Quality of life assessment in adolescent obesity: Development of a new instrument for economic evaluation

Oyeyemi Abodunde Oluboyede

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds, School of Medicine and Health

May 2013

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others

Acknowledgements

The completion of this work would not have been possible without the support I received from a number of sources.

I wish to express my gratitude to my supervisors Claire Hulme, Andrew Hill, Aki Tsuchiya, Chris McCabe and the thesis advisory group. I could not have done without their knowledge, guidance, and encouragement. A special thanks to David Meads and Stephane Hess for assisting my understanding of Rasch analysis and stated preference experiments. I am especially grateful to my friends and colleagues in the Academic Unit of Health Economics for their unfailing support throughout the thesis. Special thanks also go to Sandy Tubeuf for proof reading this entire thesis.

This work was funded through the National Institute for Health Research doctoral fellowship programme. I am indebted to the individuals who reviewed my proposal and the interview panel for believing in me enough to give me the opportunity to undertake this project. I would like to thank all the individuals who participated in this study. A special thanks to the adolescents who eagerly gave up their time to share their experiences with me. I would like to extend my gratitude to the weight management organisations that provided invaluable help with the recruitment of young people (MEND, More Life, and Watch It). A special thanks to Carolyn, Gloria, Rob, Sarah, and Win who really did make me feel like part of the team.

My family, particularly my husband Dan, have come on the PhD rollercoaster ride with me. I am grateful to you for not running for the hills...and let's just say I would not have blamed you at times! Your love and support has been a vital source of strength and motivation. I can't leave out that pesky dog, Bas, who in his own way was absolutely the perfect PhD companion.

Finally, thanks to the inspirational quote that always managed to get me through the toughest times:

'Everything will be okay in the end. If it's not okay it's not the end'
-Unknown

Abstract

Obesity, particularly in the younger population, is a major public health concern. Because of this, numerous public health initiatives and specialised weight management interventions are available for this population. Economic evaluations of weight management interventions can be utilised to inform resource allocation decision making. But in order to provide fully informed economic evaluations of weight management interventions, a valid and reliable tool to a) measure weight specific quality of life (QoL) and b) assign preference values to different aspects of weight specific QoL is necessary. Currently no existing tools meet these crucial requirements, and this thesis aims to fill this gap.

The thesis utilised a multiple methodological approach in the development of a new weight specific instrument. The development of the tool comprised of four distinct studies. The first was informed by the existing literature and aimed to identify items through a) qualitative interviews with adolescents and b) discussion with specialists in the field of adolescent obesity. A long list of potential items for inclusion in the new instrument was crafted in the first study. The aim of the second study was the identification of a reduced item set by performing psychometric assessments and Rasch analysis. At the end of the first two studies the new instrument was created. The third study aimed to assess its measurement properties through psychometric analyses. The final feasibility valuation study addressed the derivation of preference values for the states described by the instrument.

The results of the empirical studies taken together demonstrated that it is feasible to identify the impact of weight status on QoL using adolescents' views. It has been possible to create the **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (**WAItE**), consisting of seven items each associated with five response options. It was not possible to obtain a scoring algorithm for the states described by the WAItE when three variants of discrete choice experiments were implemented. The WAItE can be used to investigate the benefits associated with alternative weight management interventions.

Table of Contents

List of Tables

List of Figures

Abbreviations

BAROS	Bariatric Analysis and Reporting Outcome System
BMI	Body Mass Index
CFI	Comparative Fit Index
CFA	Confirmatory Factor Analysis
CHIP	Child Health Illness Profile
CHQ	Child Health Questionnaire
CHU-9D	Child Health Questionnaire-9 Dimensions
CUA	Cost Utility Analysis
DCE	Discrete Choice Experiment
DfE	Department for Education
DH	Department of Health
DIF	Differential-Item Functioning
EFA	Exploratory Factor Analysis
EQ-5D	EuroQol-5D questionnaire (5 dimensions with 3 level response option)
EQ-5D-Y	EuroQol-5D questionnaire – Youth version
FDA	Food and Drug Administration
HSE	Health Survey for England
HRQoL	Health Related Quality of Life
HTA	Health Technology Assessment
HUI	Health Utilities Index
IOTF	International Obesity Task Force
IWQoL-Kids	Impact of Weight on Quality of Life-Kids
IWQoL-Lite	Impact of Weight on Quality of Life-Lite
LLDS	Long list descriptive system
MNL	Multinomial logit model
NAO	National Audit Office
NCMP	National Child Measurement Programme
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NOO	National Obesity Observatory
OPOI	Oxford Paediatric Obesity Instrument
PAF	Principal Axis Factors
PSI	Person Separation Index
PROMs	Patient Reported Outcome Measures
QALY	Quality Adjusted Life Year
QoL	Quality of Life
RMSEA	Root Mean Square Error of Approximation
RUT	Random Utility Theory
RUM	Random Utility Model
SF-36	Short Form-36
SG	Standard Gamble
SP	Stated Preference
SD	Standard deviation
TTO	Time Trade Off
TLI	Tucker Lewis Index
WAItE	Weight-specific Adolescent Instrument for Economic-evaluation
WHO	World Health Organisation
UK90	The British 1990 Growth Reference
VAS	Visual Analogue Scale
YQOL-W	Youth Quality of Life Instrument-Weight module

Chapter 1
Overview of the thesis

Obesity and overweight have been described as an escalating global epidemic by the World Health Organisation (WHO, 2012, 2013a). In adults, obesity is defined as a Body Mass Index (BMI) of 30 or more and overweight individuals have a BMI of between 25 and 30. For children and adolescents under the age of 19 years, growth reference charts are used to define weight status (NOO, 2009, 2011a). Due to the long term health risks associated with childhood obesity, including respiratory diseases, type II diabetes and the emotional and psychological effects of being obese, public health initiatives and specialised weight management interventions have been developed specifically for the adolescent population in the UK. These include community-based interventions that have the potential to address both health and lifestyle behaviours, as well as surgical or pharmaceutical interventions.

In order to provide a rigorous assessment of different weight management interventions, decision makers take into consideration data derived from economic evaluations. The recently updated *Guide to the methods of technology appraisal* provided by the National Institute for Health and Care Excellence (NICE, 2013), specifies that the evaluation of cost effectiveness of competing technologies should be translated into quality-adjusted life years (QALYs - an index of survival that takes into account a) life expectancy and b) quality of life during this time) (NICE, 2013). The translation of benefits derived from weight management interventions using existing generic and weight specific instruments into valid and reliable QALY estimates has been limited thus far. A key reason for this is that the QALY estimates cannot be generated from existing weight specific instruments as they are not preference-based. Additionally, evidence shows that not all dimensions of QoL are equally important to overweight and obese people (Hauber *et al*., 2010) therefore, the use of non-preference based instruments in the evaluation of weight management interventions would be inaccurate. Moreover it has been suggested that generic instruments are not as sensitive to changes in BMI as weight specific instruments, thus existing generic preference based instruments may be missing important domains of quality of life (QoL) affected by weight status (Brazier *et al*., 2004). There are few existing weight specific instruments that are appropriate for the younger population and none provide the crucial information required for the calculation of QALYs (i.e. utility or preference values).

This thesis addresses the measurement and valuation of weight specific QoL in the adolescent population. The empirical studies within the thesis describe the development of a new weight specific instrument suitable for the measurement of benefits derived from weight management interventions.

1.1 Aims and objectives of the thesis

Aim:

To construct a weight specific QoL instrument suitable for the elicitation of preference values for the calculation of QALYs within the 11 to 18 age group.

Objectives:

1. To utilise information from: a) the existing literature; b) qualitative interviews with adolescents; and c) consultation with specialists in the field of adolescent obesity; to identify the aspects of QoL affected by weight status in order to create a weight specific long list descriptive system (LLDS)
2. To utilise quantitative methods to identify a reduced item set, from the LLDS, generating a new reduced form weight specific descriptive system appropriate for preference elicitation and undertake preliminary testing of the measurement properties of the finalised instrument
3. To undertake a methodological study, facilitating the derivation of preference values, informed by members of the general public utilising a subset of states defined by the new weight specific descriptive system
4. To use econometric modelling to derive preference weights for all feasible states from the descriptive system

1.2 Structure of the thesis

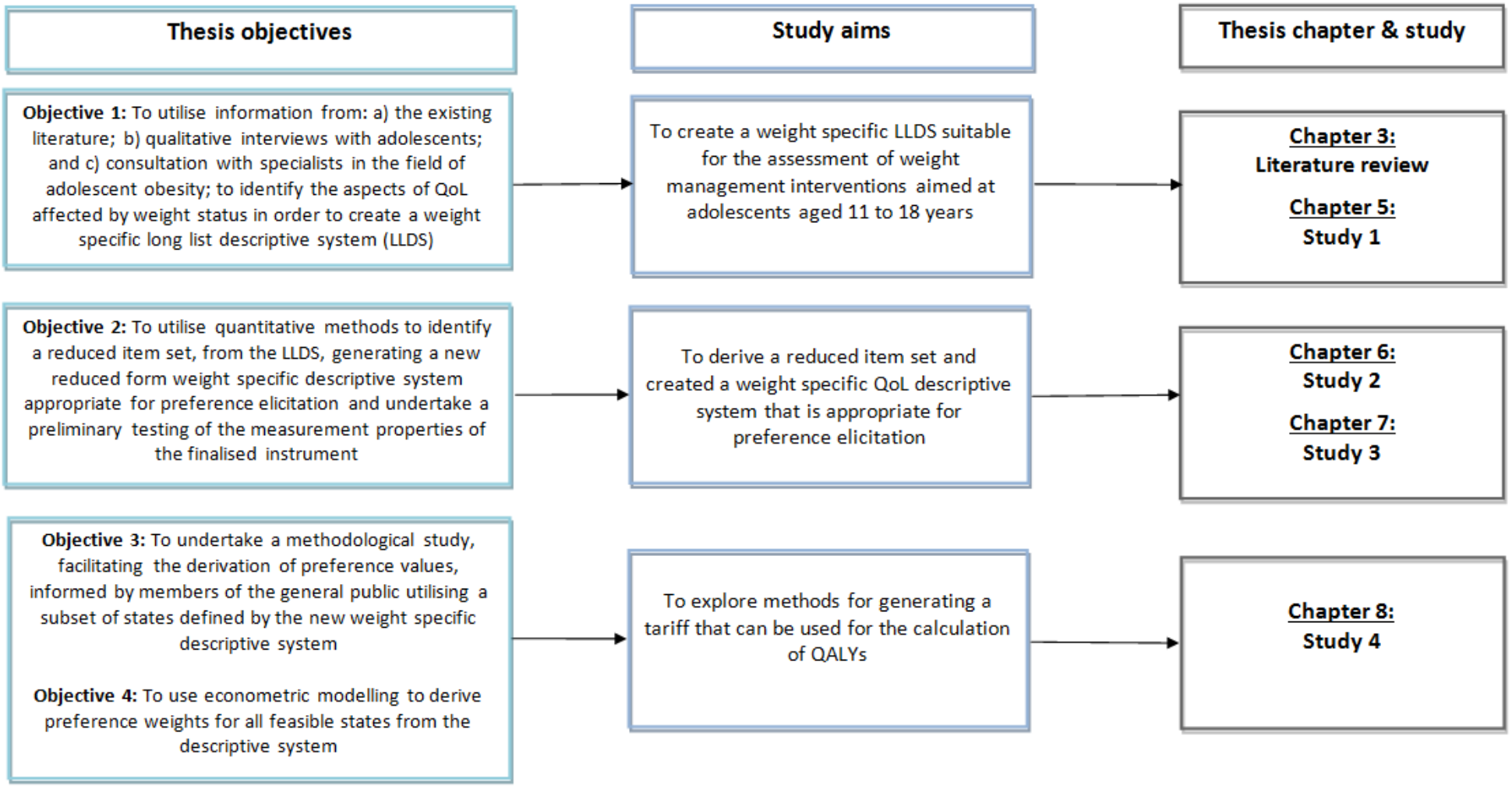
Chapter 2 provides the context for this research. Clinical descriptions of obese and overweight categories are given and the prevalence of the problem of obesity including the increasing numbers of overweight individuals who may be at risk of becoming obese. The chapter includes a discussion of the problems that are linked with obesity and the associated implications for the National Health Service (NHS) in the UK. The role of economic evaluation in the allocation of scarce resources is presented, plus the implications of measurement and valuation of benefits derived from weight management interventions. The final section of the chapter provides discussion regarding the significance of carrying out fully informed economic evaluations of weight management interventions and why a new preference based weight specific measure is needed for the adolescent population. Understanding the state of play in relation to obesity in general, and specifically in the adolescent population, and in the context of resource allocation sets the foundation for this research. The next chapter builds on the findings of the current literature.

A literature review of instruments used in the obese and overweight adolescent population is presented in Chapter 3. The chapter utilises two existing reviews to provide an overview of the impact of above normal weight status on the lives of adolescents. It details existing generic and weight specific QoL measures that have been used with obese and overweight adolescents. Instruments are identified from the literature review and are assessed for their measurement properties and suitability for carrying out economic evaluation of weight management interventions. The chapter highlights the limitations of the existing literature, in providing valid and reliable QALY calculations particularly for the adolescent population. Consequently, this curtails the ability to carry out fully informed cost utility analysis (CUA) of weight management interventions.

Before presenting the work from each of the empirical studies undertaken in the thesis, Chapter 4 presents key steps involved in the development of a preference based measure and the associated methodological challenges. The methodology, analysis, and results of the empirical studies are then presented in the following chapters. Information gathered from Chapters 3 and 4 informed the design of the empirical studies which undertook the creation of a new weight specific measure suited to the adolescent population.

