (23)/ Uganda)	+ Developed from HUI2	NR	- No correlation between children and proxies using VAS scores and HUI3 and VAS scores, Pearson correlation =0.133 and 0.380 respectively; +High correlation	NR	NR		- Not highly correlated (Pearson correlation=0.488) between HUI3 and VAS scores.		NR	+ Large treatment effect (Wilcoxon signed-rank test for VAS=0.43) observed in acutely ill children.	+*Feasible:* only 4 of 68 children unable to complete the HUI3 due to deafness and/or learning disability.	+ (adaptation involved translation of the HUI3); + (cognitive interviews with parents and children)

			(0.848) between children and proxies using HUI3							
EQ-5D-Y (Scott et al 2017 (24)/ South Africa)	+ Developed from adult EQ-5D version	+All dimensions (Kappa 0.365-0.653) except usual activities (kappa=0.199); + VAS scores ICC=0.77	NR	+ Nine clinical therapists found it to have right content	- No evidence of correlation between VAS and composite scores of different groups except the acutely ill	+ Only evident in acutely ill between EQ-5D-Y and PedsQL	+On dimension: Able to discriminate across groups with different health states (acutely ill v healthy children ($p<0.001$); -On composite scores: unable to discriminate +On VAS scores: able to discrimin	+ + able to detect medium effect size for both the chronic ill and acute ill children for both the composite and VAS scores	+ All children completed the EQ-5D-Y and was found to be quick and easy to apply by six of nine therapists who took part in the study	- (no translation was required as the English version was used); + (the original EQ-5D-Y established this during development which included South Africa)

							ate between groups (p<0.001)			
EQ-5D-Y (Jelsma & Ramma 2010 (25) / South Africa)	+ Developed from adult EQ-5D version	NR	- No agreement between Children and their proxies on all domains (Kappa= 0.01-0.60)	NR	NR	NR	-Children with disabilities didn't perceive their HRQoL worse compared to those without	NR	+ Of 62 respondents, there were only 5 missing responses	- (none required was in English); + (the original EQ-5D-Y established this during development)
EQ-5D-Y EQ5D (Jelsma 2010 (26)/South Africa)	+ Developed from adult EQ-5D version	NR	NR	NR	+ EQ-5D-Y performed better than EQ-5D	NR	NR	NR	+ + More missing responses on the EQ-5D compared to EQ-5D-Y both for dimension and VAS, Chi-squared=9.404; d.f.=1; (P=0.002and Chi-squared=26.159; d.f.=1; (P<0.001) respectively	- (none required was in English); + (the original EQ-5D-Y established this during development which included South Africa)

EQ-5D-Y (Wille et al 2010 (20)/ South Africa)	+ Developed from adult EQ-5D version	NR	NR	NR	NR	NR	NR	NR	+ + Children and adolescents satisfactoril y understood EQ-5D-Y than EQ-5D. Also, better accepted and proved more feasible than the EQ-5D	- (none required was in English); + (the original EQ-5D-Y established this during development which included South Africa)
EQ-5D-Y (Ravens-sieberer et al 2010 (19)/ South Africa)	+ Developed from adult EQ-5D version	NR	NR	NR	NR	+ There was evidence of convergent validity between the EQ-5D-Y and three other measures: KIDSCREEN-10 HRQoL Index, KIDSCREEN-27 and the Life Satisfaction Ladder.	+ There was evidence of known-group validity between groups with different health conditions	NR	+ Missing or inappropriat e responses on the EQ-5D-Y was 2%	- (none required was in English); + (the original EQ-5D-Y established this during development which included South Africa)

*Internal consistency not reported by any study

*Psychometric evaluation for both measures was extracted from one article provided by the developer;

+: indicate meeting minimum review criteria; -: indicate not meeting minimum review criteria; NR: indicate not reported;

HUI2 Health Utility Mark Index 2; HUI3 Health Utility Mark Index 3; QWB-SA Quality of Well-Being Self-Administered; 16D 16Dimension; 17D 17Dimension; AQoL-6D Assessment of Quality-of-Life 6Dimension adolescent; CHU-9D Child Health Utility 9 Dimension; EQ-5D-Y EuroQol 5Dimension Youth; AHUM Assessment of Health Utility Measurement; CH-6D Child Health 6Dimension;
ICC: Intra-class correlation coefficient; MID: meaningfully important difference

10.3 Appendix 3 CHU-9D translation request and translation quotation

Monday, February 22, 2021 at 11:34:28 AM South Africa Standard Time

Subject: RE: New translation - 2493-TR-0006
Date: Tuesday, 23 October 2018 at 16:58:32 South Africa Standard Time
From: Leary, Meg
To: Lucky-Gift Ngwira, Katherine J Stevens
CC: Clara Mukuria, Donna Rowen
Attachments: image003.gif, image004.gif, image005.gif, 23OCT2018_2493-TR-0006_University of Sheffield_CHU9D_Chichewa_v1.0.pdf

Hi Katherine and Lucky,

Thank you so very much for your patience as we conducted our research. We have concluded that the requested scope will be feasible for our team. Therefore please find attached a quote for this work.

I included final reports for additional quality documentation and for your benefit, but if costs are a concern I can certainly remove this – just let me know.

Please feel free to reach out with any questions.

Best,

Meg Leary
Senior Project Analyst, Language Services
Commercialisation & Outcomes

Tel: +1 919 294 2224
Email: Meg.Leary@iconplc.com
Web: www.iconplc.com

10.4 Appendix 4 EuroQol EQ-5D-Y and EQ-5D-Y-5L translation permission

Tuesday, October 16, 2018 at 10:29:56 AM Central Africa Time

Subject: RE: Malawi EQ-5D-Y translation - TRF2098
Date: Monday, 8 October 2018 at 16:15:03 Central Africa Time
From: Nalinie Banarsi
To: Lucky-Gift Ngwira
CC: Bianca Smit, Jennifer Jelsma
Attachments: image004.png, image005.png, image006.png, image007.png, 1a. EQ Guidelines for EQ-5D-3L_5L_Y Full translation_April 2018.pdf, 3. EQ Definition of Concepts 3L_5L_Y_May 2018.pdf, IN PROGRESS_Malawi (Nyanja - Chichewa) EQ-5D-Y Paper Self complete v1.0.0 (ID 71367).docx, Y version colour coded for trx purposes (ID 58620).docx, 2a. EQ Cultural Adaptation Procedures_Grid_April 2018.pdf

Dear Lucky,

Thank you for registering your study on our website. I confirm safe receipt.

As mentioned in previous emails, VMC has granted you permission to develop the **Malawi (Nyanja-Chichewa) EQ-5D-Y Paper Self-Complete** questionnaire. Your translation request number is: **TRF2098**. From now on, always refer to this number if you have any questions related to the translation project.

Before we will start with this project, I would like to recap and inform you about the following important items:

1. Our legal team will contact you shortly to set up your translation agreement. This agreement enables you to be the designated translator and it will regulate the new EQ-5D questionnaire. Please be aware that you will receive a separate email from the legal team. I would kindly like to ask you to keep both subjects (i.e. TRF2098 and translation agreement) separate. This is to avoid any miscommunication.
2. It is truly vital to make yourself acquainted with our translation guidelines. Therefore please read the attached *1a.* translation guideline really carefully. To support the *1a.* guideline, I have also attached the *3.* EQ-5D Definition of Concepts document, the Malawi (Nyanja-Chichewa) EQ-5D-Y Paper Self-Complete template and a colour coded Youth Self-Complete questionnaire.
3. We require the following translation methodology: dual forward and back translation, cognitive debriefing n~8 respondents and proofreading. For more information please see the attached *2a.* grid.
4. Only after your acknowledgement of understanding (above # 2 and 3), we will move forward to the next stage.

Please do not hesitate to contact me if you have any questions or concerns.

Kind regards,

Nalinie Banarsi
Legal Counsel / VMC Project Manager
EuroQol Research Foundation

EQ-5D

T +31 88 4400191 | E banarsi@euroqol.org | www.euroqol.org | Marten Meesweg 107 | 3068 AV Rotterdam | The Netherlands

Upcoming OOO: 24 October – 14 November 2018

10.5 Appendix 5 Mapi Trust approval for the translation of the PedsQL 4.0 child and teen self-report into Chichewa (Malawi)

Sunday, March 21, 2021 at 8:34:57 PM South Africa Standard Time

Subject: ePROVIDE™: Request #165063_PedsQL_Malawi
Date: Monday, 10 December 2018 at 15:47:51 South Africa Standard Time
From: Bellet, Alexandra
To: Lucky Ngwira
Attachments: image001.gif, image002.png

Dear Lucy,

Thank you for your interest in the PedsQL and for submitting your request in ePROVIDE. I am pleased to be your contact again for your request.

Please be kindly informed that the PedsQL is distributed by Mapi Research Trust. We do not have any Nyanja for Malawi versions available.
Could you please confirm you will like to translate the PedsQL ?

If so I will send you a User Agreement and a Translation Agreement. Upon reception of these completed documents, I will send you linguistic validation guidance, the scoring, and the original questionnaire in English for the USA.

Please kindly note that no fees apply as you are a non-funded academic user.
Should you have any questions please do not hesitate to contact me directly,

Best regards,
Alexandra Bellet
Customer Relation Specialist

alexandra.bellet@mapi-trust.org
www.mapi-trust.org, https://eprovide.mapi-trust.org/

Please, note my email address has changed to alexandra.bellet@mapi-trust.org Please update your directory with my new information.

Mapi Research Trust

From: eProvide
Sent: vendredi 7 décembre 2018 11:50
To: Lngwira@mlw.mw
Subject: ePROVIDE™: Request #165063 is New

Dear User,

We have received your request with the following details:

Request

- **Type of request** : (not defined)
- **Status** : New
- **Subject** : translation
- **Description** : Linguistic validation of PedsQL Generic Core Scales
- **Instruments** :

10.6 Appendix 6 Parent/Guardian information leaflet for cognitive interviews

Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mkulu wakafukufukuyu: Lucky Gift Ngwira, MLW Clin Res Programme & Liverpool Sch of Trop Medicine, P.O. Box 30096, Chichiri, Blantyre 3. MALAWI. Tel: 01876444
Email: Lngwira @mlw.mw

Mawu oyamba

Mukupemphedwa ngati kuli kothekera kuti mwana wanu atenge nawo mbali mu kafukufuku wokhudzana ndi umoyo wabwino. Chonde werengani mosamala uthengawu kapena umene wawerengedwa kwa inu. Chikalata chimodzi mudzapatsidwa kuti musunge. Ngati mukufuna kudziwa zambiri chonde khalani omasuka kutifunsa ife.

Kodi kafukufukuyu ndi wa chani?

Kafukufukuyu akukhudzana ndi mafunso amene amafunsidwa kuti tidziwe mmene moyo wabwino ulili ndinso kudziwa ngati mafunsowa ali oyenelera kwathu kuno. Otenga nawo mbali mu kafukufukuyu tidzawafunsa kuti atiuze chimene umoyo umatanthauza komanso m'mene akumvera za umoyo wawo.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Nkutheka kuti mwana wanu amapanga maphunziro ake pa imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa mwana wanga?

Tidzafunsa mwana wanu kukhala nawo mu makambitsirano a mwana wanu ndi wochititsa kafukufuku kufuna kudziwa m'mene ana akutha kumvera mafunso amene amafunsidwa kuti tidziwe za moyo wabwino wa ana. Pakali pano, mafunsowa anakonzedwa ku mayiko aku ulaya, ndipo tikufuna chithandizo cha ana kuti tiwunike ngati mafunsowa ali oyenelera kwathu kuno. Mafunso ake ndi wokhudza m'mene moyo wa mwana wanu uliri kuphatikizapo ngati akumva kupweteka kapena ululu ndiponso m 'mene akumvera mthupi.

Mwana wanga adzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambilanazi zidzatenga utali wosaposera ola limodzi. Chifukwa cha ichi mwana wanu adzapatsidwa cholembera komanso kope.