Chapter 5 describes the qualitative methodology used for data collection and analysis that resulted in the creation of the weight specific LLDS. Chapter 6 details the second empirical study in which *item reduction* was carried out. In line with constraints of preference based measures the weight specific LLDS was reduced in order to make it appropriate for the elicitation of preference values. Preliminary assessments of the measurement properties of the new instrument were undertaken and are presented in Chapter 7. In order to ensure that the descriptive system is suitable for CUA of weight management interventions, preference values need to be elicited for all states outlined by the system. The final empirical valuation study presented in Chapter 8 assessed the application of discrete choice experiment (DCE) methods in facilitating the derivation of utility values for QoL states described by the new descriptive system. The chapter provides a description of the design and modelling processes which were undertaken. **Figure 1.1** provides a summary flowchart linking each of the research objectives with the write up of the literature review and the empirical studies contained in the relevant thesis chapters. Chapter 9 provides a critical assessment of the empirical work undertaken within the thesis providing a discussion regarding the strengths and limitations of the work and the implications of these on the research findings. Consideration is also given to the policy implications if extensive adoption the new measure is undertaken and future research that could be carried out in the area.

Figure 1:1 Overview of the thesis



Chapter 2
Background

2.1 Introduction

Worldwide obesity has more than doubled since 1980. Overweight and obesity are the fifth leading risk for global deaths (WHO, 2012, 2013a). The prevalence of obesity in childhood and adolescence has been rising over the last three decades, particularly in first world countries (Ogden *et al*., 2012). The WHO recently estimated the prevalence of overweight and obesity (BMI (kg/m^2) ≥ 25), in the UK for individuals aged 15 years and over, reporting rates of 67.8% and 63.8% for obese males (BMI ≥ 30 = 23.7%) and females (BMI ≥ 30 = 26.3%) respectively (WHO, 2013b). Recent estimates, using data from the Health Survey for England (HSE)[1], suggest that around three in ten boys and girls aged 2 to 15 were classed as either overweight or obese in 2011 (31% and 28% respectively) (NOO, 2013). Despite some evidence for a deceleration of rising obesity figures, in some high-income countries, there are still historically high rates of obesity (Wang *et al*., 2011).

2.2 The link between childhood and adulthood obesity

Childhood obesity tracks into obesity in adulthood. Predictors of obesity in adulthood include weight gain that occurs between the ages of 2 and 5 years and being overweight by 8 years of age (Freedman *et al*., 2005 and Rowlinson, 2011). The high prevalence of obesity on population health is far-reaching; societies are burdened by premature mortality, co-morbidity associated with many chronic disorders, and negative effects on HRQoL. These negative consequences have both immediate and long-term health consequences as overweight and obese children and adolescents are more likely to become obese adults, and have a higher risk of morbidity, disability, and premature mortality in adulthood (Dietz, 1998).

The thesis focuses on the adolescent population because of: a) the high prevalence of obesity in childhood and adolescence, b) the many adverse health effects associated with obesity in the younger population and c) the fact that overweight children are more likely to grow up to be overweight adults.

[1] The HSE, originally set-up in 1991, is an annual cross-sectional survey utilising a representative population sample of approximately 16,000 adults and 4,000 children living in England. Socioeconomic data and information on health and health-related behaviours and measures height and weight are collected within the survey.

2.3 QoL and health implications of obesity

There has been a strong emphasis on the impact of obesity and overweight on the younger population, due to its significant consequences on health in both the short and longer term (NOO, 2011b). Childhood obesity can lead to life-threatening conditions including: diabetes, poor pulmonary function, heart disease, sleep *apnoea*, cancer; and other disorders such as liver disease, advanced growth, early maturity, polycystic ovary disease, orthopaedic complications, eating disorders (anorexia and bulimia), skin infections, asthma, and other respiratory problems (Dietz 1998, Lobstein *et al*., 2004, Luttikhuis *et al*., 2009). Furthermore, overweight children and adolescents are known to become targets of stigma and discrimination (Dietz 1998). At present, there is a growing interest in emotional health and well-being in obesity. Studies have also shown that increased weight status above the healthy range is likely to have a significantly negative impact on psychological physical and social dimensions of QoL (Lobstein *et al*., 2004, NOO 2011b, Tsiros *et al*., 2009 and Griffiths *et al*., 2010). In one study, severely obese children rated their QoL as low as children with cancer on chemotherapy (Schwimmer *et al*., 2003). **Table 2.1** summarises the effect of obesity on different aspects of health and wellbeing.

Table 2:1 Summary of different conditions associated with obesity in young people

Systems of the body	Conditions
Blood system (cardiovascular or circulatory system)	• Heart disease • Poor pulmonary function
Endocrine system (provides communication within the body using hormones made by endocrine glands)	• Advanced growth • Diabetes • Early maturity
Digestive system (specifically the gastrointestinal tract including the stomach and intestine)	• Liver disease
Musculoskeletal system	• Orthopaedic complications
Reproductive system	• Polycystic ovary disease
Respiratory system	• Asthma • Sleep *apnoea*
Integumentary system (the organ system that protects the body from various kinds of damage such as the skin, hair and nails)	• Skin infections
Other problems	• Anxiety • Depression • Increased risk of eating disorders such as anorexia and bulimia • Increased risk of coronary heart disease and cancer later in life • Poor quality of life • Poor self-esteem • Stigma, teasing, marginalisation, distorted peer relationships

2.4 Governmental focus on obesity in young people

Concerns for the negative consequences of childhood obesity have led to the development of numerous obesity prevention initiatives by public sector organisations. In the UK the National Child Measurement Programme (NCMP), established in 2006, is an important part of the Government's work programme on childhood and adolescent obesity, and is operated jointly by the Department of Health (DH) and the Department for Education (DfE) (NOO, 2013). A recent review by Oyebode & Mindell (2013) illustrated the close link between obesity policy making and monitoring in England and data collected from nationally surveys, such as the HSE. This shows the UK government's motivation to address childhood and adolescent obesity and, by implication, prevent future adult obesity.

2.5 Cost burden of obesity

Several chronic and acute health conditions are associated with above normal weight status, which not only have a negative effect on the QoL of the individuals, and lead to increased health care costs and lost productivity (Wang *et al*., 2011). Obesity currently has significant cost implications for the UK National Health Service (NHS). A current estimate of the direct costs of obesity (including costs incurred by excess use of ambulatory care, hospitalisation, drugs, radiological or laboratory tests, and long term care) and associated costs of co-morbidities to the NHS was £4.2 billion in 2007 and it was forecasted that the continuing rise in obesity will add £5·5 billion in medical costs to the NHS by 2050 (Butland *et al*., 2007). An estimate of £15.8 billion was calculated for the wider economic costs of obesity (Butland *et al*., 2007), based on indirect costs associated with factors such as: decreased years of disability-free life, increased mortality before retirement, early retirement, disability pensions, and work absenteeism or loss of productivity for example (Wang *et al*., 2011).

2.6 Measurement of obesity in young people

In adults, obesity is defined by a Body Mass Index (BMI) of 30 or more and overweight individuals have a BMI between 25 and 30 (WHO, 2013a). BMI is calculated by taking a person's weight in kilograms (kg) and dividing this by the square of their height in meters (m). The identification of obesity and overweight in young people aged 18 years and under is more complicated than for adults because, as they grow, the height of children and adolescents changes and thus there is a dynamic relationship between weight and height until adult height is achieved. This makes it more difficult to categorise the weight status of children and adolescents. Therefore, to work out whether the BMI of children and adolescents is too high

or too low, both the age and sex of the individual need to be taken into account as BMI changes considerably between birth and adulthood (NOO 2009). As a result of this, a growth reference chart is used to define the weight status of individuals aged 18 years and under. The raw BMI score is compared with thresholds that vary according to age and gender which are usually derived from a reference population (a child growth reference). BMI thresholds are frequently defined in terms of a specific BMI z-score / BMI-Standard Deviation Score (SDS), or centile, on a child growth reference. A BMI z-score or BMI-SDS indicates how many units (of the standard deviation) a child's BMI is above or below the average BMI value for their age group and sex[2]. Once a child's BMI centile or z score has been calculated, this figure can then be checked to see whether it is above or below the defined thresholds for the child growth reference used (NOO, 2011a).

In England, the British 1990 (UK90) growth reference is the most commonly used reference in the UK for population monitoring and clinical assessment in young people (aged 4 to 18 years) (Cole *et al*., 1995, Cole *et al*., 2007 and Cole *et al*., 1998). The classification of weight status is based on population level data and is produced by calculating the percentage of boys and girls who are in the 2nd (underweight), 85th (overweight), and 95th (obese) centiles[3]. In other countries, more frequently used growth references include: the International Obesity Task Force (IOTF) thresholds (Cole *et al*., 2000), the World Health Organization Growth Reference and the Centre for Disease Control Growth Reference. As the UK90 is recommended for use in England, it will be used in the definition of weight status for adolescents, henceforth, in combination with the population monitoring cut-off points of BMI. Other measurements such as waist circumference and skin thickness can be collected to indicate a person's weight status; however, none of these are as widely used as BMI, additionally, equivalent growth references do not exist for these other measures (NOO, 2009).

[2] For instance, a z score of 1.5 indicates that a child is 1.5 standard deviations above the average value, and a z-score of –1.5 indicates a child is 1.5 standard deviations below the average value (NOO 2011a (p. 3)).

[3] There are also clinical cut-offs utilised in the clinical setting (i.e. in a hospital setting for example). Here overweight and obese weight status is identified as those individuals falling in the 91st and 98th centiles.

2.7 Management of obesity in young people

Despite the overwhelming evidence showing the need to reduce obesity, no clear consensus on effective management policy or programme based strategies has been reached. The changes needed to reverse the obesity epidemic are likely to require various long term interventions targeting: individual behaviour change through interventions in schools, homes, and workplaces and; sector change within agriculture, food services, education, transportation, and urban planning (Steven *et al*., 2011). In 2006 NICE developed the first national guidance on the prevention, identification, assessment and management of overweight and obesity in adults and children in England and Wales (NICE, 2006). The guidance provided recommendations on the clinical management of overweight and obesity in the NHS, and advice on the prevention of overweight and obesity. **Table 2.2** provides a summary of the clinical interventions that were recommended, for the management of overweight and obesity in the younger population, in this guidance document. In 2011 the clinical guideline was reviewed; however, there were no changes to the original recommendations for management and prevention (NICE, 2011).

Table 2:2 Interventions for prevention and management of obesity in young people*

Interventions	Details
Lifestyle interventions: Should be multicomponent interventions provided by Weight management programmes	• Weight management programmes should include behaviour change strategies to increase young people's physical activity levels or decrease inactivity, improve eating behaviour, quality and reduce energy intake. • The aim of weight management programmes for children and young people may be either weight maintenance or weight loss, depending on their age and stage of growth • Parents of overweight or obese children and young people should be encouraged to lose weight if they are also overweight or obese
Behavioural interventions: Should be delivered with the support of an appropriately trained professional	• These should include strategies, that are appropriate for the young people (e.g. stimulus control, self monitoring, goal setting, rewards for reaching goals and problem solving) • Parents should be encouraged to give praise and to role-model desired behaviours
Physical activity: Young people should be encouraged to increase their physical activity even if they do not lose weight as a result	• Young people should be encouraged to do at least 60 minutes of moderate activity each day • Those who are already overweight may need to do more than 60 minutes' activity • Sedentary behaviours, such as sitting watching television, using a computer or playing video games should be reduced • Support and encouragement to do more exercise or structured physical activity in the day. The choice of activity should be appropriate to the ability and confidence of the individual
Dietary advice: Dietary changes should be individualised, tailored to food preferences	• A dietary approach alone is not recommended for young people and it is essential that any dietary recommendations are part of a multicomponent intervention • Any dietary changes should be age appropriate and consistent with healthy eating advice

and allow for flexible approaches to reducing calorie intake	• For overweight and obese children and adolescents, total energy intake should be below their energy expenditure and changes should be sustainable
Pharmacological interventions: Drug treatment should be considered only after dietary, exercise and behavioural approaches have been started and evaluated	• Drug treatment is not generally recommended for individuals younger than 12 years • In children younger than 12 years, drug treatment may be used only in exceptional circumstances, if severe life-threatening comorbidities (such as sleep *apnoea* or raised intracranial pressure) are present. Prescribing should be started and monitored only in specialist paediatric settings • In children aged 12 years and older, treatment with orlistat is recommended only if physical comorbidities (such as orthopaedic problems or sleep *apnoea*) or severe psychological comorbidities are present. Treatment should be started in a specialist paediatric setting, by multidisciplinary teams with experience of prescribing in this age group
Surgical interventions: Only available for young people with morbid obesity	• Surgical intervention is not generally recommended in children or young people • Bariatric surgery may be considered for young people only in exceptional circumstances, and if they have achieved or nearly achieved physiological maturity • Surgery for obesity should be undertaken only by a multidisciplinary team that have specialist paediatric expertise • All young people should have had a comprehensive psychological, education, family, and social assessment before undergoing bariatric surgery. A full medical evaluation including genetic screening or assessment should be made before surgery to exclude rare, treatable causes of the obesity

*Informed by the NICE 2006 clinical guideline document (NICE, 2006)

The evidence base for obesity research is growing with the development of databases and reviews of randomised controlled trials of preventive and treatment interventions (Steven *et al*., 2011). A recent Cochrane review investigated interventions for treating obesity in individuals aged 18 years and under (Luttikhuis *et al*., 2009). There were 64 studies that were examined in the review including: 54 studies on lifestyle treatments (with a focus on diet, physical activity, or behaviour change) and 10 studies on drug treatment to help overweight and obese children and their families with weight control (no surgical treatment studies met the inclusion criteria for inclusion in the review). Findings showed that lifestyle programs can reduce the level of overweight in child and adolescent obesity at six and twelve months after the beginning of the program. In moderate to severely obese adolescents, a reduction in overweight was found when either the drug orlistat or the drug sibutramine[4] were given in addition to a lifestyle program. The authors concluded that there was limited information on the long-term outcome of obesity treatment in children and adolescents and there was a need for this to be examined in high quality studies (Luttikhuis *et al*., 2009). Findings from a more recent review, of the treatment of childhood obesity, reiterated the benefits of lifestyle

[4] NICE recommendations for the use of sibutramine have been withdrawn since the publication of the 2006 guidance (NICE, 2006).

programs comprising family based treatment interventions that incorporate physical activity, diet and psychological components (Staniford *et al*., 2013).