Chinsinsi ndi kasungidwe ka uthenga

Uthenga womwe mwana wanu ati adzapeleke udzatapidwa ndi makina otapila ndi cholinga choti uthengawu usasowe. Pambuyo pamafunso, uthengawu uzatanthauzidwa kupita kuchilankhulo cha chingelezi ndipo zidzasungidwa mmakina a kompyuta mosagwiritsa ntchito dzina la mwana wanu. Uthengawu adzatha kuufikila ndi anzathu okha ogwira ntchito mukafukufukuyu. Uthenga wotapidwawu tidzausunga kufikira pamene tidzautembenuze kukhala m'mapepala. Zotembenuzidwa m'mapepala zizasungidwa kwa zana zisanu zokha, koma ngati ukuvomereza tidzasunga kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:

Kafufukuyu si wamankhwala operekedwa kuchipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wa mwana wanu. Komabe zitha kutheka kuti mwina pamakambitsirano athu mwana atha kutopa kapena kukhumudwa, ndipo titakhala kuti tawona zimenzi tizasiira pompo macheza athu, ndi kuwafotokozera aphunzitsi kapena inuyo.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?

Tidzakupemphani kuti musayine kapena chidindo cha chala chikalata chopereka chilolezo kuti mwana wanu alowe nawo mukafukufukuyi. Mwana wanu naye adzafunsidwa ngati ali okondweretsedwa kulowa nawo mukafukufuku ndipo atapereka chilolezocho, adzalowa mumakambitsirano a pa gulu kapena a munthu payekha.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?

Zopezeka za kafukufukuyu zidzatha kupezeka kudzera mmisonkhano yomwe idzakonzedwe ndi a Malawi Liverpool Wellcome Trust. (MLW). Zotsatila zi zidzasindikizidwa nso ndipo zidzatha kupezeka kwa inu ngati mungafunse- Chonde gwiritsani ntchito ma nambala omwe apelekedwa pansi pa tsamba la uthengali kuti mupemphe uthenga wa zotsatira za kafukufukuyu.

Chitachitike ndi chani ngati sindikufuna kuti mwana kutenga nawo mbali mukafukufukuyu?

Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chanu ndinso cha mwana wanu. Maganizo anu azalemekezedwa ndipo izi sizidzakhala ndi vuto pa moyo ndi chithandizo chanu kapena mwana wanu cha zaumoyo.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?

Mungathe kutenga mpaka sabata imodzi kupereka chilolezo. Komanso mutafuna kuti mudziwe zambiri, chonde funsani a Lucky G Ngwira pa 0999-076-932. Ngati mungakhale ndi

funso lilonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene mwakhudzidwira ndi kafukufuku, chonde masukani polumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.7 Appendix 7 Children information leaflet for cognitive interviews

2d.C- The VOICE! study Children Information Leaflet (Cognitive Interviews)

UTHENGA KWA ANA (zaka 8 mpaka 12)

Mutu wa Kafukufukuyu: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mawu oyamba

Mukupemphedwa kulowa nawo mu kafukufuku. Kafukufuku ndi njira imodzi yopezera mayankho ku mafunso amene tilibe mayankho.

Kodi kafukufukuyu ndi wa chiyani?

Kafukufukuyu akufuna kupeza ngati mafunso amene timafunsa munthu akadwala ndi oyenera. Tikufunsa chimene moyo wabwino umathanthauza komanso m'mene ukumvera mthupi mwako.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mu sukulu zimene zili gawo la bungwe la Malawi-Liverpool-Wellcome Trust pansi pa dipatimenti ya Sayansi CoMm. Dipatimentiyi ikulimbikitsa ana a mu sukulu zozungulira m'zinda wa Blantyre kutenga maphunziro a sayansi. Nkutheka kuti umayimba sukulu mu imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Ngati ungakonde kulowa nawo mu kafukufukuyi,	
Choyamba	**Tidzapempha iwe ndi bambo, mai kapena amene umakhala nawo kuti ulembe dzina lako pa fomu kusonyeza kuti ukupereka chilolezo.**
Chachiwiri	**Udzapemphedwa kuyankha mafunso ochepa.**
Chachitatu	**Kenako tidzafuna kucheza nawe zokhudzana ndi mafunsowo.**
Chachinayi	**Zokambiranazi zidzakhala pakati pa iwe ndi wochititsa kafukufuku basi.**

Kodi ndiyenera kunena inde?

Ayi, sichoncho. Zili ndi iwe! Ungonena ngati sukufuna kutenga nawo mbali. Zimenezi sizidzakhudza m'mene ukuchitira mukalasi kapena thandizo lako lakuchipatala ukadwala.

Kodi ndidzakhala mu kafukufuku kwa nthawi yaitali bwanji?

Zokambiranazi pakati pa iwe ndi wochititsa kafukufuku zidzatenga pakati pa mphindi 30 ndi ola. Ukhoza kupumira nthawi iliyonse. Potenga nawo mbali udzapatsidwa cholembera komanso kope.

Kodi munthu wina azadziwa kuti ndikupanga nawo kafukufukuyu?

Ndi okhawo ali gulu lakafukufuku wathu amene azadziwe. Palibe winanso amene azadziwe chifukwa tidzakupatsa nambala imene udzagwiritse ntchito nthawi ya zokambiranazo.

Chidzachitike nchiyani pa zimene ndidzanene?

Tidzalemba komanso kuika mawu amenewo mu malipoti. Malipotiwa adzagwiritsidwa ntchito ndi ma dokotala, anamwino (ma nesi) komanso ochita kafukufuku. Tikhoza kukupatsa lipotili utafuna. Uthengawu tizausunga zaka zokwana zisanu zokha koma titha kudzausunga kwa nthawi yochulukirapo utavomereza.

Ndingapite kwa ndani kuti ndidziwe zambiri za kafukufukuyi?

Utha kutidziwitsa m'masiku asanu ngati ukufuna kutenga nawo mbali m u kafukufukuyu. Ngati uli ndi funso lina lililonse, bambo, mai kapena amene umakhala nawo atha kukuyankha. Ukhozanso kupempha ochitisa kafukufukuyu kuti akuyankhe.

Kuti umve zambiri, funsa mkulu wa kafukufukuyi:
Lucky Gift Ngwira,
Malawi-Liverpool-Wellcome Trust,
P.O. Box 30096, Chichiri, Blantyre 3. MALAWI.
Tel: 01876444
Email: Lngwira @mlw.mw

10.8 Appendix 8 Adolescent Information leaflet-for cognitive interviews

UTHENGA KWA ACHINYAMATA (zaka 13 mpaka 17)

Mutu wa kafukufuku: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mawu oyamba

Tikukupempha ngati kuli kotheka kuti utenge nawo mbali mu kafukufuku.

Kodi kafukufukuyu ndi chiyani?

Kafukufukuyu ndi njira imozi yopezera mayankho ku mafunso amene tilibe mayankho.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Nkutheka kuti mwina umapanga maphunziro ako pa imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Tidzakupempha kukhala nawo mu makambitsirano okhudza mafunso amene atanthauzidwa kuchoka m'chingelezi kubwera m'chichewa. Makambitsiranowa adzakhala pakati pa a iweyo ndi wochititsa kafukufuku kuti timve maganizo ako ku mafunso otanthauzidwawo ngati ali oyenelera kwathu kuno. Tidzakufunsanso za maganizo pa chimene umoyo umatanthauza, kuphatikizapo umoyo wako ngati ukumva kupweteka kapena ululu ndiponso m 'mene ukumvera mthupi.

Kodi ndidzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambilanazi zidzatenga utali wosaposera ola limodzi.

Chinsinsi ndi kasungidwe ka uthenga

Udzapatsidwa nambala nthawi ya makambitsirano kuti dzina lako litsagwiritsidwe ntchito. Zokambiranazo zidzatapidwa ndi makina otapila ndi cholinga choti uthengawu usasowe. Pambuyo pamafunso, uthengawu uzatanthauzidwa ndi kuika m'mitu ikulu ikulu ndipo zimenezi zidzasungidwa mmakina a kompyuta. Anzathu okha ogwira ntchito mukafukufukuyu ndi amene adzatha kuufikila uthengawu. Uthenga wotapidwawu tidzausunga kufikira pamene tidzautembenuze kukhala m'mapepala. Zotembenuzidwa

m’mapepala zizasungidwa kwa zaka zisanu, koma ngati ukuvomereza titha kudzasunga kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:
Kafufukuyu si wamankhwala operekedwa kuzipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wako. Komabe, timadziwa kuti pamakambitsirano ena atha kutopa kapena kukhumudwa. Utafuna kusiya zokambiranazo utha kutero nthawi iliyonse.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?
Tidzakupempha kuti usayine kapena chidindo cha chala chopereka chilolezo kuti ukuvomereza kulowa nawo mukafukufukuyi. Makolo ako nawo adzafunsidwa ngati akuvomereza iweyo kulowa nawo mukafukufuku.

Malipiro
Kutenga mbali mukafukufuku suzalipilidwa china chilichonse. Komabe, udzapatsidwa kope ndi cholembera.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?
Zopezeka za kafukufukuyu zidzasindikizidwa mu malipoti amene madokotala, anamwino (ma nesi) ndi ena ogwira ntchito zakafukufuku amagwiritsa. Zimenzi utha kuyitanitsa utafuna kukhala nazo pogwiritsira ntchito ma nambala amene ali m’musimu.

Chitachitike ndi chani ngati sindikufuna kutenga nawo mbali mukafukufukuyu?
Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chako. Maganizo ako azalemekezedwa ndipo izi sizidzakhala ndi vuto pa zasukulu yako, moyo ndi chithandizo chako.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?
Ungathe kutenga mpaka sabata imodzi kupereka chilolezo chofuna kulowa nawo mukafukufuku. Komanso utafuna kuti mdziwe zambiri zokhudzana ndi kafukufukuyi, chonde funsa a Lucky G Ngwira pa 0999-076-932. Ngati ungakhale ndi funso lina lililonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene wakhudzidwira ndi kafukufuku, utha kulumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.9 Appendix 9 Consent/Assent Form

5b. Chi- The VOICE! study children assent form

Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Principal Investigator: Lucky Gift Ngwira, MLW Clin Res Programme & Liverpool Sch of Trop Medicine, P.O. Box 30096, Chichiri, Blantyre 3. MALAWI. Tel: 01876444 Email: Lngwira @mlw.mw

		Inde	Ayi
1	Ndawerenga, kapena wina wandiwerengera kalata ya uthenga		
2	Ndinapatsidwa mpata wofunsa mafunso ndipo ndakhutitsidwa ndi mayankho amene		
3	Ndikuvomereza kuti nditha kutenga nawo mbali komanso kuti ndikhoza kusintha maganizo pa nthawi ina ili yonse.		
4	Ndamvetsetsa kuti mau anga adzalembedwa mmalipoti a kafukufuku koma dzina langa silidzagwiritsidwa ntchito mmalipotiwa		
5	Ndikuvomereza mwa kufuna kwanga, kutenga nawo mbaili mu kafukufukuyu		
6	Ndikuvomereza mwa kufuna kwanga kwanga kuti mwana wanga atenge nawo mbali mu kafukufuku		
7	Ndikuvomereza kuti mwana wanga atha kufunsidwa mafunso maulendo awiri		

Dzina la otenga Mbali:	**Sayini:**	**Tsiku:**
Dzina la kholo/oyang'anira:	**Sayini/ Chidindo cha chala:**	**Tsiku:**
Dzina la mboni:	**Sayini:**	**Tsiku:**
Dzina la ogwira ntchito:	**Sayini:**	**Tsiku:**

10.10 Appendix 10 Malawi College of Medicine research ethics approval

CERTIFICATE OF ETHICS APPROVAL

This is to certify that the College of Medicine Research and Ethics Committee (COMREC) has reviewed and approved a study entitled:

P.10/18/2509 - Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents:Adaptation for use in economic evaluation in sub-Saharan Africa. The VOICE! Study version 2.0 dated 20190111 by Mr. Lucky Ngwira

On 08-Feb-19

As you proceed with the implementation of your study, we would like you to adhere to international ethical guidelines, national guidelines and all requirements by COMREC some of which are indicated on the next page for your study

08-Feb-19

Dr. YB. Mlombe - Chairperson (COMREC)

Date

10.11 Appendix 11 LSTM Research ethics approval

Mr. Lucky Gift Ngwira
Malawi Liverpool Wellcome Trust
P.O. Box 3009
Chichiri, Blantyre
Malawi

LSTM
LIVERPOOL SCHOOL OF TROPICAL MEDICINE
Pembroke Place,
Liverpool, L3 5QA, UK
Tel: +44(0)151 705 3100
Fax: +44(0)151 705 3370
www.lstmed.ac.uk

Tuesday, 15 October 2019

Dear Mr. Ngwira,

Re. Research Protocol (19-045) *Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. The VOICE! study*

Thank you for your letter of 14th October 2019 responding to the action points raised by the Committee and providing the necessary in-country approvals for this project. I can confirm that the protocol now has formal ethical approval from the LSTM Research Ethics Committee.

The approval is for a fixed period of three years and will therefore expire on 14th October 2022 The Committee may suspend or withdraw ethical approval at any time if appropriate.

Approval is conditional upon:

- Continued adherence to all in-country ethical requirements.
- Notification of all amendments to the protocol for approval before implementation.
- Notification of when the project actually starts.
- Provision of an annual update to the Committee.
  Failure to do so could result in suspension of the study without further notice.
- Reporting of new information relevant to patient safety to the Committee
- Provision of Data Monitoring Committee reports (if applicable) to the Committee

Failure to comply with these requirements is a breach of the LSTM Research Code of Conduct and will result in withdrawal of approval and may lead to disciplinary action. The Committee would also like to receive copies of the final report once the study is completed. Please quote your Ethics Reference number with all correspondence.

Yours sincerely

Professor Graham Devereux
Chair
LSTM Research Ethics Committee

10.12 Appendix 12 Letter of approval from the Blantyre district health office

Telephone: Blantyre 0 1875332 / 01 877 401
Fax. 01 875 430 / 01 872 331

Communication should be addressed to:
The District Health Officer

In reply please quote No. .....

DISTRICT HEALTH OFFICE
P/BAG 65
BLANTYRE
MALAWI

REF. NO. : BT DHO/MED/9

24th September 2018

The Chairman
College of Medicine
Private Bag 360
Chichiri
BLANTYRE 3

Dear Sir/ Madam

RE : LETTER OF SUPPORT FOR A STUDY TITLED « MULT –ATTRIBUTE UTILITY INSTRUMENT IN CHILDREN AND ADOLESCENTS : ADAPTATION FOR USE IN ECONOMIC EVALUATION IN SUB-SAHARAN AFRICA

Blantyre District Health Office renders its support to the above mentioned study which is to be conducted by Lucky Ngwira as part of his PhD studies.

The study touches an area which is not well studied in our setting. It is for this reason that this project has the full of our office.

We look forward to the findings from this study.

Yours faithfully

Dr. M. Jere
DISTRICT MEDICAL OFFICER

10.13 Appendix 13 Demographic and response form for cognitive interviews

Table 10.3 Demographic and response form for cognitive interviews

1. Demographic Form

INTERVIEW IDENTIFICATION		
A01	Interviewer name	_____
A02	Date of assessment	☐☐ / ☐☐☐ / 2019 d d / m m m
A03	Name of facility	_____
PARTICIPANT IDENTIFICATION		
A04	Participant ID	0 0 ☐ /

A. SOCIODEMOGRAPHIC INFORMATION			
A05	Zaka/ Age		☐☐
A06	Sex	☐1 =Male	☐2 =Female
A07	Kumene mukukhala Residence	_____	
A08	Mumatha kulemba ndi kuwerenga? Can you read or write?	☐1 =Yes	☐2 =No
B. GENERAL UNDERSTANDING			

Please read and complete the EQ-5D-Y (without help). (Replace for PedsQoL)

The EQ-5D-Y is a brief questionnaire on health that has been translated from English, and we want to check that it is clear and well written in the Chichewa language

Answer questions carefully but there are no right or wrong answers

After completing the questionnaire, we will ask some questions about what you think about the questionnaire.

After completing the questionnaire, record time it took to complete the questionnaire

A09	Nthawi Time		

C. GENERAL IMPRESSION

Is it clear, easy to understand, easy to answer?	Is it too long?	Are the instructions clear?	i).Which areas should be improved to enhance comprehension and ii).how?	Can you give an example of someone with this level of health (e.g. moderate problems washing or dressing self)?	Interviewer assessment of performance

D. INDEPTH ANALYSIS OF WHOLE QUESTIONNAIRE

For each instruction in the EQ-5D (replace for PedsQoL), ask respondents:

	Chikalata cha mafunso a za Umoyo	Chichewa cha ku Malawi	Kufotokoza za umoyo wako pa tsiku la LERO	Pansi pa mutu ulionse, chonga bokosi LIMODZI limene likufotokoza bwino za umoyo wako pa tsiku la LERO	Umoyo wako uli bwino bwanji pa tsiku la LERO	Tikufuna tidziwe m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino pa tsiku la LERO	Mzere uwu walembedwa manambala kuyambira pa 0 mpaka 100	100 akutanth-auza umoyo wabwino koposa umene ungauga-nizire	0 akutanth-auza umoyo umene suli bwino koposa umene ungauga-nizire	Chonde lemba chizindikiro cha X pa mzere pamene pakuonetsa m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino pa tsiku la LERO	Umoyo wabwino koposa umene ungauga-nizire	Umoyo umene si uli koposa umene ungauga-nizire
Are there words that you find difficult to understand? Which words?				Interviewer assessment of performance								
Did you find it childish? (only applies to 12-15yr old)												

Would you reword it in another way?												
Would you change anything in the instructions to make it easy to understand?												
Would you add or remove anything to the instructions?												
What does the following mean to you?												
'umoyo'												
moyo												
chonga bokosi LIMODZI												

LERO	
Umoyo wabwino /si uli bwino koposa umene ungauga-nizire	
What were you thinking about when you read instructions?	

For each dimension question and response option, ask the respondents:

Second consensus version	Do you understand the question? N/Y What does it mean for you?	What were you thinking about when you answered the question? Were you thinking about any particular aspect of your health or your current situation?	What would you change to make this question better and easy to understand? Why?	What does each of the words/phrase mean to you?	Additional comments
1) Chikalata cha mafunso a za Umoyo					

2) Chichewa cha ku Malawi					
3) Kufotokoza za umoyo wako pa tsiku la LERO					
4) Pansi pa mutu ulionse, chonga bokosi LIMODZI limene likufotokoza bwino za umoyo wako pa tsiku la LERO					
5) MAYENDEDWE (kutha kuyenda)					
6) Ndilibe mavuto alionse poyenda					
7) Ndili ndi mavuto pang'ono poyenda					
Question Number And Theme	Clarity of the instruction and suggestions for improvements	Interviewer assessment of performance	Self -assessment of performance		
8) Ndili ndi mavuto aakulu poyenda		How about 'Ndili ndi mavuto ambiri poyenda?'			Would you prefer use of 'mavuto ambiri'

					or 'mavuto aakulu?'
9) KUDZISAMALIRA NDEKHA					
10) Ndilibe mavuto posamba kapena kudziveka ndekha					
11) Ndili ndi mavuto pang'ono posamba kapena kudziveka ndekha					
Question Number And Theme	Clarity of the instruction and suggestions for improvements	Interviewer assessment of performance	Self -assessment of performance		
12) Ndili ndi mavuto aakulu posamba kapena kudziveka ndekha					Would you prefer use of 'mavuto ambiri' or 'mavuto aakulu?'
13) KUCHITA ZINTHU ZA TSIKU NDI TSIKU (mwachitsanzo, kupita ku sukulu, kuchita zimene					

ndimakonda, masewero olimbitsa thupi, kusewera, kuchita zinthu ndi banja langa kapena anzanga)					
14) Ndilibe mavuto alionse pochita zinthu zanga za tsiku ndi tsiku					
15) Ndili ndi mavuto pang'ono pochita zinthu zanga za tsiku ndi tsiku					
16) Ndili ndi mavuto aakulu pochita zinthu zanga za tsiku ndi tsiku					Would you prefer use of 'mavuto ambiri' or 'mavuto aakulu?'
17) KUMVA KUPWETEKA KAPENA KUPHWANYA M'THUPI					What word could you use that describes itching, nausea, dizziness?

KOSOWETSA MTENDERE					How do you understand ululu? Is it the same as kumva kupweteka?
18) Sindikumva kupweteka kapena kuphwanya m'thupi					
19) Ndikumva kupweteka pang'ono kapena kuphwanya m'thupi pang'ono					How would you understand 'Ndikumva kuphwanya mthupi mwapakatikati?'
20) Ndikumva kupweteka kwambiri kapena kuphwanya m'thupi kwambiri					
21) KUDANDAULA, KUMVA CHISONI KAPENA KUSASANGALALA					
22) Sindikudandaula, sindikumva chisoni					

kapena osasangalala					
23) Ndikudandaula pang'ono, ndikumva chisoni pang'ono kapena ndikusangalala pang'ono					
24) Ndikundaula kwambiri, ndikumva chisoni kwambiri kapena sindikusangalala					
25) Umoyo wako uli bwino bwanji pa tsiku la LERO					
26) Tikufuna tidziwe m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino pa tsiku la LERO					

27) Mzere uwu walembedwa manambala kuyambira pa 0 mpaka 100					
28) 100 akutanthauza umoyo wabwino koposa umene ungauganizire					
29) 0 akutanthauza umoyo umene suli bwino koposa umene ungauganizire					
30) Chonde lemba chizindikiro cha X pa mzere pamene pakuonetsa m'mene umoyo wako ulili bwino kapena m'mene umoyo wako sulili bwino pa tsiku la LERO					

31) Umoyo wabwino koposa umene ungauganizire					
32) Umoyo umene si uli bwino koposa umene ungauganizire					

33) What does each of the words/phrase mean to you?
a. I have no problems washing or dressing myself
b. 'a lot of' (What would you prefer? Use of ambiri or aakulu?)

c. ‘very’
34) SUMMARY
35)

10.14 Appendix 14 EurQol Group EQ-5D-Y-5L translation approval

Tuesday, September 7, 2021 at 9:18:56 PM South Africa Standard Time

Subject: RE: TRF2171 RE: Cognitive report for the EQ-5D-Y 5L
Date: Monday, 13 January 2020 at 15:45:32 South Africa Standard Time
From: EuroQol - VMC Office
To: Jennifer Jelsma, Lucky Ngwira
CC: EuroQol - VMC Office
Attachments: image003.png, image004.png, image005.png

Dear Lucky,

Congratulations with completing the Malawi (Chichewa) EQ-5D-Y-5L Paper Self-Complete version.

I confirm safe receipt of the final version and the translation report. It's now being processed for internal review. For your information, I have changed the font to Arial (as Comic Sans is not the default EQ-5D-Y font anymore) and changed the font size accordingly. I will forward a copy to you once the version is effective.

As always, it was a pleasure working with you.

Kind regards,

Nalinie Banarsi
Legal Counsel / VMC Project Manager
EuroQol Research Foundation

EQ-5D

T +31 88 4400191 | E banarsi@euroqol.org | www.euroqol.org | Marten Meesweg 107 | 3068 AV Rotterdam | The Netherlands

From: Jennifer Jelsma <jennifer.jelsma@uct.ac.za>
Sent: Tuesday, 24 December 2019 05:12
To: Lucky Ngwira <lngwira@mlw.mw>
Cc: EuroQol - VMC Office <vmc@euroqol.org>
Subject: RE: TRF2171 RE: Cognitive report for the EQ-5D-Y 5L

Dear Lucky,

Congratulations on a job well done!

I hope you have peaceful and happy Christmas season and a successful New Year. Look forward to our continuing collaboration.