2.8 Outcomes and economic evaluation of obesity

It is increasingly recognised that traditional bio-medically defined outcomes such as clinical and laboratory measures need to be complemented by measures that focus on the patient's concerns in order to evaluate interventions and identify more appropriate forms of health care (Fitzpatrick *et al*., 1998). Interest in Patient Reported Outcome Measures (PROMs) has been fuelled by the increased prevalence of chronic conditions such as obesity, where the objectives of interventions are to stop, minimise, or reverse decline in physical or psychological functioning, for example. Thus, PROMs have been designed to measure the impact of health interventions on different aspects of a given individual's life. A range of tools have emerged assessing QoL from the patient's perspective including: questionnaires, interview schedules, and rating scales. They have in common the objective of assessing health status and illness. At present there are a number of terms used to describe different PROMs. The full spectrum of measures described below will collectively be referred to as *instruments*.

A summary of the most commonly used types of instruments to assess the health status of individuals is listed below. They differ in content and also in their primary intended purpose. Some instruments have characteristics of more than one category or have evolved over time in their intended uses. Definitions have been taken from Fitzpatrick *et al*. (1998; p.3).

- *Disease-specific* – These have been developed in order to provide the patient's perception of a specific disease or health problem such as asthma or obesity
- *Dimension-specific* – Instruments that assess one specific aspect of health status. The most common type of dimension-specific measure is one that evaluates aspects of psychological well-being (e.g. anxiety & depression)
- *Generic* – Instruments of this nature are intended to capture a very broad range of aspects of health status and the consequences of illness, and therefore to be relevant to a wide range of patient groups. The content of such questionnaires has been deliberately designed to be widely appropriate
- *Summary items* – Instruments that invite respondents to summarise diverse aspects of their health status by means of one or a number of questions (within a particular dimension). A popular summary item asks individuals how they perceive their health:

'How is your health in general? Would you say it is 'very good', 'good', 'fair', 'bad', 'very bad'?'

- *Preference based* - This form of instrument uses preference-based methods to value QoL health states. Health states consist of a descriptive system made up of dimensions, items within dimensions and response categories for each item. Health states are valued by individuals using preference elicitation techniques to derive scores for dimensions of health status based on summing responses to questionnaire items, with the possibility of dimension scores being summed in turn

The focus of this thesis is to create a new *disease-specific* instrument. The primary intended purpose of the instrument is the measurement of benefits derived from weight management interventions and the utilisation of *preference based* techniques in the valuation of states described by the instrument.

Policy makers are increasingly asking not only whether an intervention works, but also evidence to assess benefits in relation to costs of health care so that better use is made of limited resources (NICE, 2013). Evidence of the benefit from interventions, as perceived by patients, carers, health care professionals and by society as a whole, is therefore needed. There is little evidence on the cost effectiveness of weight management interventions aimed at the younger population, partly because of a lack of suitable outcome measures that are amenable to health economic evaluations (NICE, 2013). Some studies have conducted reviews to evaluate the cost effectiveness of weight management interventions, such as: bariatric surgery, pharmacotherapy and multi-component (Castaneda-Gonzalez *et al*., 2010, Hussain & Bloom, 2011, Klarenbach *et al*., 2010, Lehnert *et al*., 2012, O'Meara *et al*., 2002, Padwal *et al*., 2011, Picot *et al*., 2012, Salem *et al*., 2005). The relevance of the findings from the studies included in the reviews, to children and adolescents living in the UK, are limited because the study populations: a) were mainly from a non-UK setting and/or b) were predominantly composed of adults aged 18 years plus. Of the eighteen studies included in the most recent review by Lehnert *et al*. (2012) only two were based in the UK and only one of these included participants aged 16 years and under. This justifies the original misgivings, highlighted in the NICE guidance (NICE, 2006), that the evidence on the cost effectiveness of prevention and management strategies for adolescent obesity is yet to be fully addressed. '*It would be valuable to run cost-effectiveness studies in parallel with clinical trials, so that patient-level data can be collected*' (NICE, 2006 (*p.*63)). In order to allow greater comparison between types of intervention and improve assumptions made in cost-effectiveness analyses, there is a need for the reporting of more information from QoL questionnaires throughout the intervention

and follow-up period to aid the assessment of the value of weight management interventions aimed at the younger population.

Different types of economic evaluation exist, and the unit of measurement employed to assess benefits differs within these analyses (see Drummond, 1996 for more detail), the thesis is concerned with a specific type of evaluation known as Cost-Utility-Analysis (CUA). Here benefits are assessed using preference based instruments. Preference based instruments have been developed based on economics and decision theory in order to provide an estimate of individuals' preferences for different health outcomes (Drummond, 1996 and Bakker & van der Linden, 1995). More detailed discussion of CUA of weight management interventions is provided in **Section 2.8.3** below.

2.8.1 The measurement of outcomes in economic evaluation

Assessing the impact of weight and changes in weight on Health Related Quality of Life (HRQoL) or more generally QoL is important for CUA. The concept of QoL or HRQoL needs to be identified and defined before assessing the benefits derived from weight management interventions. Preference based instruments, describing different health states (descriptive systems), are used to depict different health states in the population of interest. These instruments are made up of items (single questions), levels associated with each item (response scales) and dimensions. Dimensions consist of the aggregation of items and their associated levels that are concerned with a specific aspect of health, for example, physical, emotional or social functioning. The descriptive system allows the classification, or portrayal, of different health states. Once the definition of QoL or HRQoL is identified other key decisions need to be made in developing the content of the descriptive system. For example: a) what methods should be used to create and refine the wording of items and; b) who are the relevant populations for developing the content of the descriptive system?

The EQ-5D (EuroQol Group, 1990), a commonly utilised preference based instrument for conducting CUA (Dolan *et al*., 1996), is used as an example to illustrate the aforementioned concepts. It is a self-reported questionnaire consisting of two pages comprising the EQ-5D descriptive system and the EQ-VAS (though sometimes only the former is utilised in empirical studies). In terms of the descriptive system, EQ-5D health states are defined according to five dimensions (each containing a single item): mobility, self care, usual activities, pain/discomfort and anxiety/depression. There are now two adult versions of the EQ-5D. The older version has three levels; however, a five-level version of the instrument has recently been launched (Herdman *et al*., 2011). In the earlier version, each of the five dimensions has three levels of

severity, which essentially equate to 'no problems', 'some problems' and 'severe problems' (i.e. 3 response categories – see **Figure 2.1** for the EQ-5D™ health questionnaire descriptive system). In the item relating to the mobility dimension of the EQ-5D, individuals are asked to select one out of three, the statement which best describes their own health: 1) no problems walking about, 2) some problems walking about or 3) confined to bed. An individual's response to all five of the items in the EQ-5D generates a description for a particular health state. The three-level descriptive system thus generates 243 theoretically possible health states. Each state has a five-digit identifier that signifies the level of severity on each dimension; for example, 11111 is a state with 'no problems' on all five dimensions, whereas state 33333 has 'severe problems' on all dimensions.

Figure 2:1 EQ-5D™ health questionnaire (EuroQol Group, 1990)

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility	
I have no problems in walking about	❑
I have some problems in walking about	❑
I am confined to bed	❑
Self-Care	
I have no problems with self-care	❑
I have some problems washing or dressing myself	❑
I am unable to wash or dress myself	❑
Usual Activities (e.g. *work, study, housework, family or leisure activities*)	
I have no problems with performing my usual activities	❑
I have some problems with performing my usual activities	❑
I am unable to perform my usual activities	❑
Pain/Discomfort	
I have no pain or discomfort	❑
I have moderate pain or discomfort	❑
I have extreme pain or discomfort	❑
Anxiety/Depression	
I am not anxious or depressed	❑
I am moderately anxious or depressed	❑
I am extremely anxious or depressed	❑

2.8.2 Health state valuation

The consideration of benefits (i.e. QALY improvements derived from competing interventions) for CUA requires: i) the identification of a descriptive system to measure health states and ii) the valuation of those health states. In order to compare the benefits derived from competing interventions, and how it impacts upon different dimensions of health, a valuation of the health state is undertaken by utilising stated preference (SP) elicitation methods. There are different methods that can be used to elicit health state preference values (a measure of strength of preference that an individual assigns to a particular health state). The traditional approaches, utilising cardinal techniques, include: the Visual Analogue Scale (Huskisson, 1974), Time Trade Off (Torrance *et al*., 1972) and Standard Gamble (based on the axioms of expected utility theory, see von-Neumann & Morgenstern, 1947). A description of each method is given below; further details are available from a number of sources, see Drummond *et al*. (1996) for example.

Visual analogue scale (VAS) - Respondents are asked to rate a health state on a scale from 0 to 100 (which is then transformed into a 0 to 1 scale), with 0 representing 'the worst imaginable health state' and 100 representing 'the best imaginable health state'. Say an individual gives a hypothetical dysfunctional health state a value of 50, the utility value assigned to the given health state is 0.5 (50 divided by 100)[5].

Time trade off (TTO) - Respondents are asked to choose between remaining in a state of ill health for a pre-defined period of time (say 10 years), or having perfect health for a shorter length of time, both followed by immediate death. The amount of time in perfect health is altered until the individual reaches indifference between the two choices. Once a point of indifference is reached a health state preference value can be calculated. For example if an individual is indifferent between being in a hypothetical dysfunctional health state for ten years or being in full health for five years, both followed by immediate death, the preference value assigned to this hypothetical state is 0.5 (5 divided by 10).

[5] Assuming that the respondent thought the worst imaginable health state is equivalent to being dead.

Standard gamble (SG) - Respondents are asked to choose between two alternatives: one certain and the other uncertain (the lottery). In the first alternative the individual remains in ill health with the certainty of staying alive (the certain choice). In the second alternative an individual faces an uncertain outcome, with a probability of restoration to perfect health (for the duration of their remaining years followed by death), and the probability of immediate death (the lottery - one minus the probability of perfect health). The probability of immediate death and perfect health is changed in the second alternative until the individual is indifferent between the two alternatives. As with TTO, once the individual reaches indifference between the two choices the health state preference value can be calculated. In the situation where an individual is indifferent between taking a 0.5 chance of immediate death compared to living the rest of their life in a hypothetical dysfunctional health state, the preference value assigned to that health state is 0.5 (one minus the probability of immediate death).

These methods place all possible health states on a scale that includes zero (the state of *being dead* or a state valued as equal to being dead) and one (the state of being in *full health* or the *best imaginable health state* – as described by the HRQoL or QoL measure). States perceived as worse than being dead can also be valued using these three methods. Usually, a subset of all the possible health states from the descriptive system is valued; then econometric modelling is used to assign values to the remaining states. Once values have been assigned to all possible health states within a given HRQoL or QoL measure, the preference based instrument is derived (see Dolan *et al*., 1996 and McCabe *et al*., 2005 for example). Since all states can in principle be measured on such a common scale, even states referring to very different conditions can be compared with each other.

Alternative ordinal valuation methods have also been developed in the elicitation of preference values including ranking and discrete choice experiments (DCE).

Ranking – Individuals are asked to place health states in order of severity. The states perceived as most and least severe are at opposite ends and states in between are ordered in terms of severity between these two extremes. Ranking is often used as a warm-up exercise in valuation studies. Although ordinal data collected from the ranking of health states have been included in a number of health-state valuation exercises, rank responses are not typically used to estimate cardinal values needed for CUA. However, the use of ordinal rank responses to derive cardinal values for health states is emerging as a new method for deriving preference valuations for health states (for example see McCabe *et al*., 2006 and Craig *et al*., 2009 for further details).

Discrete Choice Experiment (DCE) – In the case of DCEs, individuals are presented with a DCE exercise, consisting of a number of choice tasks made up of two or more profiles. Profiles can describe different health states comprising combinations of dimensions of QoL (attributes) and associated response levels (a profile can also include a survival attribute). Individuals are asked to indicate their preference over the alternative profiles presented to them. The aim of a DCE valuation study is to estimate individuals' preferences over different attributes by identifying the different levels at which they trade good levels (less dysfunction) for attributes that are most important to them for poor levels (more dysfunction) in attributes that are less important to them. Reviews by Lancsar & Louvier (2008) and Ryan & Gerard (2003) provide a summary of the use of DCEs in the health literature. The addition of a survival attribute to a DCE health profile has recently been employed in order to generate preference values anchored on the full health dead scale (Bansback *et al*., 2012). A more detailed description of the DCE elicitation method is provided in **Chapter 8**.

2.8.3 CUA of weight management interventions

To allow consistent decision making across appraisals, as part of their technology appraisal process, economic evaluations for NICE are formulated according to a "reference case" (NICE, 2013). This states that the quality-adjusted-life-year (QALYs), which combines both length and quality of life into a single summary measure, should be used. The QALY should capture the health benefits of competing interventions in terms of incremental QALYs gained. NICE identifies the EQ-5D as the appropriate source for the application of pre-defined utility weights for the QALYs. NICE do say that other condition specific measures can be used if there is sufficient justification because in some cases, the applicability of the EQ-5D is limited. In terms of the assessment of weight management interventions, the EQ-5D may not fully reflect the effect of weight status or changes in weight status on QoL. For example, a study conducted with the adult population that utilised a generic preference based instrument to derive preference weights for a weight specific instrument, suggested that important dimensions influenced by weight status such as *public distress* and *sexual functioning* are not accounted for when generic instruments are utilised (Brazier *et al*., 2004). Whilst condition specific instruments may be more applicable for the assessment of weight status and changes in weight status on QoL, the majority of existing weight specific QoL measures lack preference weights and thus they cannot be used to calculate QALYs.