All the best,

Jennifer

From: Lucky Ngwira <lngwira@mlw.mw>
Sent: Friday, December 20, 2019 6:49 PM
To: Jennifer Jelsma <jennifer.jelsma@uct.ac.za>; saraheq2018@outlook.com; EuroQol - VMC Office <vmc@euroqol.org>; banarsi@euroqol.org

10.15 Appendix 15 MAPI Trust PedsQL 4.0 translation approval

Subject: RE: ePROVIDE™: Request #165063_PedsQL_Malawi
Date: Friday, 27 December 2019 at 16:32:59 South Africa Standard Time
From: Bellet, Alexandra
To: Lucky Ngwira
Attachments: image011.jpg, image012.jpg, image013.jpg, image014.jpg, image015.gif, image016.gif, image017.gif, image018.png, image019.png, image020.gif, image021.png

Dear Lucky,

Many thanks for providing the translations, I confirm safe receipt.
Again, thank you for your hard work which is very useful for us, in Mapi Research Trust, and for future researchers.

The versions are approved, you may send us the clean versions. Thank you.
I wish you all the best for the end of the year.

Best regards,

Alexandra Bellet
Client Services Assoc II, PCS
Commercialisation & Outcomes

Tel : (+33) 4 72 13 37 47
Alexandra.Bellet@mapi-trust.org
www.mapi-trust.org ; https://eprovide.mapi-trust.org/

Mapi Research Trust

Upcoming author webinars:
Mar 12th - 10am ET Register here
Author Webinar on Patient-Reported Outcomes Assessments:
Spotlight on the Quality of Life Disease Impact Scale (QDIS®)

From: Lucky Ngwira <lngwira@mlw.mw>
Sent: lundi 9 décembre 2019 15:25
To: Bellet, Alexandra <Alexandra.Bellet@mapi-trust.org>
Subject: Re: ePROVIDE™: Request #165063_PedsQL_Malawi

Dear Alexandra,

Please find attached the post-proof-reading version which are now being submitted as the pre-final version both with tracked changes and the clean versions.

I intend to provide a full translation report once these versions get approved by your office.

I look forward to hearing from you soon.

Regards,
Luckh

10.16 Appendix16 Parent/Guardian information sheet for psychometric evaluation

Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mkulu wakafukufukuyu: Lucky Gift Ngwira, Malawi-Liverpool-Wellcome Trust Clinical Research Programme & Liverpool School of Trop Medicine, P.O. Box 30096, Chichiri, Blantyre 3. MALAWI. Tel: 01876444 Email: Lngwira @mlw.mw

Mawu oyamba

Mukupemphedwa ngati kuli kotheka kuti mwana wanu atenge nawo mbali mu kafukufuku wokhudzana ndi umoyo wabwino. Chonde werengani mosamala uthengawu kapena umene wawerengedwa kwa inu. Chikalalata chimodzi mudzapatsidwa kuti musunge. Ngati mukufuna kudziwa zambiri chonde khalani omasuka kutifunsa ife.

Kodi kafukufukuyu ndi wa chiyani?

Kafukufukuyu ndiwofuna kudziwa ngati mafunso amene amafunsidwa kuti tidziwe mmene moyo wabwino ulili ali oyenelera kwathu kuno. Otenga nawo mbali mu kafukufukuyu tidzawapempha kuti atidziwitse chimene umoyo umatanthauza komanso m'mene akumvelera za umoyo wawo.

Chifukwa chiyani mwana wanga akupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Mwana wanu amapanga maphunziro ake pa imodzi ya sukuluzi.

Chidzachitike ndi chiyani kwa mwana wanga?

Mwana wanu komanso ana ena adzafunsidwa kuyankha mafunso a payekha payekha. Mafunsowa ndi okhuza moyo wabwino ndipo adzafunsa m'mene moyo wake uliri, ngati akumva kupweteka kapena ululu ndiponso m 'mene akumvera mthupi

Mwana wanga adzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambiranazi zidzachitika kamodzi kapena kawiri (masabata awiri kuchokera pa zokambirana zoyamba) ndipo sipazatenga nthawi yoposera mphindi makumi atatu.

Chinsinsi ndi kasungidwe ka uthenga

Uthenga womwe mwana wanu ati adzapeleke udzalowetsedwa mmakina a kompyuta pogwiritsa ntchito nambala m'malo mwa dzina ndi cholinga choti uthengawo ukhale wachinsisi. Uthengawu udzasungidwa mu kabati kwa nthawi zonse ndipo okha ogwira ntchito mukafukufukuyu ndi amene azakhoza kufikira uthengawu. Mayankho a mwana wanu komanso ana ena olowa nawo mukafukufukuyi azasonkhanitsidwa pamodzi ndikusindikidzidwa mumalipoti amene sazaonetsa dzina lawo kapena thandizo lakuchipatala limene akulandira. Uthengawu uzasungidwa kwa zana zisanu zokha, koma ngati muvomereza tidzasunga zolembedwa pa kompyuta kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:
Kafufukuyu si wamankhwala operekedwa kuchipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wa mwana wanu. Komabe zitha kutheka kuti mwina pamakambitsirano athu mwana atha kutopa kapena kukhumudwa, ndipo titakhala kuti tawona zimenzi tizasiira pompo macheza athu, ndi kuwafotokozera aphunzitsi kapena inuyo.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?
Tidzakupemphani kuti musayine kapena kudinda chidindo cha chala chikalata chopeleka chilolezo. Mwana wanu naye adzafunsidwa ngati ali okondweretsedwa kulowa nawo mukafukufuku. Atapeleka chilolezocho, adzakhala nawo mu gulu la ana amene ati azafunsidwe mafunso okhudza moyo wabwino.

Malipiro
Kutenga mbali mukafukufuku simuzalipila china chilichonse. Sitidzakulipilani kupatula kupatsa mwana wanu kope ndi cholembera nthawi iliyonse angazalowe mukafukufukuyu.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?
Zotsatira za kafukufukuyu zidzatha kupezeka kudzera mmisonkhano yomwe idzakonzedwe ndi a Malawi Liverpool Wellcome Trust. (MLW). Zotsatila zi zidzasindikizidwa nso ndipo zidzatha kupezeka kwa inu ngati mungafunse- Chonde gwiritsani ntchito ma nambala omwe apelekedwa pansi pa tsamba la uthengali kuti mupemphe uthenga wa zotsatira za kafukufukuyu.

Chitachitike ndi chani ngati sinditenga nawo mbali mukafukufukuyu?
Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chanu ndinso cha mwana wanu. Maganizo anu azalemekezedwa ndipo izi sizidzakhala ndi vuto pa moyo kapena chithandizo chanu ndi cha mwana wanu cha zaumoyo.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?

Mungathe kutenga mpaka sabata imodzi kupereka chilolezo. Komanso mutafuna kuti mudziwe zambiri, chonde funsani a Lucky G Ngwira pa 0999-076-932. Ngati mungakhale ndi funso lilonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene mwakhudzidwira ndi kafukufuku, chonde masukani polumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.17 Appendix 17Children information sheet for psychometric evaluation

UTHENGA KWA ANA (zaka 6 mpaka 12)

Mutu wa Kafukufukuyu: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. ***The VOICE! study***

Mawu oyamba

Mukupemphedwa kulowa nawo mu kafukufuku. Kafukufuku ndi njira imodzi yopezera mayankho ku mafunso amene tilibe mayankho.

Kodi kafukufukuyu ndi wa chiyani?

Kafukufukuyu akufuna kupeza m'mene umamvera ukadwala kapena ukakhala kuti ukudwala. Tikufunsa chimene moyo wabwino umathanthauza komanso m'mene ukumvera mthupi mwako.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mu sukulu zimene zili gawo la bungwe la Malawi-Liverpool-Wellcome Trust pansi pa dipatimenti ya Sayansi CoM. Dipatimenti ya Science CoM ikulimibikitsa ana a mu sukulu zozungulira m'zinda wa Blantyre kutenga maphunziro a sayansi. Nkutheka kuti umayimba sukulu mu imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Ngati ungakonde kulowa nawo mu kafukufukuyi,	
Choyamba	**Tidzapempha iwe ndi bambo, mai kapena amene umakhala nawo kuti ulembe dzina lako pa fomu kusonyeza kuti ukupereka chilolezo.**
Chachiwiri	**Udzapemphedwa kuyankha mafunso ochepa.**
Chachitatu	**Tidzakupempha kuchita zimenezi kamodzi kapena kawiri.**

Kodi ndiyenera kunena inde?

Ayi, sichoncho. Zili ndi iwe! Ungonena ngati sukufuna kutenga nawo mbali. Ukatenga nawo mbali kenako kusintha maganizo utha kusiya nthawi ina iliyonse. Zimenezi sizidzakhudza m'mene ukuchitira mukalasi kapena thandizo lakuchipatala ukadwala.

Kodi ndidzakhala mu kafukufuku kwa nthawi yaitali bwanji?

Zokambiranazi zizakhala kamodzi kokha kapenanso kawiri. Nthawi zonse zimenzi zidzatenga mphindi 30. Ukhoza kupumira ukafuna. Nthawi iliyonse udzatenge nawo mbali mukafukufukuyu udzapatsidwa cholembera komanso kope.

Kodi munthu wina azadziwa kuti ndikupanga nawo kafukufukuyu?

Ndi okhawo ali gulu lakafukufuku wathu amene azadziwe. Palibe winanso amene azadziwe chifukwa tidzakupatsa nambala imene udzagwiritse ntchito nthawi ya zokambiranazo.

Chidzachitike nchiyani pa zimene nditi ndidzanene?

Tidzalemba uthengawo mu malipoti amene madokotala, anamwino (ma nesi) komanso ochita kafukufuku angadzagwiritse ntchito. Tikhoza kukupatsa lipotili utafuna. Uthengawu tizausunga zaka zokwana zisanu zokha koma utafuna kuti tizazisunge kwa nthawi yochulukirapo ukhoza kunena.

Ndingapite kwa ndani kuti ndidziwe zambiri zakafukufukuyi?

Utha kutidziwitsa m'masiku asanu ngati ukufuna kutenga nawo mbali mu kafukufukuyu. Ngati uli ndi funso lina lililonse, mai, bambo kapena amene umakhala nawo atha kukuyankha. Ukhozanso kupempha ochitisa kafukufukuyu kuti akuyankhe.

Kuti umve zambiri, funsa mkulu wa kafukufukuyi:
Lucky Gift Ngwira,
Malawi-Liverpool-Wellcome Trust,
P.O. Box 30096, Chichiri, Blantyre 3. MALAWI.
Tel: 01876444
Email: Lngwira @mlw.mw

10.18 Appendix 18 Adolescent information sheet for psychometric evaluation

UTHENGA KWA ACHINYAMATA (zaka 13 mpaka 17)

Mutu wa kafukufuku: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mawu oyamba

Tikukupempha ngati kuli kotheka kuti utenge nawo mbali mu kafukufuku.

Kodi kafukufuku ndi chiyani?

Kafukufukuyu ndi njira imozi yopezera mayankho ku mafunso amene tilibe mayankho.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Nkutheka kuti mwina umapanga maphunziro ako pa imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Tidzakupempha kuyankha mafunso okhuza umoyo wako ngati ukumva kupweteka kapena ululu ndiponso m 'mene ukumvera mthupi.

Kodi ndidzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambilanazi zidzatenga utali wosaposera theka la ola. Komanso nkutheka kuti tidzakufunsanso mafunso awa pasanathe sabata ziwiri.

Chinsinsi ndi kasungidwe ka uthenga

Udzapatsidwa nambala nthawi ya makambitsirano kuti dzina lako litsagwiritsidwe ntchito. Zokambiranazo zidzaikidwa m'mitu ikulu ikulu ndipo zimenezi zidzasungidwa mmakina a kompyuta. Anzathu okha ogwira ntchito mukafukufukuyu ndi amene adzatha kuufikila uthengawu. Uthengawu udzasungidwa kwa zaka zisanu, koma ngati ukuvomereza titha kudzasunga kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:

Kafufukuyu si wamankhwala operekedwa kuzipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wako. Komabe, timadziwa kuti ena atha kukhumudwa ndi mafunso okhuza za umoyo. Utafuna kusiya kuyankha mafunsowa utha kutero nthawi iliyonse.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?
Tidzakupempha kuti usayine kapena chidindo cha chala chopereka chilolezo kuti ukuvomereza kulowa nawo mukafukufukuyi. Makolo ako nawo adzafunsidwa ngati akuvomereza iweyo kulowa nawo mukafukufuku.

Malipiro
Kutenga mbali mukafukufuku suzalipilidwa china chilichonse. Komabe, udzapatsidwa kope ndi cholembera nthawi iliyonse utadzatenge nawo mbali.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?
Zopezeka za kafukufukuyu zidzasindikizidwa mu malipoti amene madokotala, anamwino (ma nesi) ndi ena ogwira ntchito zakafukufuku amagwiritsa. Zimenzi utha kuyitanitsa utafuna kukhala nazo pogwiritsira ntchito ma nambala amene ali m'musimu.

Chitachitike ndi chani ngati sindikufuna kutenga nawo mbali mukafukufukuyu?
Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chako. Maganizo ako azalemekezedwa ndipo izi sizidzakhala ndi vuto pa zasukulu yako, moyo ndi chithandizo chako.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?
Ungathe kutenga mpaka sabata imodzi kupereka chilolezo chofuna kulowa nawo mukafukufuku. Komanso utafuna kuti mdziwe zambiri zokhudzana ndi kafukufukuyi, chonde funsa a Lucky G Ngwira pa 0999-076-932. Ngati ungakhale ndi funso lina lililonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene wakhudzidwira ndi kafukufuku, utha kulumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.19 Appendix 19 Demographic data for the psychometric evaluation

7926434016

EQ-5D-Y-5L: Mafunso a za umoyo

A. INTERVIEWER IDENTIFICATION

E01 Interviewer ID

E02 Date of completion - - 2 0 2

Date Format: DD-MON-YYYY

B. PARTICIPANT IDENTIFICATION

E03 Participant ID

E04 Age

E05 Sex ☐ Male(1) ☐ Female(2)

E06 Grade/Standard ☐ Standard(1) ☐ Form(2)

E07 Class ☐ 1 (1) ☐ 2 (2) ☐ 3 (3) ☐ 4 (4) ☐ 5 (5) ☐ 6 (6)

C. MEDICAL HISTORY (E08-E15 for children seen with carer/clinician. E08-E09 for school/community)

E08 Have you been seen by a doctor or burse at a hospital becasue you were not feeling well in the last month?
☐ Yes (1)
☐ No (2)

E09 If yes, do you visit this hospital or clinic to see the doctor or nurse every week or month?
☐ Yes (1)
☐ No (2)

E10 What is/was the main reason that brought you to the hospital/hospitalized today?
☐ Fever (1) ☐ Diarrhoea (4)
☐ Convulsion (2) ☐ Pain (5)
☐ Vomiting (3) ☐ Other (6), specify below
☐ N/A (7) [if completing in school]

E11 Do you suffer from any other condition/symptom other than what has been presented today?
☐ Yes (1) What is the condition?
☐ No (2) Go to E15

E12 How long have you had this other condition/symptom?
☐ <= 2 wks (1) ☐ 3-6 months (4)
☐ 2-4 wks (2) ☐ 6-12 months (5)
☐ 1-3 months (3) ☐ >1 year (6)

E13 Do you see a doctor/visit the hospital for this other condition/symptom?
☐ Yes (1) If yes, how often? ☐ Daily (1) ☐ Weekly (2) ☐ Monthly (3)
☐ No (2)

E14 Do you receive some medication for this other condition/symptom?
☐ Yes (1) If yes, how often? ☐ Daily (1) ☐ Weekly (2) ☐ Monthly (3)
☐ No (2)

E15 Status (Check in health passport to see if the child being treated has been treated for below conditions in last 12m)

	Yes (1)	No (2)
HIV	☐	☐
TB	☐	☐
Asthma	☐	☐
Epilepsy	☐	☐
Oncology	☐	☐
Diabetes	☐	☐
Other, specify	☐	☐

10.20 Appendix 20 Detailed description of pre-1995 adult generic health measures

Table 10.4 Detailed description of pre-1995 adult generic health measures

Measure*	SIP	QWB	SF-36	Rosser Kind	12 Dimensions	DUHP	HUI1	NHP	COOP	EQ-5D-3L
Ref	Bergner et al 1976; Bergner et al 1981	Kaplan et al 1976	Ware & sherbourne 1992	Rosser 1976	Sintonen 1981	Parkerson et al 1990	Torrance et al 1982	Hunt et al 1985	Nelson et al 1987; Nelson et al 1990	The EuroQol Group 1990
Year developed	1976	1976		1976	1981	1990	1982	1985	1987	1990
Construct	health status	symptoms and functioning	health status	measure of illness	measuring and valuing health states	measure of health and dysfunction	assessment of health status	population survey tool (from a quality of life measure that focussed on physical, mental and social health - Hunt et al 1980) for clinical and epidemiological research. The profile can be used to measure general perceived	measurement of health and functional status measure (efficiently assessing function in routine office practice)	health status, HRQoL

								health status or specific conditions of ill-health.		
Conceptual framework	start: 1972 with the formulation of procedures to collect statements describing sickness-related behavioural dysfunction from patients, healthy carers and health care professionals using open ended forms. Data was collected from several outpatient clinics of a university hospital, a hospital walk-in	Step 1: speciality by speciality review of medical reference works. Step 2: Ways (big or small) in one's behaviour as well as role performance is affected by diseases and injuries on functional status were listed; Step 3: matching of standard survey items with disruptions in role performance	Seven-year project. Reviewed the content of various measures used in measuring limitation in physical, social, and role functioning, general mental health and general health perceptions	Sixty doctors from various specialities helped in the development of the measure. The doctors were asked which criteria they use to decide which of two patients is more ill. Two major constructs emerged based on the principle of paired comparisons: i). observed disability; and ii) subjective distress that included pain, depression,	analysis of authoritative Finnish documents on health policy identified health=quantity (length of life) + quality. QoL=perceived health, psycho-physical functioning, and social functioning	shorten the DUHP-63. Original DUHP had 63 items that measured generic functional health status along four dimensions: symptom status (26 items on physical and psychologic symptoms), physical function (9 items on disability and capacity), social function (5 items on	developed from literature review and epidemiological surveys by a team of experts. Thereafter 84 children and parent pairs (same gender and living in same household) evaluated these to determine which were most important out of 15 (age of onset of health	Statements were collected from over 700 people describing the typical effects of ill-health - social, psychological, behavioural and physical - for example, 'I sleep badly'. 'I've lost interest in sex', 'I find it hard to walk about'. 2200 statements were initially collected to formulate concepts, and these were reduced to 138 after	The COOP was developed from reviewing clinical work by researchers such as Katz and Goldman (12,15), medical philosophers such as Dubos (18) who advocated that functional status was central in medicine, and health status	literature review and knowledge of researchers

clinic, a private partnership practice, and a prepaid group practice. Data were examined independently and then in a consensus where duplicates were eliminated. Original 312 item, then revised SIP with 189 items, and final version with 136-items. (Gilson et al 1975; Bergner et al 1976). finish: six-year project of field testing to come up

and other activities- the following sub-scales were created: mobility, physical activity, and social activity. 1974-6 went through validity and reliability testing in San Diego, USA

and anxiety. 29 combinations of eight states of disability and four states of distress.

social role performance), and emotional function (23 items on self-esteem). The conceptual framework for DUHP were based on the three WHO dimensions. Items were selected based on investigators' experience (face validity) in research, health promotion, and patient care, and based on psychometric properties

problem, cause of health problem, happiness, hearing, learning ability, mobility, name of disease or disorder, pain or discomfort, physical activity, play, school performance, self-care, sight, speech, use of limbs) potential attributes identified in the review. six attributes rated highly: sensory and communica

checking for redundancy, colloquialisms and ambiguity. After further field testing for two years (1976-78) the number of items was reduced to 82. In 1978 a survey tool was developed containing 38 items in six areas. Profile has two parts: part 1- contain 38 items in six areas; part2- consists of 7 statements about daily life affected by health

investigators such as Ware (18). The COOP measurement was named chart because the measure was handed to patients as a chart or was placed on a wall in the examination room.

	with a final product					as suggested by Ware et al 1980 like item score variance, item-convergent validity, item-discriminant validity. overall perceived health measure based on personal health status score was also included.	tion ability (vision, hearing and speech), happiness, self-care, pain or discomfort, learning and school ability, and physical activity ability. attribute levels were defined to cover the full range of possible disabilities as well as abilities			
Self-reported/ proxy/ interviewer administered	interviewer administered but participant answering only questions	interviewer administered	self-reported/ interviewer administered	interviewer administered	self-administered	self-report	self-administered	self-administered	self-reported	self-reported/ interviewer administered
Items	136	43 + symptoms	36	12	57	17	23	45 (38 for part 1-health	9	5

								problems on physical, mental and social health; and 7 for part 2-daily life activities)		
Domains[#]	2 dimensions (with 7 categories) + 5 independent categories	3	8	2	12	6 health measures; 4 dysfunction measures (anxiety, depression, pain, and self-esteem)	4 attributes	6 areas for part 1 (38 items) and daily life impact (7 statements) for part 2	9	5
	Independent categories:	mobility	physical functioning	disability (8)	perceived health	physical health	physical function	sleep (5 items)	physical condition	mobility
	sleep and rest	physical activity	role limitations because of physical health problems	distress (4)	breathing	mental health	role function	physical mobility (8 items)	emotional condition	selfcare
	eating	social activity	social functioning		sleeping	social health	social-emotional function	energy (3 items)	daily work	usual activities
	work	symptoms	general mental health (psychological distress and		speaking (communication)	general health (all 15 items, 5 each for physical,	health problems	pain (8 items)	social activities	pain/discomfort

	psychological well-being)		mental and social health)				
home management	role limitations because of emotional problems	moving	perceived health		emotional reactions (9 items)	pain	anxiety/depression
recreation and pastimes	bodily pain	eating	self-esteem	N.B. Within the skin (social	social isolation (5 items)	change in condition	
mobility	vitality (energy/fatigue)	incontinence		interaction outside the skin).		social support	
1.Physical:	general health perceptions	hearing	anxiety	Physical function: mobility and	how health impact daily life activities	quality of life (quality of life)	
ambulation mobility		seeing intellectual or mental functioning	depression pain	physical activity (6 items), Role			
body care and movement		working	disability	function: self-care and role			
2.Psychosocial:		social participation		activity (5 items), Social-emotional function: emotional well-being			

						and social activity (4 items), health problem (8 items)		
	social interaction alertness behaviour emotional behaviour communication							
Purpose	health care should reduce sickness and its effect on daily living. behavioural impact of sickness in terms of dysfunction and not assess levels of positive functioning.	refined classification system that would cover all possible functional status disruptions	most of the health status measures were quite lengthy with more than 100 items. The short form (SF) was supposed to be fill this gap, a standardized short form that was comprehensive and psychometrically valid.	come up with a reliable (through test reliability) classification system for diagnoses or symptoms of illness. 7 states in the disability x 4 distress=28 + unconscious state in the disability gives total 29health states	?	6 health measures based upon the three WHO dimensions: physical, mental, and social health. developed from Duke health profile (DUHP-63) a 63-item questionnaire measuring	evaluation of neonatal intensive care (HUI1- Boyle, Torrance et al 1983), modified for use in childhood cancer (HUI2- Torrance, Feeny et al 1996), developed and scaled	to develop a population survey tool from an existing instrument tool. Has two parts: profile part 1: comprises 38 statements regarding health problems in six areas: sleep, physical mobility,	Core: physical, social, and role functioning; emotional status; and overall health. Additional charts were developed according to need/purp

functional health in four dimensions: symptom status (26 items on physical and psychologic symptoms), physical function (9 items on disability days and capacity to use upper and lower extremities), social function (5 items on social role performanc e), and emotional function (23 items on self-esteem).	for the adult pop (HUI3- Feeny, Furlong et al 2002)	energy, pain, emotional reactions and social isolation; Part 2 of profile consists of 7 statements relating to how health problems affect daily life areas: paid employment, jobs around the house, social life, personal relationships, sex life, hobbies and interests, and holidays.	ose: i) clinical need-->pain and change in health (depend on need for clinician); health-related need--> social support and quality of life. Current COOP system has nine charts