2.9 The knowledge gap

Currently, there are no valid and reliable preference based weight specific measures of QoL available where the content has been informed by adolescents. As adolescents will be the users of the instrument, they would be the most relevant population to inform its content. If, as suggested, the burden of obesity to young people is significant then the prioritisation of weight-management interventions is essential. However, the lack of an appropriate instrument impedes the accurate estimation of the value of alternative weight management interventions for this population. This highlights the need for a new preference based weight specific instrument to assess the full impact of interventions aimed at adolescents who are obese and overweight.

Empirical research shows that increased weight status above the healthy range has a significantly negative impact on QoL and HRQoL (Tsiros *et al*., 2009 and Griffiths *et al*., 2010). It is also known that for the younger population the impact of obesity is particularly troublesome as it is likely to carry on into adulthood (Freedman *et al*., 2005 and Rowlinson, 2011). Evidence shows that the direct costs of obesity and its associated co-morbidities are linked with significant cost implications for the NHS (Butland *et al*., 2007). The existing literature is limited in providing us with utility values associated with weight specific health states, particularly for the younger population. Thus we are limited in our ability to carry out fully informed CUA of weight management interventions, aimed at this population as the evidence suggests that generic measures of QoL do not perform as well as weight specific instruments, in discriminating between different BMI subgroups on several dimensions (Kolotkin *et al*., 2006).

In order to make fully informed resource allocation decisions relating to health care provision, decision makers take into consideration information derived from CUA. This analysis requires valid and reliable information on the costs and benefits of competing interventions. For decision making regarding weight management interventions, provided to the younger population, decision makers would benefit from having accurate information on the relevant dimensions of QoL affected by changes in weight status. In order to identify whether changes in QoL signify: a). no difference in QoL, b). a decline in QoL or, c). an improvement in QoL; preference weights are also necessary.

2.9.1 The value of bridging the knowledge gap

Accurate preference based measurement of outcomes that are of value to obese and overweight adolescents and that is useful to decision makers will allow economic evaluations to provide a fully informed evaluation of the effectiveness of interventions aimed at this population. Adolescents will benefit as interventions will be able to address aspects of their QoL that are relevant to them, as the impact of weight and weight changes in the adolescent population is only beginning to be fully assessed from the point of view of adolescents themselves (see Doyle, 2011 and Morales *et al*., 2011). Decision makers will benefit as they will be able to better identify and justify interventions that are cost effective. In the proposed research: a weight specific QoL descriptive system for adolescents (aged 11-18) will be created, utilising the preferences of adolescents, the existing literature, and the views of specialists in the field of adolescent obesity. As such, this research will provide a means for better informed commissioning decisions within the NHS, government, and local authority more widely, by undertaking empirical studies based in the UK setting. It will facilitate a more complete assessment of the impact of adolescent obesity, into future economic evaluations, by integrating adolescents' views into the research process. Overall the thesis will provide a further impetus to addressing obesity within NHS and, wider government policy.

2.9.2 Conclusion

This chapter provided an introduction into obesity and the issues around the CUA of weight management interventions. The next chapter reviews the literature on existing instruments that have been used to measure the QoL of obese and overweight adolescents and assesses their suitability for use in economic evaluation of weight management interventions.

Chapter 3
Literature review

3.1 Introduction

This chapter presents a review of the literature, the purpose of which is to achieve four key objectives: Firstly, to assess the relationship between above normal weight status and QoL in the adolescent population. There is no agreed upon precise definition of adolescence to date due to the difficulties related to identifying the physical, psychological, and cultural expressions of development that may begin earlier in life and end later in life. Different studies use different age ranges to refer to the adolescent population and this can complicate the evaluation of literature for this group. In current review the population of interest falls between 11 and 18 (inclusive). A pragmatic decision was made to target the younger population due to the many negative consequences of obesity in both the short and longer term (as discussed in **Chapter 2**). It was felt that 11 to 18 year olds would be better suited to being involved in the qualitative interviews to elicit aspects of life affected by weight status, whilst younger age groups might struggle to communicate their views. Secondly, to adopt the methods utilised in two recent reviews identifying existing instruments previously used in this population. Thirdly, to identify instruments that have not been used in this population but, could be relevant (i.e. instruments that have been recently developed). Finally, to assess the measurement properties of existing weight specific instruments that are appropriate for the adolescent population. In addressing these objectives the literature review will verify: a) whether obesity has any impact upon QoL, b) what impacts are observed and c) what subgroups are most affected. The review will also facilitate the identification of existing weight specific QoL measures, both non preference and preference based, for the obese and overweight adolescent population. This information is important in informing the development of a new tool that is valid and reliable for economic evaluation of weight management services.

3.2 Methods

3.2.1 Relationship between weight status and QoL

In the first instance, the findings from recent reviews by Tsiros *et al*. (2009) and Griffiths *et al*. (2010), who investigated the relationship between HRQoL and QoL and weight status in the child and adolescent population, were evaluated and summarised to gain insight into this relationship (the relationship with self-esteem was also assessed in the Griffiths *et al*. (2010) review) and a summary table describing the two reviews is provided in **Table 3.1**.

Table 3:1 Summary of the reviews by Tsiros *et al*. (2009) and Griffiths *et al*. (2010)

	Tsiros *et al*. (2009)	**Griffiths *et al*. (2010)**
Aims	To assess the effect of overweight and obese weight status on HRQoL in child and adolescent populations and utilise a pooled analyses (where feasible) of studies presenting Paediatric Quality of Life Inventory (PedsQL) scores	To examine the relationship between paediatric obesity and self-esteem or QoL and assessment of the possible effects of age, gender, ethnicity, and presence of other medical conditions on obesity
Inclusion criteria	Studies assessing the following: population under 21 years of age, weight status was assessed and finally, HRQoL was assessed either by parent proxy or self-report	Studies that used validated multi-dimensional measures of self-esteem or QoL, and those using an agreed definition of childhood obesity (i.e. the review only looked at obese-normal weight differences, not overweight - normal weight differences)
Search period	From inception to March 2008	From 1994 to March 2009
Databases searched	(1) Medline (2) EMBASE (3) Web of Science (4) Cochrane Library (5) CINAHL (6) AMED (7) PubMed	(1) Medline (2) EMBASE (3) PsycINFO (4) DARE (5) HTA
No. studies included in the review	**28**	**QoL = 25** (Self-esteem=17, Total=42)

3.2.2 Literature search

Following the assessment of the Tsiros *et al*. (2009) and Griffiths *et al*. (2010) reviews, two literature searches were conducted and aimed to address specific questions.

Literature search A: identification of the instruments used in the measurement of QoL in obese and overweight adolescents (undertaken in December 2009 / January 2010)

- Which existing preference based weight specific instruments are available?
- Which existing generic and weight specific multidimensional QoL or HRQoL instruments have been used with obese or overweight adolescents and how sensitive are they to weight status or changes in weight status?

Literature search B: assessment of the measurement properties of weight specific instruments developed for the adolescents (undertaken in January 2012)

- What new preference based or non preference based multidimensional weight specific measures have been developed for the adolescent population since literature search A was undertaken?
- What is the evidence on the psychometric performance of existing weight specific instruments developed for adolescents?

This work built upon the Tsiros *et al.* (2009) and Griffiths *et al.* (2010) reviews as: a) both reviews employed a specific framework of QoL or HRQoL, b) the Griffiths *et al.* (2010) review was restricted to studies that used *validated* multidimensional measures of QoL, and studies using an agreed definition of childhood obesity (studies involving overweight adolescents were excluded), c) the reviews do not assess the sensitivity of weight specific and generic instruments to weight status or changes in weight status and d) neither of the reviews assessed the measurement properties of existing instruments. Finally, in order to identify relevant instruments that may not have been picked up in the search because they have only recently been developed, specialists in the field of PROMs were consulted.

A sensitive review was undertaken in literature search A in order to ensure that all potentially relevant instruments were identified. This was complemented by literature search B, a specific search, with aimed to identify only weight specific instruments principally used with adolescents aged 11 to 18 years. In the latter only the relevant existing weight specific instruments, including those identified from literature search A, were evaluated. The search strategy for both literature searches was developed by the candidate who also ran the searches and collated the findings. Advice on the development of search strategy was provided by an information specialist. The findings from literature search B were screened by

both the candidate and the primary supervisor. Further details of the methods undertaken in the reviews are presented below, followed by the results. Appraisals of the overall findings are discussed along with the implications for the research in the final section.

3.2.2.1 Search strategy

The searches involved a systematic review of the literature of (a) multidimensional QoL or HRQoL and weight specific measures that have been used with (b) young people who are (c) overweight or obese. Comprehensive searches of key electronic databases were undertaken (see **Table 3.2**).

Table 3:2 Electronic databases included in the literature searches

Databases searched	**Literature search A:** assessment of generic and weight specific instruments that have been used with obese or overweight adolescents	**Literature search B:** assessment of the measurement properties of weight specific instruments designed for adolescents
MEDLINE & MEDLINE in Process	✓	✓
EMBASE Classic & EMBASE	✓	✓
AMED	✓	✓
CINAHL	✓	✓
PsycInfo	✓	
The Cochrane Library	✓	
SPORTDiscus		✓

The search strategy was devised with the consultation of search strategies from recent reviews (Toris *et al*., 2009 and Griffith *et al*., 2010) and finally with support from an information specialist. This helped produce the appropriate key words and combination of key words to run through electronic databases (see example search strategy in **Appendix 3.1**, results of all the searches that were undertaken are available upon request).

3.2.2.2 Review strategy

Search results were combined, stored, and sorted in an EndNote library (Version – X3). Search results from each database were recorded and duplicates identified and removed in order to obtain the final list of references to be assessed for inclusion in each review. The screening of retrieved papers was two-staged. Titles and abstracts were first screened. The full papers of studies that pass this initial review process were then screened. Screening was in accordance to the inclusion and exclusion criteria given below. Screening for literature search A was undertaken by the candidate and the screening for literature search B was independently carried out by two reviewers (the candidate and one supervisor).

3.2.2.3 Inclusion and exclusion criteria

3.2.2.3.1 Literature search A - QoL & weight specific instruments

Date of publication: No date limits were set.

Language of publication: Only studies published in English were included.

Type of outcome measure: Multidimensional non preference based and preference based generic or weight specific QoL outcome measures were included in the review. Studies using measures that were specifically designed for the adult population were excluded from the review as were studies utilising unidimensional measures of QoL such as self-esteem, physical activity, disordered eating or disease specific QoL instruments that were not weight specific such as asthma or sleep related QoL measures. No restrictions were placed on the mode of administration i.e. self report, proxy report, or interviewer administered. Studies reporting outcome using only clinical rating scales were also excluded.

Study setting: The search included studies regardless of whether they evaluated use of drug therapy or related interventions in the treatment of obesity / overweight. No restrictions were imposed on the use of a comparator. Studies were included in the literature review if more than one or no interventions were being assessed.

Subjects: Studies were included within the literature search if the population of interest were obese or overweight 11-18 year olds. If most of the participants were aged less than 11 years old or over 18 years old the study was excluded from the review as the majority of the study population was outside of the age range of the current review. The only exception to this rule was where a preference based weight specific instrument was identified. A specific application of an obesity / overweight measure (for example using BMI) was not necessary. No restrictions were placed on the gender of individuals included in the study. Studies assessing obesity as well as other co-morbidities (such as sleep *apnoea*, diabetes, or cystic fibrosis) were also excluded.

3.2.2.3.2 Literature search B – Measurement properties

The criteria above were applied to the findings from literature search B additionally studies were included in the review if:

a) The primary aim of the study was to evaluate the measurement properties of a weight specific instrument
b) Instruments were developed specifically for the younger population (incorporating ages 11 to 18 years)
c) Conference abstracts were included in the review. Relevant instruments identified from conference abstracts, were assessed. Authors of relevant conference abstracts were contacted

Studies were excluded from the review if they assessed the use of or the development of generic QoL.

3.2.2.4 Evaluation of studies

The reference lists of studies shortlisted in review A were assessed to identify the original studies that created the generic and weight specific instruments and a data extraction process was applied once all these references were collated.

The methodological quality of the studies included in the measurement properties review (review B) was assessed using the ‘‘COnsensus-based Standards for the selection of health status Measurement INstruments’’ checklist (COSMIN checklist, Mokkink *et al*., 2010a). This checklist consists of nine boxes with standards for how each measurement property should be assessed. Each item was scored on a four-point rating scale (i.e. “poor”, “fair”, “good,” or “excellent”). The COSMIN taxonomy and definitions (Mokkink *et al*., 2010b) were employed to decide which measurement properties were evaluated in a study and which corresponding boxes were completed (see **Appendix 3.2**). An overall score for the methodological quality of a study was determined by taking the lowest rating of any of the categories in the checklist.

3.3 Results

3.3.1 Relationship between weight status and QoL

In the majority of their conclusions, there are no major differences in the findings of the Tsiros *et al*. (2009) and Griffiths *et al*. (2010) reviews. Both are in agreement in terms of the important overall impact of weight on QoL. Both reviews found that a *BMI above healthy normal limits* had a negative impact on the QoL of adolescents. In terms of how QoL is affected by above normal weight status, significantly lower scores in obese or overweight children & adolescents compared with their lean counterparts were reported mostly in the physical and social functioning dimensions of QoL in the two reviews. Evidence also demonstrated decrements in emotional functioning.

One point of contention between the two reviews was in their conclusions about the impact of obese weight status on school functioning. Tsiros *et al.* (2009) find limited evidence to support a negative relationship. The authors found only two studies indicating that school functioning scores were significantly different when comparing obese and lean samples. Both studies obtained child and parent reports using the Paediatric Quality of Life Inventory (PedsQL) (Varni et al., 2001)[6]. One study found that school functioning was significantly impaired in the parent proxy but not child reports, whereas the second study found a significant inverse relationship between BMI z-score and school functioning subsets of HRQoL in both the child and parent reports. Griffiths *et al.* (2010), however, find more evidence to support it. Six studies out of fifteen cross-sectional studies in the Griffiths *et al*. (2010) review reported lower school functioning in obese children and / or adolescents, all the studies utilised the PedsQL to measure HRQoL (these were based on either both child reports, and parent proxy reports or both).