Scoring	category (sub-scale score)	Items scored by the category method contained a numerical rating scale on each page, described as being constructed of 11 equal intervals. For magnitude estimation, a standard item representing the upper extreme of the scale was given a score of 1000.	Scored using a 2-step process: step1: items are score on 0 to 100 range so that the lowest score gets 0 and highest gets 100, representing a percentage score that can possibly be achieved. In step 2: the scores are averaged to create the 8 scale scores. Items with missing values are not included in scale score calculation, so that scale score represent the average for all items that	predictive model to measure individual's expectation of future illness. Each of the eight and four levels is valued to produce a matrix of scores for 29 states since the "unconscious" disability level has one score only. The Each number represents the value or "utility" of being in a particular health state, where a score of 1-00 reflects full health (no disability and	This could be done for example by first asking the subjects to assess their health status on the dimensions and then assess their overall health status on a O-100 or O-10 scale as used in this study, with 100 or 10 being assigned to complete health.	scored 0.0 to 100.0. For health measures, high scores=good health; for dysfunction measures, high scores=poor health. Item scores for items, and measure scores for physical health (items 8,9,10,11,12), mental health (items 1,4,5,13,14), social (items 2,6,7,15,16), general health (all 15items), perceived health (item	utility function	For part 1 items in each domain are weighted, the higher the score the greater the severity of perceived problems in that area-maximum score is 100 per area; total score is derived by averaging the six domain scores. The profile takes into account negative aspects of health only i.e., distress and can therefore not be used to assess positive feelings of well-being	five response choices representing a five-point ordinal scale. High scores representing unfavourable scores on the measure.	utility function

were answered.	no distress) and a score of 0 00 represents death. Two states have negative scores and are considered to be worse than death. The direction of the Rosser score is therefore opposite to other measures, e.g. The NHP	3- I am basically a healthy person), self-esteem (items 1,2,4,6,7), anxiety (items 2,5,7,10,12, 14) depression (items 4,5,10,12,13), pain (item 11), disability (item 17). 0-1, 1.0 indicating the highest level of health.

* SIP: sickness index profile; QWB: Quality well-being; HUI1: Health utility index mark 1; NHP: Nottingham health profile; COOP: The COOP chart; DUHP: Duke Health Profile; #domains in bold

10.21 Appendix 21 Detailed description of pre-1995 childhood generic health measures

Table 10.5 Detailed description of pre-1995 childhood generic health measures

Measure*	CHSCS	CHRS	CHQ-PF28/PF50/PF98	KINDL	EQ-5D-Y[$]
Ref	Hester NO et al	Maylath et al 1990	Landgraf et al 1998; Landgraf and Abetz	Ravens-Sieberer & Bullinger 1998	Wille et al 2010
Year	1984	1990	1994	1994	2010
Age range	7-13yrs	9-12ys		8-16yrs	8-15yrs
Construct	health and health related behaviors	self-rating of general health	measuring and comparing health	quality of life	HRQoL, health status
Conceptual framework	Its development was based on four beliefs about perceptions of children regarding "healthy" and "unhealthy" children: individual's health self-concepts are developed during childhood, this self-concept affects individual's health behaviour, iii). knowledge of an individual's health self-concept affect is of value to health professionals for planning and evaluating interventions, and iv. health self-concept	22 items were derived from a General Health Ratings Index. Factor analysis was used in creation of this index from Rand's health insurance experiment study which was composed of an adult population. Firstly, a fourth-grade teacher was consulted during the drafting stage regarding difficulty of reading material. The instrument was then group administered to 25 second through six graders using a convenience sample. The children were asked to circle all the words that	Developed same structure as the SF-36 but since it was targeted at children, scales that affect child's health functioning on the family. Child Health Questionnaire (CHQ) was first developed with 98 item scale for adolescents, but due to industry demands because of its length and practical length- a 50 item CHQ parent form (CHQ-PF50 for pre-school children aged 4-5yrs and older) short form parent-completed was empirically derived. The 28 item CHQ-PF28 has same 13 scales but only with a subset of the items of the CHQ-PF50.	conceptual framework where four components of quality of life (psychological well-being, social relationships, physical function and everyday life activities) were explored in interviews with from several classes. The measure contains 40 items which were constructed in two pilot studies (n=28 children each). The measure was then psychometrically evaluated for reliability, validity and sensitivity.	From EQ-5D-3L which was developed from literature review, general knowledge of developers and existing generic health status measures.

could be measured. The constructs were obtained from a field research study of perceptions of health and unhealthy children by school-age children. An open-statement questionnaire (what they think a healthy child is like, and ii). what they thought a not healthy child is like) was administered to 225 school age children 225 (aged 6-13yr). The responses were collapsed into 21 categories, 12 of which were used to developed CHSCS. Ten of sub scales had five items each, whereas emotional and healthy subscales had six and two items respectively, so that the total items were 58. Each item had four response options from true (score=1) to false (score=4) so that total lowest score=58

they difficult to comprehend. As a third step, the 22 items were pilot tested in three private elementary schools among 137 fourth to six grade children. Through statistical analysis, five items were removed to improve the instruments' reliability, thereby remaining with a 17-item scale.

and highest =232. Content validity involved expert (in measurement, child development, phenomenology, research, and early childhood education, and school nursing) review. Secondly, 40 children (5-13yrs) reviewed the instrument. Finally, an expert in instrument development (Harter) reviewed measure for structure and parallelism. Further empirical testing including reliability and validity was accomplished through a sample of 940 children. Critical analyses of the findings led to the conclusions that the CHSCS had moderate stability and high internal consistency reliability, evidence of content validity, but no evidence of construct

validity.

Self-reported/ proxy/ interviewer administered	self-reported	self-reported	self-reported CHQ-CF87 for adolescents; parent-completed version (includes four concepts designed to measure the emotional impact of the child's health)	self-reported	self-reported
Items	45	17	87/50/28	40	5
Domains[#]	12 categories	?	2	4	5
	Nutrition	1. According to the doctors I've seen, my health is now	physical functioning	Functional capacity in everyday life (11)	walking around
	Physical health	2. I seem to get sick a little easier than other people.	psychosocial functioning and well-being	psychological well-being (11)	looking after oneself
	sleep	3. I feel better now than I ever have before.		physical state (9)	usual activities
	dental health	4. I will probably be sick a lot in the future.		social relationships (9)	pain or discomfort
	friends	5. Most people get sick a little easier than I do.			worried, sad or unhappy
	healthiness	6. I am somewhat ill.			
	family	7. In the future, I expect to have better health than other people I know.			
	play	8. I'm not as healthy now as I used to be.			
	activity and exercise	9. My body seems to fight off illness very well.			

	personal grooming emotional nonspecific	10. I'm as healthy as anybody I know. 11. I think my health will be worse in the future than it is now. 12. My health is excellent. 13. I expect to have a very healthy life. 14. I have been feeling bad lately. 15. When there is something going around, I usually catch it. 16. Doctors say that I am now in poor health. 17. I feel about as good now as I ever			
Purpose	health self-construct measure based on a continuum scale (negative to positive) as opposed to dichotomous scale (healthy or relatively ill), so that illness was a composite score of one's overall health and not necessarily the opposite of health.	develop and validate a child's self-rating general health using a multiitem instrument	to develop a first norm-based interpretation of children's physical and emotional functional status and well-being (driven by 1948 WHO definition of health)	psychometrically developed self-reported measure to assess HRQoL in children	describing and valuing health

Scoring	a high score on the CHSCS indicated a self-concept which was positive, lower score indicating a negative self-concept	?	scales scored using Likert's method of summated ratings. Raw scores for the profile derived by computing algebraic mean for individual answering at least half of the items for a given scale. Scores are then standardized on 0 to 100 continuum, a higher score indicating a better health and well-being. scores (for the parent version) can be summed at the concept level to give profile scores, or at overall dimension level by combining physical and psychosocial score (CHQ Summary Scores). CHQ-CF87 and CHQ-CF45 (2018) are scores at the concept/scale level only.	scored from never (1) to always (5) by per item count of scores, summarised and then transformed to 0-100 scale.	utility function

*CHSCS: Child's Health Self-Concept Scale; CHRS: Children's health ratings scale; CHQ-PF50: Child Health Questionnaire-PF50; [$]developed in 2010 (from EQ-5D-3L); [#]domains in bold

10.22 Appendix 22 Parent/Guardian information sheet for focus group discussion

Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mkulu wakafukufukuyu: Lucky Gift Ngwira, MLW Clin Res Programme & Liverpool Sch of Trop Medicine, P.O. Box 30096, Chichiri, Blantyre 3. MALAWI. Tel: 01876444 Email: Lngwira @mlw.mw

Mawu oyamba

Mukupemphedwa ngati kuli kothekera kuti mwana wanu atenge nawo mbali mu kafukufuku wokhudzana ndi umoyo wabwino. Chonde werengani mosamala uthengawu kapena umene wawerengedwa kwa inu. Chikalata chimodzi mudzapatsidwa kuti musunge. Ngati mukufuna kudziwa zambiri chonde khalani omasuka kutifunsa ife.

Kodi kafukufukuyu ndi wa chani?

Kafukufukuyu akukhudzana ndi mafunso amene amafunsidwa kuti tidziwe mmene moyo wabwino ulili ndinso kudziwa ngati mafunsowa ali oyenelera kwathu kuno. Otenga nawo mbali mu kafukufukuyu tidzawapempha kuti atidziwitse chimene umoyo umatanthauza komanso m'mene akumvera za umoyo wawo.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Nkutheka kuti mwana wanu amapanga maphunziro ake pa imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa mwana wanga?

Tidzafunsa mwana wanu kukhala nawo mu makambitsirano a kagulu ndi ane ena kufuna kudziwa m'mene ana akutha kumvera mafunso amene amafunsidwa kuti tidziwe za moyo wabwino wa ana. Pakali pano, mafunsowa anakonzedwa ku mayiko aku ulaya, ndipo tikufuna chithandizo cha ana kuti tiwunike ngati mafunsowa ali oyenelera kwathu kuno. Mafunso ake ndi wokhudza m'mene moyo wa mwana wanu uliri kuphatikizapo ngati akumva kupweteka kapena ululu ndiponso m 'mene akumvera mthupi.

Mwana wanga adzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambilanazi zidzatenga utali wosaposera ola limodzi.

Chinsinsi ndi kasungidwe ka uthenga

Uthenga womwe mwana wanu ati adzapeleke udzatapidwa ndi makina otapila ndi cholinga choti uthengawu usasowe. Pambuyo pamafunso, uthengawu uzatanthauzidwa kupita kuchilankhulo cha chingelezi ndipo zidzasungidwa mmakina a kompyuta mosagwiritsa ntchito dzina la mwana wanu. Uthengawu adzatha kuufikila ndi anzathu okha ogwira ntchito mukafukufukuyu. Uthenga wotapidwawu tidzausunga kufikira pamene tidzautembenuze kukhala m'mapepala. Zotembenuzidwa m'mapepalazi zizasungidwa kwa zana zisanu zokha, koma ngati mukuvomereza tidzasunga zolembedwa pa kompyuta kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:
Kafufukuyu si wamankhwala operekedwa kuchipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wa mwana wanu. Komabe zitha kutheka kuti mwina pamakambitsirano athu mwana atha kutopa kapena kukhumudwa, ndipo titakhala kuti tawona zimenzi tizasiira pompo macheza athu, ndi kuwafotokozera aphunzitsi kapena inuyo.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?
Tidzakupemphani kuti musayine kapena chidindo cha chala chikalata chopereka chilolezo kuti mwana wanu alowe nawo mukafukufukuyi. Mwana wanu naye adzafunsidwa ngati ali okondweretsedwa kulowa nawo mukafukufuku ndipo atapereka chilolezocho, adzalowa mumakambitsirano a pa gulu kapena a munthu payekha.