Regarding this apparent contradiction between the two reviews, it is reasonable to deduce that there is evidence of a negative impact of obesity on school functioning. In the majority of the studies reported, the same instrument was used to measure QoL – the PedsQL. However the reviews are drawing upon different literature – primarily the Griffiths *et al*. (2010) review does not include studies comparing overweight-normal weight children and adolescents; however the Tsiros *et al*. (2009) review does. This may have resulted in studies of overweight adolescents, that show non-significant associations with school functioning, being excluded from the findings of Griffiths *et al*. (2010). Other non-QoL related studies have identified a negative relationship between obese weight status and educational attainment, achievement

[6] A description of the PedsQL instrument is provided in **Table 3.4**.

and Hill *et al*. (2010) provides a discussion of some of the studies in this area. The evidence provided in the chapter supports the negative relationship between obese weight status and objective measures of school functioning. Generally, evident inconsistency in the relationship between weight status and school functioning may reflect the difference in content of instruments assessing the school functioning dimension. For example, two of the five scale questions in the PedsQL ask about absence from school while other assessments might just measure educational outcome (i.e. highest educational qualification). Other measures might focus on school achievement or performance (using school test scores) but not assess non-attendance (Griffiths *et al*., 2010).

Both reviews also assess the impact of weight loss on QoL / HRQoL of obese and overweight adolescent populations. Both find that the majority of studies that reported weight loss also reported increases in global QoL / HRQoL. A particularly interesting observation by Griffiths *et al.* (2010) is that *'evidence of psychological benefit even in the absence of weight loss has previously been reported and must therefore be associated with some feature of the environment or support network in which they are organised. This finding clearly merits further investigation. It suggests that weight management programs have the potential to equip obese youth with positive self-evaluations necessary for positive lifestyle changes that may enhance their future well-being, even if weight loss is not apparent in the short-term'* Griffiths *et al.* (2010 (*p.* 301)). This is of particular importance in terms of the assessment of the efficacy of weight management programmes; the primary outcome measure may be reduction in BMI, however, assessments of secondary outcomes, alongside this, is also paramount in order to understand the resultant benefits associated with these mulitcomponent interventions.

3.3.2 Literature search

The results of the number of hits from each of the databases included in the literature searches, the total number of hits before and after the removal of duplicates is shown in **Table3.3** (see **Appendix 3.3, 3.4 and 3.5** for more details).

Table 3:3 Results of literature searches

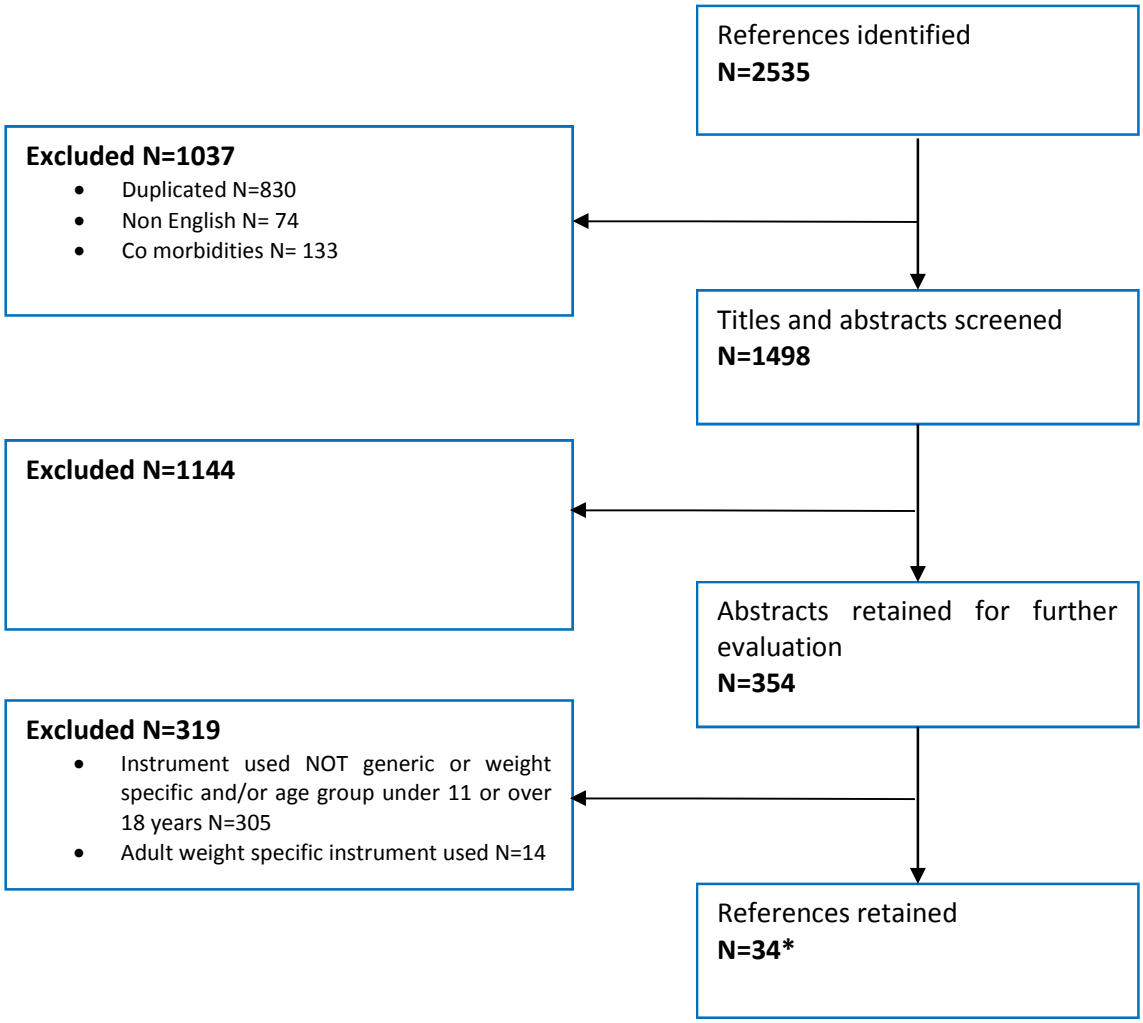
Database	Search A	Search B
	QoL and HRQoL instruments used with obese or overweight adolescents	Measurement properties of weight specific instruments designed for adolescents
	Hits	
MEDLINE and Medline in Process	579	1021
EMBASE Classic and EMBASE	462	1464
AMED	32	42
CINAHL	984	76
PsycInfo	217	NA
The Cochrane Library	261	NA
SPORTDiscus	NA	166
Total including duplicates	**2535**	**2769**
Total unique references assessed	**1705**	**2036**

3.3.2.1 Literature search A - QoL and weight specific instruments identified

The results of the literature search are given in **Figure 3.1**. Of the 2535 references identified in the initial search, 1498 unique references underwent title and abstract sifting. The application of the inclusion criteria defined above resulted in a total of 354 references retained for further assessment, from which thirty-four references were retained. Eleven instruments were identified from these studies consisting of six generic: 1) Child Health Questionnaire (CHQ, Waters *et al*., 2001), 2) KINDL (Ravens-Sieberer & Bullinger, 1998), 3) Pediatric Quality of Life Inventory (PedsQL, Varni *et al*.,2001), 4) EQ-5D-3L (EuroQol Group, 1990 and Dolan *et al*., 1996), 5) Health Utility Index (HUI, Furlong *et al*., 2001 and McCabe at al., 2005) and 6) Short Form-36 (SF-36, Ware & Sherbourne, 1992 and Brazier *et al*., 2002). Five weight specific instruments were identified 1) KINDL-Obesity module (Ravens-Sieberer *et al*., 2001), 2) Impact of Weight on Quality of Life –Kids version (IWQOL-Kids, Kolotkin *et al*., 2006), 3) Moorehead-Ardelt Quality of Life Questionnaire II (M-A-QoL Q, Moorehead *et al*., 2003), 4) Sizing Me Up (Zeller, & Modi, 2009) and 5) Sizing Them Up (Modi & Zeller, 2008). None of the weight specific instruments developed for the adolescent population were preference based. However, one preference based weight specific instrument developed for adults was found (Impact of Weight on Quality of Life – Adult version (IWQOL–Lite, Kolotkin *et al*., 2001 and Brazier *et al*., 2004) and, due to it being the only one existing, was evaluated.

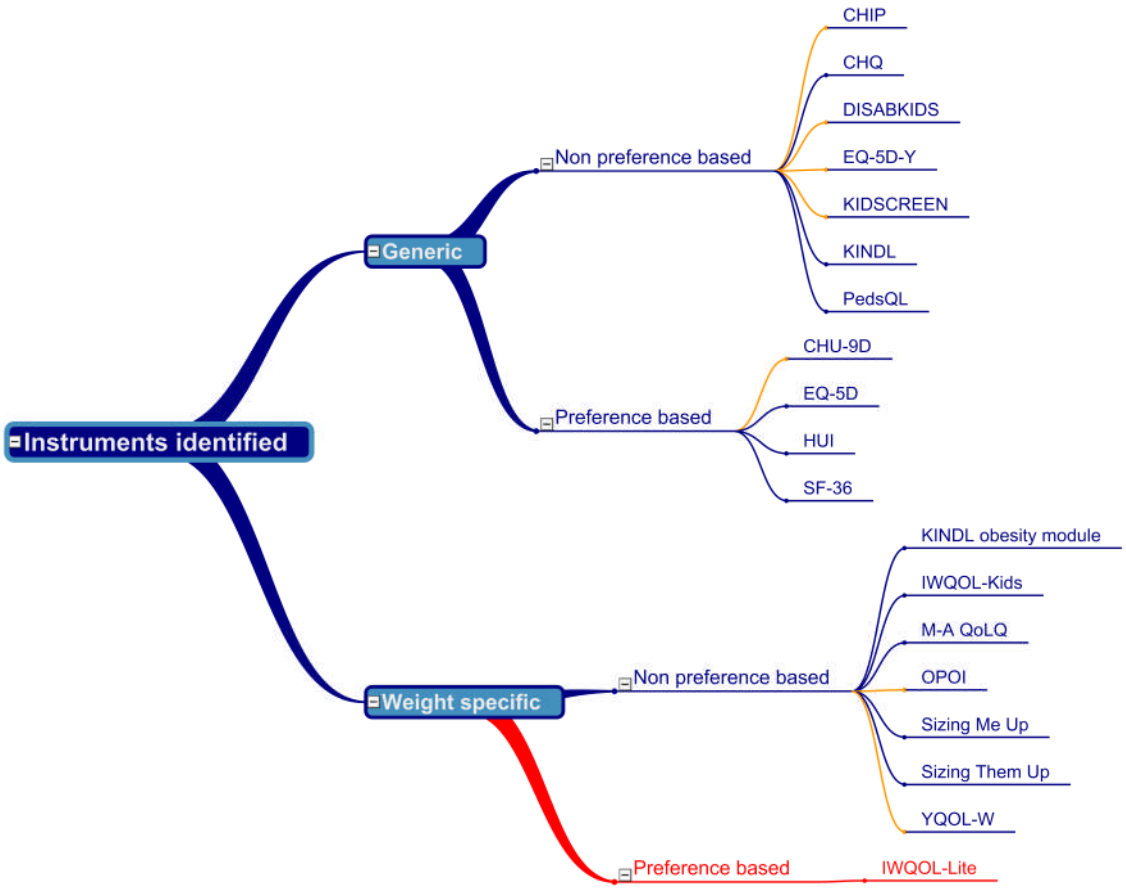
In addition, relevant instruments were identified from other sources and added to the review. The Child Health Impact Profile (CHIP, Starfield *et al*., 1993), was not picked up in the search, though it was identified in both the Tsiros *et al.* (2009) and Griffiths *et al.* (2010) reviews. This instrument was added to the review. Four newly developed generic measures were identified from consultation with specialists in PROMs (Child Health Utility – nine Dimensions (CHU-9D), Stevens, 2009 and 2012), DISABKIDS (Barrs *et al*., 2005), KIDSCREEN (The KIDSCREEN Group Europe, 2006) and EQ-5D-Y (Wille *et al*., 2010). Also added to the review were: a) two recently created weight specific instruments suitable for adolescents (Youth Quality of Life – Weight module (YQOL-W, Morales at al., 2010) and Oxford Paediatric Obesity Instrument (OPOI, Doyle *et al*., 2011) and b) a second recent study that estimated the importance weights of different dimensions of the IWQOL-Lite instrument (Hauber at al., 2010); the three studies were identified from literature search B. The pooled results of the multidimensional QoL, HRQoL and weight specific instruments that have been used with obese or overweight adolescents, are summarised in **Figure 3.2**. The seven orange connectors in the figure illustrate the instruments that were identified outside of literature search A, whilst the red connector signifies the adult weight specific preference based measure that was identified.

Figure 3:1 Results of literature search A



*One duplicated reference was identified from the thirty-five originally shortlisted references leaving thirty-four unique references that met the inclusion criteria. From these references seven generic and five weight specific instruments were used with obese and overweight adolescents.

Figure 3:2 Summary of instruments identified from the literature



Key			
Generic		**Weight specific**	
CHIP	Child Health Impact Profile	KINDL obesity module	Health-Related Quality of Life in Children and Adolescents – obesity module (originally a German-language instrument)
		IWQOL-Kids	Impact of Weight on Quality of Life –Kids version
CHQ	Child Health Questionnaire	M-A-QoL Q	Moorehead-Ardelt Quality of Life Questionnaire
DISABKIDS	Developed from the European DISABKIDS project –for use with chronic conditions	OPOI	Oxford Paediatric Obesity Instrument
EQ-5D-Y	EuroQol-5D questionnaire – Youth version	Sizing Me Up	
KIDSCREEN	Developed from the European KIDSCREEN project – the generic module		
KINDL	Health-Related Quality of Life in Children and Adolescents (originally a German-language instrument)	Sizing Them Up	
PedsQL	Paediatric Quality of Life Inventory	YQOL-W	Youth Quality of Life – Weight
Generic – Preference based		**Weight specific – Preference based**	
CHU-9D	Child Health Utility – nine Dimensions	IWQOL - Lite	Impact of Weight on Quality of Life –Adult version
EQ-5D	EuroQol-5D questionnaire (five dimensions three level)		
HUI	Health Utility Index		
SF-36	Short Form-36		

3.3.2.1.1 The preference based IWQOL-Lite instrument

There was only one preference based weight specific QoL instrument that was identified in the literature search the IWQOL-Lite instrument (IWQoL-Lite, Kolotkin *et al.,* 2001). Utility values have been derived for the IWQOL-Lite instrument in two ways: a) mapping to a generic preference based instrument (Brazier *et al*., 2004) and b) using conjoint analysis[7] (Hauber *et al*., 2010). As there are limited search findings in this research area there is merit in assessing these two studies, even though the applicability of the IWQOL-Lite to the adolescent population is questionable due to it being created for the adult population.

The IWQOL was developed by clinicians specializing in the treatment of obesity who catalogued patients' (aged 18 and over) concerns about the impact of their obesity. Items were developed based on these concerns, and verified with patients that the items were comprehensive and accurate (Kolotkin *et al.,* 1995). The IWQOL-Lite is a short form 31-item instrument of the original IWQOL questionnaire (containing 74 items) and has five dimensions: Physical Function (11 items), Self-Esteem (seven items), Sexual Life (four items), Public Distress (five items), and Work (four items) (Kolotkin *et al.,* 2001). Response options to all IWQOL-Lite items are rated by the research subject as *always true*, *usually true*, *sometimes true*, *rarely true*, or *never true*. *Always true* responses are given a score of five, whilst *never true* responses are given a score of one. A single score for each of the five different scales or a total score for the entire instrument can be calculated. Higher scores indicate poorer QoL for each subscale and also for the total scores (ranging from 31 to 155).

Estimates of preference weights for the IWQOL-Lite

In order to use the IWQOL-Lite to represent health in terms of QALYs for cost-effectiveness analysis, a measure of the value that an individual puts on an IWQOL-Lite health state is necessary. Two studies, using two different methods, have assessed the relationship between the IWQOL-Lite and the relative importance of the different dimensions contained in the measure.

[7] The conjoint analysis can loosely be described as the extrapolation of ordinal responses to cardinal preferences (Bansback 2010). In the context of the Hauber et al (2010) study the methods used can be taken to be synonymous with DCE. It should be noted however that conjoint analysis is a generic term used to describe several ways to elicit preferences which, according to Louviere et al (2010), is different to DCEs in certain contexts. Conjoint analysis relies on the theory of Conjoint Measurement whilst the DCEs are based on Random Utility Theory (Louviere et al., 2010 (see **Chapter 8** for further details on DCE)).

Mapping

Brazier *et al.* (2004) mapped preference weights onto the IWQOL-Lite instrument from the SF-6D (study participants completed the SF-36, a generic measure of HRQOL converted into the preference-based SF-6D descriptive system (Ware & Sherbourne, 1992 and Brazier *et al*., 2002). Mapping is used to link alternative QoL and HRQoL measures, through the process of constructing a statistical relationship; linking one measure of health to another is usually carried out in order to estimate preference values for states described by an instrument where none are available. Brazier et al. (2004) aimed to estimate preference weights for states described by the IWQOL-Lite using regression analysis: i.e. regressing the IWQIL-Lite scores on the SF-6D utility index, using a range of methods to achieve the best fitting, unbiased model. The analysis was based on a sample of 1794 adults aged 18 to 90 years including treatment seekers and non treatment seekers with an average BMI of 35.0 (ranging from 18.6 to 91.9). The SF-6D index was regressed on: 1) the IWQOL-Lite total score, 2) the 5 IWQOL-Lite dimension scores, 3) the 31 IWQOL-Lite item scores (using a continuous scale for responses), 4) same as 3, but with the item response entered as a dummy variable with level 1 as the baseline, and finally, 5) the best performing models from 1 to 4 with additional variables for the respondent characteristics of age, sex and BMI (Brazier *et al.*, 2004). The authors found that Model 4 was the best fitting model, the specification utilising the 31 IWQOL-Lite item scores where item response was entered as a dummy variable.

A weakness that was noted by the authors was that, while the IWQOL-Lite can be mapped into the SF-6D, some of the HRQoL information in IWQOL-Lite scores was not contained in SF-6D. The impact of the *public distress* and *sexual life* dimensions in the IWQoL-Lite might not be fully reflected in the generic measure leading to the model underestimating the impact of obesity on HRQOL, as these dimensions were non-significant variables in the models. For this mapping approach to generate a valid prediction of preference weights, the preference based descriptive system needs to provide a valid description of the condition and its treatment (Brazier *et al.,* 2004). The authors concluded, '*such a mapping exercise is always a second best exercise compared to either the direct use of the SF-6D or a valuation of the condition-specific instrument*' Brazier *et al*. (2004 (*p.* 459)).

Conjoint Analysis

Hauber *et al*. (2010) utilised a choice format conjoint analysis survey to elicit and quantify the relative importance of attributes of the IWQOL-Lite, as perceived by overweight and obese adults aged 18 years and over. Conjoint analysis is a technique for eliciting judgments or

preferences over sets of outcomes or characteristics considered by individuals when making decisions (see **Chapter 4** for more details). Hauber *et al*. (2010) did not attempt to estimate a single absolute preference-based index for the IWQOL-Lite, but rather identify relative importance weights for its dimensions. These weights are not appropriate for the estimation of QALY values used in economic evaluation of weight management interventions as they are not anchored (or absolute) weights for each state described by the IWQOL-Lite instrument.

3.3.2.1.2 Instruments used with obese or overweight adolescents

Generic

A summary of the generic non preference based instruments that were identified is provided in **Table 3.4**. Most instruments cover a wide age range, typically ranging from 8 to 18 years. The number of dimensions of QoL covered by each instrument ranged from four to eleven. Young people informed the content of all seven instruments in some way. **Table 3.5** describes the four generic preference based instruments that were identified. The dimensions covered by these instruments range from five to nine comprising symptoms and functioning based items. Of the four generic instruments that were identified, there was only one where the content was informed by the views of the younger population (children aged 7 to 11 years), in the development of the CHU-9D instrument. The CHU-9D measure was originally developed with children aged 7-11 years, since then it has been validated in an adolescent population (11-17 years) (Stevens, 2009).

Weight specific

Table 3.6 presents a summary of the seven weight specific measures that have been used with the adolescent population. All of these except for one, the Sizing them up (Modi *et al*., 2008) instrument, are administered via self report. The Moorehead-Ardelt Quality of Life Questionnaire (M-A QoL Q II) was included in the review as there were two studies identified in the literature search where it was used with the adolescent population (see **Appendix 3.5**). The validation study for the instrument, however, was undertaken on a sample of participants aged 19 to 65 and was used to evaluate the measurement properties. The remaining instruments were developed for adolescents aged 5 to 18 years. The seven instruments consisted of between six to thirty items covering three to six dimensions of weight specific QoL. The content of the two most recently created instruments (the YQOL-W and OIPO) were informed by adolescents. The content of four of the seven instruments (IWQOL-Kids, M-A QoLQ, Sizing Me Up and Sizing Them Up) were informed by researchers and clinicians. It was not clear how the content of the 12 items of the KINDL-Obesity module were created.

Table 3:4 Summary of generic non preference based instruments (N=7)

Measure and reference	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
Child Health and Illness Profile (adolescent version) CHIP - AE Starfield *et al.* (1993)	11 to 17	107 items plus an additional 46 items that are specific to disease or injury (153 total)	1. Discomfort 2. Satisfaction 3. Disorders 4. Achievement 5. Resilience / Risk avoidance There is an optional *Disorders* domain that is available	Yes – but does not clearly explain how information from adolescents was used in deriving the descriptive system	The conceptualisation of adolescent health was based on a broad definition of health, on clinical experience, and on the literature on functional status and quality of life. The instrument was developed based on the literature, focus groups with children, adolescents, and parents. This was done in order to understand better what the term "health" means. A standard set of questions was posed to a set of eight groups of seven to ten adolescents and two groups of ten to twelve parents. Each with a different mixture of age, gender, race, and socioeconomic status, to learn what they thought about the meaning and indicators of good and poor health. Also the proposed dimensions and sub-dimensions were sent to more than fifty experts who represented the disciplines of adolescent and paediatric medicine, sociology anthropology, psychology, nursing, psychiatry, and health services research. Their task was to indicate missing or poorly conceptualised dimensions or sub-dimensions.
Child health questionnaire CHQ	5 to 18	Parent form PF-98, PF-50, PF- 28; Child form	CHQ-CF 87 (Child Health Questionnaire-Child Form 87): 1. Physical functioning 2. Role/social functioning (physical)	Yes	The development of the items and scales from the CHQ questionnaire were generated based on traditional qualitative techniques and methods. Their primary interest was in documenting and

Measure and reference	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
Landgraf *et al.* (1997)			3. Role/Social functioning (emotional) 4. Role/Social functioning (behavioural) 5. General health perceptions 6. Bodily pain/discomfort 7. General behaviour 8. Mental health 9. Self-esteem 10. Parental impact (emotional) 11. Parental impact (time) 12. Family functioning (family activities/cohesion) Global item: change in health		understanding the physical and psychosocial well-being of children and adolescents.
DISABKIDS Baars *et al.* (2005)	8 to 16	To be used in conjunction with KIDSCREEN - The seven condition-specific modules consist of an 'Impact' dimension and an additional dimension with a total of 10 to 12 items	*To be used in conjunction with the core generic KIDSCREEN module	Yes	The DISABKIDS group has developed a European HRQoL instrument for children and adolescents with a chronic medical condition and their parents. It includes seven chronic medical conditions: asthma, juvenile idiopathic arthritis (JIA), atopic dermatitis, cerebral palsy (CP), cystic fibrosis (CF), diabetes, and epilepsy. The generic module is provided by the KIDSCREEN Project. Central to the DISABKIDS project was the 'bottom-up' (patient-derived) nature of questionnaire construction, which was accomplished by involving children and adolescents with a chronic medical condition throughout the project. Focus groups and interviews were carried out in order to identify important HRQoL aspects from the

Measure and reference	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
					perspective of children, adolescents and their parents.
EuroQol-5D questionnaire – Youth EQ-5D – Y Wille *et al.* (2010)	8 to 18	5	Same as the five EQ-5D dimensions BUT some minor adaptations were introduced in their operationalisation to clarify the meaning of these dimensions for younger respondents 1 Mobility 2 Self-care 3 Usual activities 4 Pain/Discomfort 5 Anxiety/Depression	To a degree	An international task force revised the content of EQ-5D and wording to ensure relevance and clarity for young respondents. Children's and adolescents' understanding of the EQ-5D-Y was tested in cognitive interviews. The content of the EQ-5D dimensions proved to be appropriate for the measurement of HRQOL in young respondents. The wording of the questionnaire had to be adapted which led to small changes in the meaning of some items and answer options.
KIDSCREEN The KIDSCREEN Group (2004)	8 to 18	KIDSCREEN-52 KIDSCREEN-27 KIDSCREEN-10	The KIDSCREEN-52 instrument measures 10 HRQoL dimensions 1. Physical Well-being 2. Psychological Well-being 3. Moods and Emotions 4. Self-Perception 5. Autonomy 6. Parent Relations and Home Life 7. Social support and peers 8. School Environment 9. Social acceptance (Bullying) 10. Financial resources	Yes	The generation of the questionnaire was based on literature reviews, expert consultation, and children's focus groups across thirteen participating European countries. This was done to identify dimensions and items of HRQOL which are relevant to respondents in all countries.
KINDL Ravens-Sieberer &	8 to 16	24	1. Physical functioning 2. Emotional well-being 3. Self-esteem	Yes – based on a German population	The KINDL was derived from a conceptual model, in which the four main components of QoL, namely psychological well-being, social relationships,

Measure and reference	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
Bullinger (1998)			4. Family 5. Friends 6. School		physical function and everyday life activities, were included in interviews with children (several school classes).
Paediatric Quality of Life Inventory PedsQL Varni *et al.* (2001)	8 to 18	23 (core)	Core 1. Physical 2. Emotional 3. Social 4. School	Yes	The PedsQL was empirically derived from data collected from 291 paediatric cancer patients and their parents at various stages of treatment. The PedsQL measures the patient's and the parent's perceptions of the patient's HRQOL, as defined in terms of the impact of disease and treatment on an individual's physical, psychological, and social functioning, and by disease/treatment-specific symptoms.

Table 3:5 Summary of generic preference based QoL instruments (N=4)

Measure and study	Respondent Age range	No. of items	Name dimensions	Adolescents' involvement in deriving the descriptive system?
Child health Utility-9D CHU-9D Stevens (2009 and 2012)	7 to 17	9	1. Worried 2. Sad 3. Pain 4. Tired 5. Annoyed 6. Schoolwork 7. Sleep 8. Daily routine 9. Activities	**Yes** - Interviews were carried out with over 70 children aged 7-11 years
EuroQol-5D questionnaire EQ-5D EuroQol Group (1990) Dolan *et al*. (1996)	12 Plus	5	1 Mobility 2 Self-care 3 Usual activities 4 Pain/discomfort 5 Anxiety/depression	**No** – Dimensions of the original descriptive system were selected mainly based on examination of the content of existing health status measures and the opinions of the EuroQol group members. Existing scales included: Quality of Well Being Scale Sickness Impact Profile Nottingham Health Profile Rosser Index Additional measures that were being used by members of the EuroQol Group were also included in the analysis.
The Health Utilities Index HUI Furlong *et al*. (2001) McCabe *et al*. (2005) (the HUI self administered questionnaire is an amalgamation of the HUI 2 and 3 descriptive	5 Plus (self complete age 10 plus)	15	**HUI2** classification system 1 Sensation 2 Mobility 3 Emotion 4 Cognition 5 Self-Care 6 Pain **HUI3** classification system	**To a degree** - Originates from the aim of describing the health status of survivors of treatment for childhood cancer. Final descriptive system comes from surveys assessing individual paediatric patients using the multi-attribute health status classification system. The original content of the descriptive system was not developed with adolescents.

Measure and study	Respondent Age range	No. of items	Name dimensions	Adolescents' involvement in deriving the descriptive system?
systems)			1 Vision 2 Hearing 3 Speech 4 Ambulation 5 Dexterity 6 Emotion 7 Cognition 8 Pain	
The short from 36 SF-36 Ware & Sherbourne (1992)	14 Plus	36	1. Physical function 2. Role limitation - physical 3. Role limitation - emotional 4. Social functioning 5. Mental health 6. Energy/vitality 7. Pain 8. General health perception & change in health	**No** - The Short From 36 (SF-36) was constructed for use in the medical outcomes study (MOS). Most of these items have been adapted from instruments that have been used in research studies over the past 20 to 40 years or longer.

Table 3:6 Summary of adolescent weight specific QoL instruments (N=7)

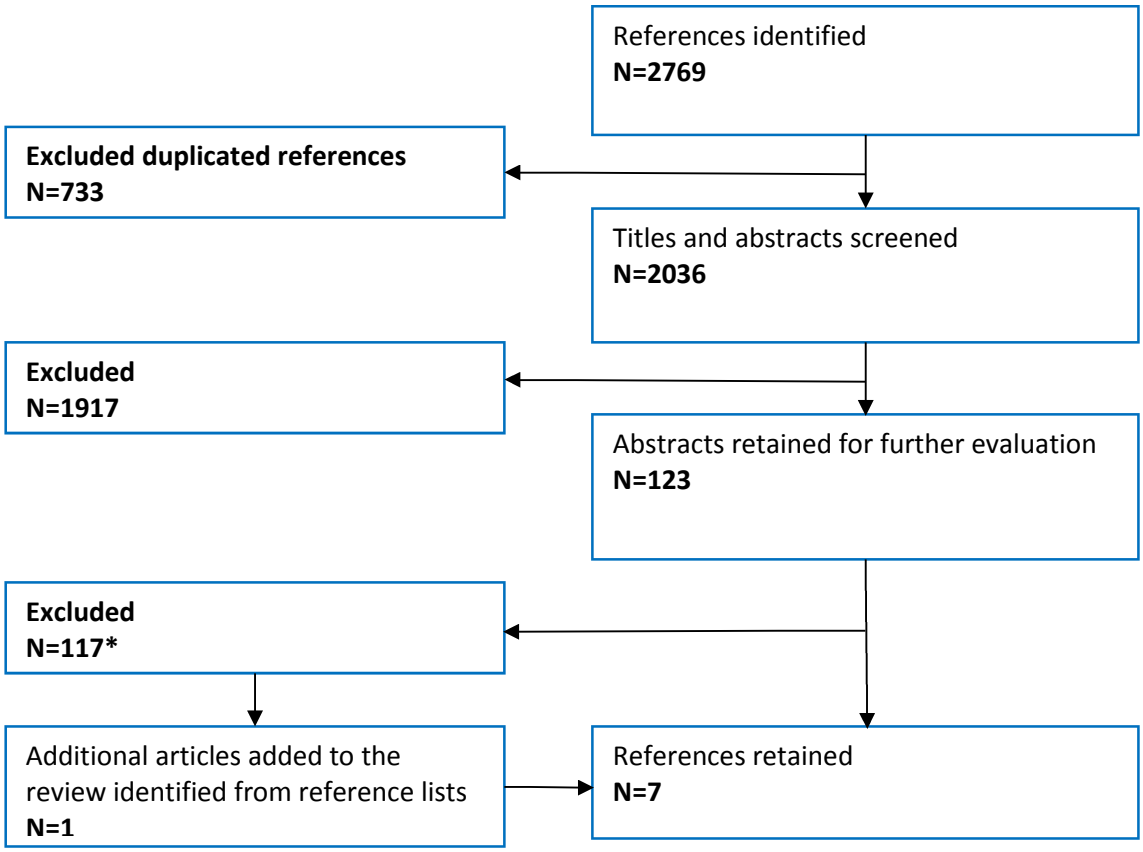
Measure and study	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
KINDL-Obesity Module Ravens-Sieberer *et al*. (2001)	8 to 16	12	1. Physical well-being 2. Emotional well-being 3. Self-esteem 4. Family 5. Friends 6. Functional aspects	Unclear	Information on how the 12 items were created was not provided.
Impact of Weight on Quality of Life – Kids IWQOL-Kids Kolotkin *et al*. (2006)	11 to 19	27	1. Physical comfort (six items) 2. Body esteem (nine items) 3. Social life (six items) 4. Family relations (six items)	No	The original IWQOL was developed in a clinical setting for moderate to severe obesity, and measured those aspects of quality of life that were identified by obese persons in treatment to be of the greatest concern. Using a literature search, clinical experience, and consultation with paediatric clinicians, the IWQOL-Kids was modelled after the IWQOL-Lite.
Moorehead-Ardelt Quality of Life Questionnaire version II M-A QoL Q II Moorehead *et al*. (2003)	Unclear	6	1. General self esteem 2. Physical activity 3. Social contacts 4. Satisfaction concerning work 5. Pleasure related to sexuality 6. Focus on eating behaviour	No	The content of the instrument was generated by clinicians in the field of bariatric surgery.
Oxford Paediatric Obesity Index OPOI Doyle (2011)	7 to 13	30	1. Physical wellbeing 2. Social wellbeing 3. Emotional wellbeing	Yes	The content of the OPOI was informed by focus group interviews with children recruited from community-based paediatric obesity programmes in the UK, with the intention of establishing what impact obesity has on the lives of children and to determine which HRQOL domains are affected. A panel of experts in paediatric obesity and PRO development reviewed the draft items.

Measure and study	Respondent Age range	No. of items	Name of dimensions	Adolescents' involvement in deriving the descriptive system?	How was the descriptive system derived?
Sizing Me Up (self complete) Zeller & Modi (2009)	5 to 13	22	1. Emotional functioning, 2. Physical functioning 3. Social avoidance, 4. Positive social attributes 5. Teasing/marginalization	No	Item content for Sizing Me Up was based on the published child obesity and HRQOL literatures and expert opinion.
Sizing Them Up (Parent proxy) Modi *et al.* (2008)	5 to 18	22	1. Emotional functioning (seven items) 2. Physical functioning (five items) 3. Teasing/marginalization (three items) 4. Positive attributes (four items) 5. Mealtime challenges (two items) 6. School functioning (one item) In addition, there were a separate set of developmentally appropriate items (six items) for parents of adolescents aged 14–18 years.	No	Item content for the original 35-item measure was based on the published child/adolescent obesity and HRQOL literature, as well as expert advice from three independent paediatric obesity clinicians and researchers.
Youth Quality-of-Life Instrument–Weight module YQOL-W Morales *et al.* (2010)	11 to 18	21	1. Self 2. Social 3. Environment	Yes	The development of YQOL-W was guided by the World Health Organisation's (WHO) conceptualisation of QoL. The content of the YQOL-W is based on over 50 in-depth interviews with African American, Mexican American and white youth (aged 11 – 18) rather than expert opinion.

3.3.2.2 Literature search B - Measurement properties of existing weight-specific measures

The findings from literature search B resulted in a total of 2036 unique hits, of which 123 articles were selected based on their title and abstract. The second round of assessment resulted in exclusion of a further 117 articles. Six articles, met the inclusion criteria; they assessed the measurement properties of four multi-attribute weight specific QoL instruments. Fifteen out of 117 excluded articles utilised relevant weight specific measures in the adolescent population. Reference checking of these articles resulted in the identification of one additional study that assessed the measurement properties of a weight specific instrument and met the inclusion criteria for the review. This study was added to the review, and thus a total of seven articles evaluating the measurement properties of five weight specific instruments were included in the review (see **Figure 3.3** and **Table 3.7**).

Figure 3:3 Results of literature search B



*Main reason for exclusion of studies:

- Relevant study but unavailable to obtain an English full text version (n=2)
- Adult or pre-adolescent participant population e.g. under 11 or over 18 years (n=18)
- Generic and non-weight-specific Qol instruments used in the study (n=55)
- The primary aim of the study was not to validate a weight specific measure (n=27)
- Relevant conference abstract but full text paper unavailable or not relevant (n=17)

Table 3:7 Summary of the studies included in the review (N=7)

Reference	Patient characteristics			Setting (Country)	Instrument
	N	Mean age (SD or range)	BMI-z score or BMI-SDS [8] (SD or range)		
Kolotkin *et al*. (2006)	642 (491 = final sample used in factor analysis)	14.0 (2.0)	1.5 (1.2)	Multiple sites where youth were participating in psychological research or clinical protocols (USA)	IWQOL-Kids
Modi & Zeller (2011)	263	15.1 (1.9)	43.1 (11.2) 2.6 (0.4)	Hospital participants seeking either weight management or bariatric surgery programme (USA)	
Moorehead *et al*. (2003)*	110	45 (19 to 65)	50 (32-92)	Patients participating in gastric bypass support groups (USA)	M-A QoL Q II
Zeller & Modi (2009)	141	9.2 (2.2)	31.8 (6.2) 2.5 (0.35)	Hospital based paediatric weight management programmes (USA)	Sizing Me Up
Modi & Zeller (2008)	220	11.6 (3.3)	36.7 (11.6) 2.6 (0.37)	Hospital based paediatric weight management programmes (USA)	Sizing Them Up
Morales *et al*. (2011)	443	14.7 (2.2)	27.1 (6.8)	Youths were recruited through community centres, schools, clinics and youth programmes (USA)	YQOL-W
Patrick *et al*. (2011)	133	15.4 (2.0)	34.0 (6.0) 2.1 (0.4)	Weight loss campers at 8 week Wellspring Camps (USA)	

*Identified from Collins *et al*., 2007 and Silberhumer *et al*., 2006. In the former the Moorehead-Ardelt Quality of Life Questionnaire II instrument was used with 9 adolescents under the age of 18 years and in the latter it was used with 50 participants aged 9 to 18 years.

[8] A BMI-z score or BMI-SDS indicates how many units (of the standard deviation) a child's BMI is above or below the average BMI value for their age group and sex. See page 9 of the thesis for further details.

Details about the investigated measurement properties and the methodological quality of the seven studies included in the review are summarised in **Table 3.8**. A summary of the results for each questionnaire is provided below.

Table 3:8 Summary of the quality and measurement property of each questionnaire

Instrument		IWQOL-Kids		M-A QoL Q II	Sizing Me Up	Sizing Them Up	YQOL-W	
Study		Kolotkin *et al.* (2006)	Modi & Zeller (2011)	Moorehead *et al.* (2003)	Zeller & Modi (2009)	Modi & Zeller (2008)	Morales *et al.* (2011)	Patrick *et al.* (2011)
Measurement properties* (E=excellent G=good F=fair P=poor)	**Internal consistency**	E		P	G	G	E	
	Measurement error		P			G		
	Reliability		G		P	G	F	
	Content validity	F		P	F	F	F	
	Structural validity	E		F	G	G	E	
	Hypotheses testing	G	F	P	F	F	P	F
	Responsive-ness					P		F

*Where the cell is blank the measurement property was not tested within the study

3.3.2.2.1 IWQOL-Kids

Two studies assessed the measurement properties of the IWQOL-Kids instrument. Kolotkin *et al*., 2006 investigated the internal consistency, content and structural validity. The Modi & Zeller, 2011 study assessed measurement error and reliability. Both of the studies investigated hypothesis testing by formulating a priori hypotheses about the correlation analyses they undertook. Neither of the two studies assessed responsiveness and thus there is no evidence on this for the IWQOL-Kids.

3.3.2.2.2 M-A QoL Q II

The Moorehead *et al*. (2003) study assessed all the measurement properties of the M-A QoL Q II apart from the measurement error, reliability, and responsiveness of the instrument. The measurement properties were assessed with an adult respondent population.

3.3.2.2.3 Sizing Me Up and Sizing Them Up

The study by Zeller & Modi (2009) assessed all but the measurement error and responsiveness of the Sizing Me Up instrument whilst Modi & Zeller (2008) investigated all the measurement properties of the parent proxy Sizing Them Up instrument. Data collection for both studies was undertaken in the same institution and over the same time period. Parents of adolescents aged 5 to 18 years enrolled in hospital based weight management programmes completed the Sizing Them Up instrument, whilst adolescents aged 5 to 13 completed the Sizing Me Up instrument.

3.3.2.2.4 YQOL-W

The measurement properties of the YQOL-W were assessed in the studies by Morales *et al*. (2011) and Patrick *et al*. (2011). All measurement properties apart from measurement error and responsiveness were assessed by the former whilst hypothesises testing and responsiveness were assessed in the latter. There is no evidence on the measurement error of the YQOL-W.

3.4 Discussion

The literature review addressed the four key objectives that were set. Firstly, the findings from two recent reviews provided evidence to support the negative impact of above normal weight status on QoL in the adolescent population. Secondly, two literature searches were undertaken that adopted search strategies that were successfully utilised in two recent reviews. Thirdly, the existing instruments that were, potentially relevant to, or had been used with, the adolescent population were identified. Finally, the measurement properties of existing weight specific instruments suitable for adolescents were assessed.

Since the Tsiros *et al*. (2009) and Griffiths *et al*. (2010) reviews were undertaken several additional weight specific instruments were identified: KINDL-Obesity module, M-A QoL Q II, OPOI Sizing Me Up, and YQOL-W, though none are preference based measures. Two studies picked up in the review identified the IWQOL-Lite, an adult weight specific instrument, where values are available (Brazier *et al*., 2004 and Hauber *et al*., 2010). The use of adult measures on a younger population is discouraged as HRQoL perceptions differ between these populations (Eiser *et al.,* 2001) and the applicability of their use in terms of content validity is uncertain. The *sexual life* and *work* dimensions contained in the IWQOL-Lite may have little relevance to the younger age group, which re-enforces the questionable applicability of using the measure on this age group. Although, Brazier *et al*. (2004) reported that whilst the

IWQOL-Lite can be mapped onto the SF-6D, there appears to be HRQoL information in the IWQOL-Lite not contained in SF-6D scores. Whilst the Hauber *et al*. (2010) study does not provide the necessary information required for the calculation of QALYs, using the IWQOL-Lite, the authors report that there is clear evidence that not all domains within the IWQOL-Lite are equally important to overweight and obese people. In addition, the results of this study confirm that some improvements within a particular domain or item are more important to subjects than other improvements within the same domain or item. This led the authors to conclude that weighting the individual items in the IWQOL-Lite by the importance of outcomes to overweight and obese subjects may provide a more meaningful evaluation of the effect of changes in weight on patient wellbeing than a non preference based measure of HRQOL.

Although a number of generic instruments suitable for adolescent have been used in the literature, a summary of the evidence has shown that *weight specific* measures are more suited to capturing changes that matter than generic measures, in studies of adult obesity (Kolotkin *et al*., 2006). Generic measures have been found to perform poorly in discriminating between different BMI subgroups on physical, psychological, and emotional dimensions of functioning relative to weight specific measures. Therefore, the value in using weight specific QoL measures to assess interventions for overweight and obese adolescents is strengthened.

3.5 Conclusion

In total seven weight specific instruments developed for use with the adolescent population were identified in the literature. All are completed via self report apart from the parent or carer proxy Sizing Them Up instrument. None of these measures however are suitable for conducting CUA. This is mainly because of the lack of preference values to enable the calculation of QALYs. The content of the two most recently developed instruments, the OPOI and YQOL-W, was informed by input from adolescents but measurement properties have been assessed in only the latter instrument. Plus, the content of the YQOL-W was informed by adolescents living in the US, whose views may differ from adolescents living in the UK. Due to all these issues, the use of an existing instrument in the generation of preference weights, though potentially a quicker option, was rejected.

The development of a new instrument that is applicable for the elicitation of preferences will address this problem as the present research aims to identify all the dimensions of QoL pertinent to adolescents who are obese or overweight. The limited evidence on the aspects of QoL important to overweight and obese adolescents were the main reason behind this

decision. It is anticipated that the newly developed instrument will be appropriate for measuring the weight specific QoL of adolescents aged 11 to 18 years. As this is such a diverse group it will not be clear whether one instrument for the entire age range will be suitable until empirical work is undertaken to inform the content of the new measure. The instrument will explicitly be designed for use in economic evaluation (meeting the constraints regarding the number of items that can be utilised in a preference valuation study, as discussed in the next chapter).

Chapter 4
Key stages for deriving a preference based measure

4.1 Introduction

The findings discussed in **Chapter 3** showed that existing instruments are not appropriate for undertaking CUA of weight management interventions targeting the adolescent population. The review highlighted the need for the creation of a new weight specific instrument with the primary intended purpose of carrying out economic evaluation. The key phases when developing a new preference based instrument for CUA involve: a) *the measurement of weight specific QoL*: the identification and collection of items (questions) that form the content of a new weight specific QoL instrument together with, b) *the valuation study*: the elicitation of preference weights for a subset of states described by the instrument informing the assignment of weights for all of the possible states. Part 1 of the three part process utilised in the thesis required the creation of a weight specific LLDS (a preliminary list of items crafted from consultations with adolescents and specialists). In Part 2 the instrument was refined making it appropriate for the elicitation of preference values, leading to the derivation of a new reduced form weight specific descriptive system. At this point psychometric testing of the measurement properties of the reduced form instrument was undertaken. Obtaining a scoring algorithm or population value set for the states defined by the new instrument, was undertaken in Part 3 and involved: a) a representative sample of the general population being asked to value a selected sub-set of states defined by the reduced from classification system and b) use of econometric models to predict utility values for all possible states defined by the classification system. Before presenting a detailed account of the empirical studies undertaken in the thesis, this chapter provides an overview of some of the key issues that arose from the three part process described above. Justifications for the key choices that were made prior to commencement of data collection are also presented. The topic of obtaining values for health states to aid resource allocation is large and involves a number of challenges and methodologies. This chapter purposefully focuses on the issues surrounding the methods that were utilised within the thesis.

4.2 Core questions for consideration

The current work takes the perspective that the instrument under development is for use in health care resource allocation decisions for adolescent weight management interventions. Therefore it needs to fully account for how weight status impacts upon different aspects of adolescents' QoL by capturing changes in dimensions that are important, appropriate, and relevant to them. A number of core decisions are necessary prior to undertaking the empirical studies that facilitate the creation of a weight specific preference based instrument. These are listed below and build upon the issues raised by Brazier *et al*. (2007):

1. What concept of QoL or HRQoL should be used?
2. What methods and whose views should used to develop the descriptive system?
3. What technique should be used to elicit preference values?
4. Who is (are) the relevant population(s) for providing the preference values?

The first two questions are related to Parts 1 and 2 of the development phase in the aforementioned three part process and need to be considered for the construction of the descriptive system. The last two questions are concerned with Part 3 where the valuation of the descriptive system is carried out.

4.2.1 What concept of QoL or HRQoL should be used?

Health has been given various definitions but one of the most well known in the field of health economics, is the statement in the Constitution of the World Health Organisation (1948); that health is *a 'state of complete physical, mental and social wellbeing, not just merely the absence of disease or infirmity'* (WHO, 1948). This definition has influenced the development of numerous QoL and HRQoL instruments used in the field of health economics, although it is a very broad concept of health and not easily operationalised. In the health literature, terms such as health, health status, HRQoL and QoL are used to mean different things by different instrument developers. Terminology is not the focus of importance here, but instead, the significance lies within the distinction between the content of the different instruments. For example, when describing the benefits associated with health care, whether social relationships are counted as HRQoL or QoL is not important. What is important is whether this domain is used to measure benefits associated with a particular health intervention (Brazier *et al*., 2007).

A key question is what aspects of Qol should be covered by the measure. A narrow definition can be viewed as an impairment or symptoms based ('within skin') approach, whilst a broader concept of QoL describing the way an impairment or disability affects a person's participation ('beyond skin') is based more on social functioning. Numerous measures adopt different approaches, for example the HUI (Furlong *et al*., 2001) utilises the first approach, whilst the latter is adopted for the SF-36 (Ware & Sherbourne, 1992). Both approaches have their strengths and weaknesses, though neither approach is thought of as superior to the other (Brazier *et al*., 2007). The descriptive system developed in the thesis will aim to contain information concerning the benefits of weight management interventions in terms of their impact on relevant and important dimensions of QoL, as perceived by adolescents. This is because it has been suggested that when designing a HRQoL instrument suitable for the younger population it is important to ensure items correspond to experiences, activities, and contexts directly relevant to the (age of the) sample (Matza *et al*., 2004). This may result in the content of the descriptive system capturing not only HRQoL dimensions, but also broader QoL dimensions (such as self-esteem). This is in line with the WHO definition of health, and the Food and Drug Administration (FDA) guidance for the development of patient reported outcome measures (FDA, 2009).

4.2.2 What methods and whose views should be used to develop the descriptive system?

In terms of the method that should be used in deriving the content of the descriptive system, two available options include: the so called, *top down* and *bottom up* approaches. A number of the outcome measures developed in the past have utilised a *top down* approach, using the literature around the views of experts (usually clinicians or researchers) and existing instruments, to inform the content of the descriptive system (Fitzpatrick *et al*., 1998). The top down approach was used to develop the content of five of the seven weight specific instruments that were identified in the literature. In recent times, the developers of new instruments have utilised a contrasting approach known as a *bottom up* approach, where-by qualitative methods were undertaken, typically one-to-one interviews and focus groups with the population under study, to inform the content of their instruments (for example see: Carlton, 2011, Doyle *et al*., 2011, Morales *et al*., 2010 and Stevens, 2009 and 2010). The benefits of the *bottom up* approach are numerous: the final measure under development will have appropriate language and terminology for the population of interest increasing the content validity. This approach should also improve responsiveness to change, as it will ensure that outcomes of relevance to the patient are included in the instrument (McColl, 2005).

The development of the content for the new weight specific instrument will combine the two approaches. The *bottom up* approach will be the primary source of informing the content of the new instrument due to its many advantages. Here qualitative interviews with adolescents will be undertaken to identify the aspects of QoL affected by weight status, facilitating the creation of the LLDS. The methods employed will be guided by the Stevens & Palfreyman (2012) study that provides a description of how qualitative methods can be used in the development of descriptive systems of preference-based measures of HRQoL. In addition to this, items crafted from the qualitative interviews with adolescents will be refined via discussions with experts in the field of adolescent obesity (the *top down* approach). This will have the advantage of potentially highlighting issues not identified from the interviews with adolescents. It is not anticipated that there will be disagreement between the views of adolescents and experts, as experts views will primarily be sought with regard to the acceptability and clarity of the items and response options generated for the LLDS.

The consideration of the issue of whose views should inform the content of a new instrument also needs to be considered. It is recommended, in the review by Fitzpatrick *et al*. (1998) of PROMs used in randomised controlled clinical trials, that the reflection of the preferences and experiences of the patient when deriving new outcome measures is important. In situations where it is not possible to directly elicit patients' own views, the authors state that people who are close to the patient are well placed to provide a proxy description of the patient's experiences and their views should be taken into account when deriving the content for a descriptive system. This will provide 'face validity' to the content of the derived measure. Face validity involves an assessment of the instrument in terms of whether it actually captures the dimensions of QoL that are relevant to the population under study.

The content of descriptive systems for weight specific measures has frequently been developed from the views of adult patients, doctors and experts in the field (IWQOL-Kids (Kolotkin *et al*., 2006), Sizing-Me-Up (Zeller & Modi, 2009) and KINDL-obesity module (Ravens-Sieberer *et al*., 2001), for example). Other potentially relevant populations that could provide information on the content for the instrument under development are: adolescent patients, parents, or primary carers, paediatric health care professionals, the general adult or adolescent population and health care decision makers. Although all the populations outlined above could provide information on what they feel are the appropriate, and relevant, dimensions for informing the content of a new weight specific instrument, the most relevant, and appropriate population to inform the content of the descriptive system are adolescents themselves, as they will be the users of the measure. It is increasingly being recognised that

young people have the capacity to participate in decision making, although they may have different competencies to adults (Mayall, 2002). Both adolescents who are treatment seekers and a general sample of adolescents will be used to inform the content of the descriptive system. Individuals with experience and expertise in the area of adolescent obesity will also have relevant and informed views about what is appropriate to consider in this context. The views of individuals who work with adolescents in weight management services will also be obtained in informing the content of the new descriptive system. To avoid the complications that might arise if there is disagreement between the views of adolescents and experts, the views of adolescents will take precedence over those of experts. The key input from experts will be to add to the LLDS item set and check that items wordings are clear and unambiguous.

4.2.3 What technique should be used for valuing health?

In **Chapter 2**, a description of each of the valuation techniques referred to below was provided. Usually VAS, TTO and SG are the elicitation techniques used with the adult population. The choice between techniques is difficult: VAS is simpler to understand and implement, but TTO and SG provide clearer anchor points at zero (the *dead* state) and one (the *full health* state – where there is no dysfunctionality in any aspect of QoL). TTO and SG also require individuals to make tradeoffs between outcomes, whereas VAS does not. In the context of informing cost per QALY analysis, both health economists and NICE favour the choice-based methods of SG or TTO (Brazier *et al*., 2007). It should however be noted that both these techniques also have limitations in their application to health state valuation. SG is based on the axioms of expected utility theory, but there is little empirical evidence to support this theory in the context of health state valuation. Additionally, there is concern over the empirical basis of the TTO technique, as duration effects (the length of time spent in a particular health state) and time preference (the general desire for an individual to have good health sooner rather than later) can have an impact on the elicitation of TTO values. Further information around the biases of these two techniques is provided by Bleichtrodt (2002).

Recent studies have utilised DCEs, as opposed to SG and TTO, in the anchoring of utility values, on the full health – dead scale (Bansback *et al*., 2012 and Norman *et al*., 2012). DCEs can be used to indirectly elicit measures of value derived from choosing a particular option from a choice set, or in this context a particular health state from the available alternatives (Bansback, 2010). Historically, a drawback of the application of DCEs for estimating utility values, suitable for use in QALY calculations, was that attributes were valued relative to each other. For example, Ryan *et al*. (2006) assumed that the best health profile, described by a preference based outcome measure for social care for older people, was equivalent to the full health state (equal to one) and the worst state was equivalent to dead (equal to zero). Assuming that the worst state described by a preference based instrument has its limitations. Studies using SG and TTO have shown that the health states individuals consider to be equal to dead varies and is dependent on the descriptive system under study (Bansback *et al*., 2010). The Bansback *et al*. (2012) and Norman *et al*. (2012) studies have shown that incorporating an additional 'years of survival' (duration) attribute to the design of a DCE facilitates the calculation of utility values that can be anchored on the full health – dead scale. Although DCEs are more complex to design and analyse, they are easier to implement without the presence of an interviewer and can be administered using an on-line survey. This means that a large amount of data can be collected over a short period of time. By requiring individuals to choose rather than rank a set of health states, imitating real world decision making where individuals are faced with the notion of sacrifice, the results of DCEs are more consistent when compared to ranking. Another essential advantage of DCEs over ranking is that they do not require the assumption of the *independence of irrelevant alternatives*: that the ordering of a pair of health states, within a choice set, is *not dependent* on the other states being considered (Bansback, 2010). The DCE valuation technique will be utilised in the thesis given the time and resource constraints and the novel approach proposed by Bansback *et al*. (2012) will be implemented.

4.2.4 Who should provide the preference values?

The last key decision that needs to be addressed regards *who* should value the states defined by the descriptive system. Preference values for health states can be obtained from a number of different sources, such as: patients, carers, health professionals and the general public. There is a paucity of research into the feasibility of asking adolescents to make