Malipiro
Kutenga mbali mukafukufuku simuzalipila china chilichonse. Sitidzakulipilani kupatula kupatsa mwana wanu kope ndi cholembera chifukwa chotenga nawo mbali mukafukufukuyu.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?
Zopezeka za kafukufukuyu zidzatha kupezeka kudzera mmisonkhano yomwe idzakonzedwe ndi a Malawi Liverpool Wellcome Trust. (MLW). Zotsatila zi zidzasindikizidwa nso ndipo zidzatha kupezeka kwa inu ngati mungafunse- Chonde gwiritsani ntchito ma nambala omwe apelekedwa pansi pa tsamba la uthengali kuti mupemphe uthenga wa zotsatira za kafukufukuyu.

Chitachitike ndi chani ngati sindikufuna kuti mwana kutenga nawo mbali mukafukufukuyu?
Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chanu ndinso cha mwana wanu. Maganizo anu azalemekezedwa ndipo izi sizidzakhala ndi vuto pa zasukulu ya mwana wanu, moyo ndi chithandizo chanu kapena mwana wanu cha zaumoyo.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?

Mungathe kutenga mpaka sabata imodzi kupereka chilolezo. Komanso mutafuna kuti mudziwe zambiri, chonde funsani a Lucky G Ngwira pa 0999-076-932. Ngati mungakhale ndi funso lilonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene mwakhudzidwira ndi kafukufuku, chonde masukani polumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.23 Appendix 23 Children information sheet for focus group discussion

2b.C- The VOICE! study Children Information Leaflet (Focus Group Discussion)

UTHENGA KWA ANA (zaka 8 mpaka 12)

Mutu wa Kafukufukuyu: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. ***The VOICE! study***

Mawu oyamba

Mukupemphedwa kulowa nawo mu kafukufuku. Kafukufuku ndi njira imodzi yopezera mayankho ku mafunso amene tilibe mayankho.

Kodi kafukufukuyu ndi wa chiyani?

Kafukufukuyu akufuna kupeza ngati mafunso amene timafunsa munthu akadwala ndi oyenera. Tikufunsa chimene moyo wabwino umathanthauza komanso m'mene ukumvera mthupi mwako.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mu sukulu zimene zili gawo la bungwe la Malawi-Liverpool-Wellcome Trust, pansi pa dipatimenti ya Sayansi Com. Dipatimentiyi ikulimbikitsa ana a mu sukulu zozungulira m'zinda wa Blantyre kutenga maphunziro a sayansi. Ndiwe m'modzi woyimba sukulu mu imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Ngati ungakonde kulowa nawo mu kafukufukuyi,	
Choyamba	**Tidzapempha iwe ndi bambo, mai ako kapena amene umakhala nawo kuti ulembe dzina lako pa fomu kusonyeza kuti ukupereka chilolezo.**
Chachiwiri	**Udzapemphedwa kukhala nawo mu zokambirana za pagulu ndi ana ena.**
Chachitatu	**Aliyense mu gululi adzapatsidwa nambala.**
Chachinayi	**Kenakono tidzakufunsani mafunso.**
Chachisanu	**Iwe ndi ana ena mudzapemphedwa kupereka maganizo anu**

Kodi ndiyenera kunena inde?

Ayi, sichoncho. Zili ndi iwe! Ungonena ngati sukufuna kutenga nawo mbali. Zimenezi sizidzakhudza m'mene ukuchitira mukalasi kapena thandizo lakuchipatala ukadwala.

Kodi ndidzakhala mu kafukufuku kwa nthawi yaitali bwanji?

Zokambiranazi pakati pa iwe ndi wochititsa kafukufuku zidzatenga pakati pa mphindi 30 ndi ola. Ukhoza kupumira nthawi iliyonse. Potenga nawo mbali udzapatsidwa cholembera komanso kope.

Kodi munthu wina azadziwa kuti ndikupanga nawo kafukufukuyu?

Ndi okhawo ali gulu lakafukufuku wathu amene azadziwe. Palibe winanso amene azadziwe chifukwa tidzakupatsa nambala imene udzagwiritse ntchito nthawi ya zokambiranazo.

Chidzachitike nchiyani pa zimene ndidzanene?

Tidzalemba komanso kuika mawu amenewo mu malipoti. Malipotiwa adzagwiritsidwa ntchito ndi ma dokotala, anamwino (ma nesi) komanso ochita kafukufuku. Tikhoza kukupatsa lipotili utafuna. Uthengawu tizausunga zaka zokwana zisanu zokha koma utafuna kuti tizazisunge kwa nthawi yochulukirapo ukhoza kunena.

Ndingapite kwa ndani kuti ndidziwe zambiri za kafukufukuyu?

Utha kutidziwitsa m'masiku asanu ngati ukufuna kutenga nawo mbali m u kafukufukuyu. Ngati uli ndi funso lina lililonse, bambo, mai kapena amene umakhala nawo atha kukuyankha. Ukhozanso kupempha ochitisa kafukufukuyu kuti akuyankhe.

Kuti umve zambiri, funsa mkulu wa kafukufukuyi:
Lucky Gift Ngwira,
Malawi-Liverpool-Wellcome Trust,
P.O. Box 30096, Chichiri, Blantyre 3. MALAWI.
Tel: 01876444
Email: Lngwira @mlw.mw

10.24 Appendix 24 Adolescent information leaflet focus group discussion

UTHENGA KWA ACHINYAMATA (zaka 13 mpaka 17)

Mutu wa kafukufuku: Multi-Attribute Utility Instrument (MAU!) in Children and Adolescents: Adaptation for use in economic evaluation in sub-Saharan Africa. *The VOICE! study*

Mawu oyamba

Tikukupempha ngati kuli kotheka kuti utenge nawo mbali mu kafukufuku.

Kodi kafukufuku ndi chiyani?

Kafukufukuyu ndi njira imozi yopezera mayankho ku mafunso amene tilibe mayankho.

Chifukwa chiyani ndikupemphedwa kutenga nawo mbali?

Kafukufukuyu akuchitika mumasukulu amene ali mu Samala Moyo komanso Sayansi kwa onse imene ikuchititsidwa ndi a Malawi-Liverpool-Wellcome Trust. Nkutheka kuti mwina umapanga maphunziro ako pa imodzi mwa sukuluzi.

Chidzachitike ndi chiyani kwa ine?

Tidzakupempha kukhala nawo mu makambitsirano a kagulu ndi ane ena. Makambitsirano a kagulu amakhala ndi anthu pakatu pa asanu ndi m'modzi kufikira khumi ndi awiri. M'makambitisiranowo tikufuna maganizo ako chimene umoyo umatanthauza, kuphatikizapo umoyo wako ngati ukumva kupweteka kapena ululu ndiponso m 'mene ukumvera mthupi.

Kodi ndidzakhala mu kafukufuku nthawi yaitali bwanji?

Zokambilanazi zidzatenga utali wosaposera ola limodzi.

Chinsinsi ndi kasungidwe ka uthenga

Udzapatsidwa nambala nthawi ya makambitsirano kuti dzina lako litsagwiritsidwe ntchito. Zokambiranazo zidzatapidwa ndi makina otapila ndi cholinga choti uthengawu usasowe. Pambuyo pamafunso, uthengawu uzatanthauzidwa ndi kuika m'mitu ikulu ikulu ndipo zimenezi zidzasungidwa mmakina a kompyuta. Anzathu okha ogwira ntchito mukafukufukuyu ndi amene adzatha kuufikila uthengawu. Uthenga wotapidwawu tidzausunga kufikira pamene tidzautembenuze kukhala m'mapepala. Zotembenuzidwa

m’mapepala zizasungidwa kwa zaka zisanu, koma ngati ukuvomereza titha kudzasunga kuti akafukufuku ena akhale nazo.

Chiwopsezo kapena zina zovuta:
Kafufukuyu si wamankhwala operekedwa kuzipatala ayi ndiye alibe chiwopsezo china chilichonse pa moyo wako. Komabe, timadziwa kuti pamakambitsirano ena atha kutopa kapena kukhumudwa. Utafuna kusiya zokambiranazo utha kutero nthawi iliyonse.

Kodi ndidzakhuzidwa bwanji ndi kafukufukuyu?
Tidzakupempha kuti usayine kapena chidindo cha chala chopereka chilolezo kuti ukuvomereza kulowa nawo mukafukufukuyi. Makolo ako nawo adzafunsidwa ngati akuvomereza iweyo kulowa nawo mukafukufuku.

Malipiro
Kutenga mbali mukafukufuku suzalipilidwa china chilichonse. Komabe, udzapatsidwa kope ndi cholembera.

Kodi ndizadziwa zonse zotsatira zokhudza kafukufukuyu?
Zopezeka za kafukufukuyu zidzasindikizidwa mu malipoti amene madokotala, anamwino (ma nesi) ndi ena ogwira ntchito zakafukufuku amagwiritsa. Zimenzi utha kuyitanitsa utafuna kukhala nazo pogwiritsira ntchito ma nambala amene ali m’musimu.

Chitachitike ndi chani ngati sindikufuna kutenga nawo mbali mukafukufukuyu?
Kutenga nawo mbali mukafukufukuyu chidzakhala chifuno chako. Maganizo ako azalemekezedwa ndipo izi sizidzakhala ndi vuto pa zasukulu yako, moyo ndi chithandizo chako.

Ndingapite kwa ndani ngati ngati ndikufuna kudziwa zambiri?
Ungathe kutenga mpaka sabata imodzi kupereka chilolezo chofuna kulowa nawo mukafukufuku. Komanso utafuna kuti mdziwe zambiri zokhudzana ndi kafukufukuyi, chonde funsa a Lucky G Ngwira pa 0999-076-932. Ngati ungakhale ndi funso lina lililonse lokhudza ufulu wa anthu otenga nawo mbali mu kafukufuku, kapena zokhudza mmene wakhudzidwira ndi kafukufuku, utha kulumikizana ndi a bungwe loona za ufulu wa anthu mmakafukufuku la sukulu ya ukachenjede la College of Medicine Research Ethics Committee Secretariat pa nambala iyi: 01-877-245 kapena 01-877-291.

10.25 Appendix 25 Focus group discussion topic guide

Principal Investigator: Lucky Gift Ngwira, Malawi-Liverpool-Wellcome Trust Clinical Research Programme & Liverpool School of Tropical Medicine, P.O. Box 30096, Chichiri, Blantyre 3. MALAWI. Tel: 01876444 Email: Lngwira @mlw.mw

PART 1: FGD topic guide:

Opening statements:
Thank you for taking time to speak with me. As you may remember, my name is _____. We have asked you to participate in this exercise because we are studying the understanding of health and health-related constructs for children and adolescents in Malawi. For people to accurately state about their quality of life, they have to explain the aspects to life that matters to them. We are aiming to develop aspects of health that matters to you. As discussed during the consent discussion, we will ask your views and opinions on what these could be. This process is likely to take about one hour to complete. Feel free to let me know if you need a break at any time. You can also stop the interview if you do not want to continue the discussion. Before we begin do you have any questions?

Methodology:

i. Start by having ice-breakers (10min):
   1) For ex. know the other person (name, where they stay, class, career goal)
   2) Then introduce the other person
   3) Now that we know each other, we move on to our topic for today

ii. **Concept 1: health**
   1) How do you understand health? What is health?
   2) What do you think about moyo?
   3) What do you understand about thanzi?
   4) Is thanzi +moyo=umoyo?
   5) What about wellbeing?
      i. Is it the same or different?
   6) How about wellness?
      i. Is it the dame or different?
   7) Are there any other words?
   8) Do you/we use these words a lot?
   9) Can you give me an example when you used this word? —*teasing out context*

iii. **Concept 2: Thinking about health (umoyo) and whether it does has different parts**

1) Use physical building block task to develop this one
    i. Put down the first building block which is umoyo-
    ii. Ask the children what would be the next building block? –let them build on the first one
        - After they have, ask why?
    iii. Have you seen someone experience this aspect of health?
    iv. What did you observe with them? – *this is to get incidents and determine if they understand what they're saying*
    v. How does this affect your school attendance?
    vi. How did having this aspect of health make you feel?
2) Do the same exercise as 1) above for thanzi and moyo

iv. **Concept 3: Importance of dimensions relative to each other**

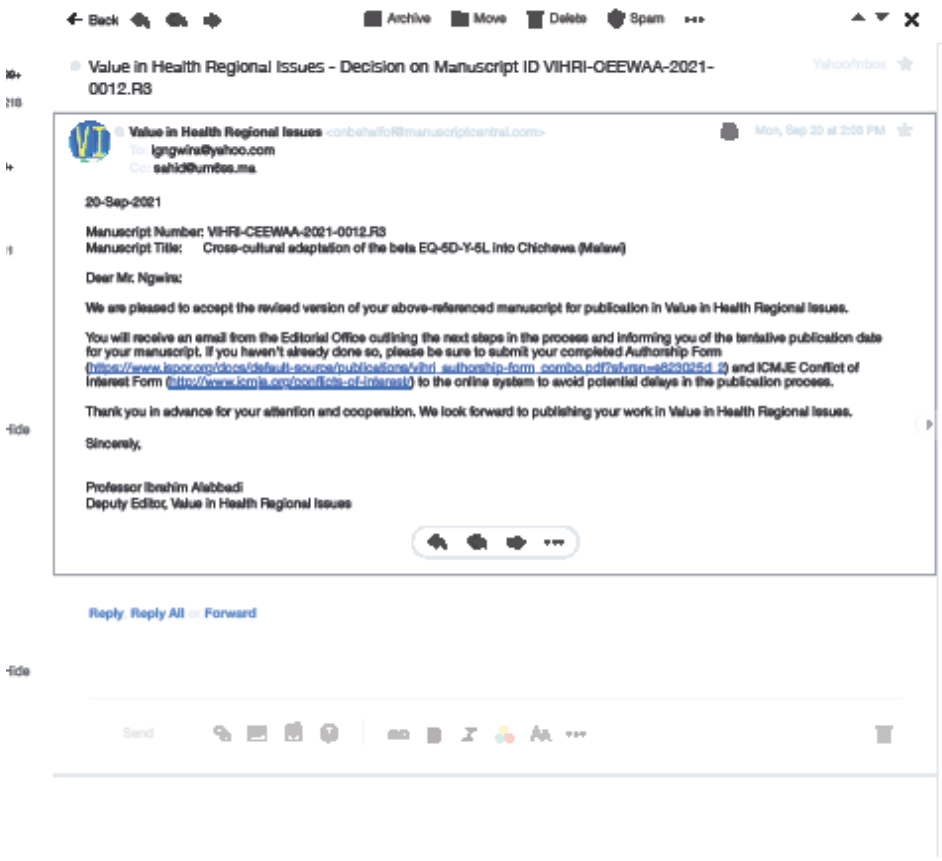
1) this particular dimension of health that you mentioned, how important is that to you as a child or an adolescent?
2) Allow all to express themselves
3) How about this other aspect – would you consider it more important? etc

v. **END: summary wrap-up**

1) Reflect back on what has been discussed
    i. The main things I heard are X, Y, Z- is that true?

10.26 List of publications

10.26.1 Cross-cultural adaptation of the beta EQ-5D-Y-5L into Chichewa (Malawi)



Back Archive Move Delete Spam

Value in Health Regional Issues - Decision on Manuscript ID VIHRI-OEEWAA-2021-0012.R3

Value in Health Regional Issues <onbehalfof@manuscriptcentral.com>
lgngwira@yahoo.com
sahid@um6ss.ma

20-Sep-2021

Manuscript Number: VIHRI-CEEWAA-2021-0012.R3
Manuscript Title: Cross-cultural adaptation of the beta EQ-5D-Y-5L into Chichewa (Malawi)

Dear Mr. Ngwira:

We are pleased to accept the revised version of your above-referenced manuscript for publication in Value in Health Regional Issues.

You will receive an email from the Editorial Office outlining the next steps in the process and informing you of the tentative publication date for your manuscript. If you haven't already done so, please be sure to submit your completed Authorship Form (https://www.ispor.org/docs/default-source/publications/vihri_authorship-form_combo.pdf?sfvrsn=e859265d_2) and ICMJE Conflict of Interest Form (http://www.icmje.org/coi_disclosure.pdf) to the online system to avoid potential delays in the publication process.

Thank you in advance for your attention and cooperation. We look forward to publishing your work in Value in Health Regional Issues.

Sincerely,

Professor Ibrahim Alabbadi
Deputy Editor, Value in Health Regional Issues

Reply Reply All Forward

10.26.2 Systematic literature review of preference-based health-related quality of life measures applied and validated for use in childhood and adolescent populations in sub-Saharan Africa

ELSEVIER

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/vhri

A Systematic Literature Review of Preference-Based Health-Related Quality-of-Life Measures Applied and Validated for Use in Childhood and Adolescent Populations in Sub-Saharan Africa

Lucky G. Ngwira, MPH, Kamran Khan, MSc, Hendramoorthy Maheswaran, MBBS, PhD, Linda Sande, MA, Linda Nyondo-Mipando, PhD, Sarah C. Smith, PhD, Stavros Petrou, PhD, Louis Niessen, PhD

ABSTRACT

Objectives: Consideration of health status in children and adolescents now includes broader concepts such as health-related quality-of-life (HRQoL). Globally, there is a need for relevant preference-based HRQoL measures (PBMs) for use in children and adolescents, yet measurement of HRQoL in these groups presents particular challenges. This article systematically reviews the available generic childhood PBMs and their application and cross-cultural validation in sub-Saharan African (sSA).

Methods: A systematic review of published literature from January 1, 1990, to February 8, 2017, was conducted using MEDLINE (through OvidSP), EMBASE (OvidSP), EconLit (EBSCOhost), PsycINFO, Web of Science, and PubMed.

Results: A total of 220 full-text articles were included in a qualitative synthesis. Ten generic childhood PBMs were identified, of which 9 were adapted from adult versions and only 1 was developed specifically for children. None of the measures were originally developed in sSA or other resource-constrained settings. The Health Utilities Index Mark 3 (HUI3) and the EQ-5D-Y were the only measures that had been applied in sSA settings. Further, the HUI3 and the EQ-5D-Y were the only generic childhood PBM that attempted to establish cross-cultural validation in sSA. Five of the 6 of these validation studies were conducted using the EQ-5D-Y in a single country, South Africa.

Conclusions: The findings show that application of generic childhood PBMs in sSA settings has hitherto been limited to the HUI3 and EQ-5D-Y. Most adaptations of existing measures take an absolutist approach, which assumes that measures can be used across cultures. Nevertheless, there is also need to ensure linguistic and conceptual equivalence and undertake validation across a range of sSA cultural contexts.

Keywords: childhood, PBM, HRQoL, validation, sSA.

VALUE IN HEALTH REGIONAL ISSUES. 2021; 25(C):37–47

Introduction

In sub-Saharan Africa (sSA), that is, the region below the Sahara desert, there has been a remarkable decline in mortality rates in infants and under-5-year-olds since 2000,[1] partly driven by policy responses to the Millennium Development Goals.[2] In 2015, the United Nations set 17 sustainable development goals as a replacement for Millennium Development Goals for nations to achieve by 2030. Goal number 3 focuses on good health and well-being, specifically targeting child and maternal mortality and chronic disease reduction.[3] As childhood mortality continues to fall, and increasing numbers of children survive into adolescence and beyond, future interventions will increasingly focus on improving wider outcomes, including children's health-related quality of life (HRQoL)".[4]

Preference-based HRQoL measures (PBMs) are designed to "assess patient preferences across broad areas including symptoms, physical functioning, work and social activities, and mental we being."[5] PBMs have increased in prominence over recent ye because of their use in economic evaluations, which often info policy.[6] This can only be meaningful if PBMs are relevant, valid, a robust. Nevertheless, most generic PBMs have been developed the English language[7] and, as the development of new measures time and resource intensive, use of PBMs in other contexts l mostly relied on translating and/or adapting existing measur within an implicit assumption that adaptation produces an equival measure.[8] Adaptation of an existing measure requires an iterat process of translation and back translation with a qualitative revi to establish both linguistic and conceptual equivalence.[7] In ad tion, the adapted measure needs to be revalidated using stand psychometric methods to reestablish validity and reliability for t new form of the measure in the new context.

Several reviews have assessed the development and appli tion of childhood PBMs and non-PBMs (restricted to ages 6-

10.26.3 A new approach to assessing children's interpretation of severity qualifiers in a multi-attribute utility instrument -the EQ-5D-Y-5L: developing and testing

The Patient - Patient-Centered Outcomes Research
https://doi.org/10.1007/s40271-021-00496-1

ORIGINAL RESEARCH ARTICLE

A New Approach to Assessing Children's Interpretation of Severity Qualifiers in a Multi-Attribute Utility Instrument–The EQ-5D-Y-5L: Development and Testing

Sarah Derrett[1,2] · Mike Herdman[2,3] · Lucky G. Ngwira[4,5] · Elizabeth Yohe Moore[6] · Jennifer Jelsma[2,7]

Accepted: 23 January 2021

Abstract

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

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

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

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

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

1 Introduction

Patient-reported outcome measures (PROMs) assess the health status or health-related quality of life of patients (or individuals from the general population) at, or over, a specified time [1]. PROMs can be condition specific (assessing particular health conditions, e.g. diabetes) or generic (assessing health across a wide range of health conditions and in healthy people). PROMs are used in clinical research, policy-making, and population survey contexts [2–4]. For example, PROMs may be used in clinical trials of new interventions and, in particular population subgroups over time to assess group or population-level health status changes. One type of PROM, multi-attribute utility instruments (MAUIs), are often used in cost-utility analyses (CUAs) to inform policy makers' decisions about the costs and benefits associated with, for example, the introduction of new clinical or pharmaceutical treatments [5]. With MAUIs, preference weights (utilities) are estimated for all combinations of health states measured by the instrument. Estimated utilities are derived using valuation studies, usually undertaken with

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

Extended author information available on the last page of the article

Published online: 02 March 2021

10.26.4 Comparing the performance of the EQ-5D-Y-3L and EQ-5D-Y-5L across age groups

Comparing the performance of the EQ-5D-Y-3L and EQ-5D-Y-5L across age groups

Janine Verstraete*[1], Lucky Gift Ngwira*[2,3], Des Scott[4], Hendramoorthy Maheswaran[5], Stavros Petrou[6], Sarah C. Smith[7], Louis Niessen[3, 8],

*These authors contributed equally as co-first authors

[1] Department of Paediatrics and Child Health, Division of Pulmonology, Cape Town, South Africa. janine.verstraete@uct.ac.za

[2] Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi. lngwira@mlw.mw (Corresponding Author)

[2] Department of clinical sciences, Liverpool School of Tropical Medicince, Liverpool, UK.

Abstract

Objectives: The aim of this study is a head-to-head comparison of the performance of the EQ-5D-Y-3L (Y-3L) and the EQ-5D-Y-5L (Y-5L) across the age groups in children and adolescents from sub-Saharan Africa.

Methods: Children/adolescents aged 8-10 years, 11-12 years and 13-15 years with a range of health conditions including an acute or chronic condition and those from the general population completed the Y-5L, Y-3L. Performance of the Y-5L and Y-3L was determined by comparing feasibility, redistribution of dimension responses, discriminatory power, convergent validity, and test-retest reliability.

Results: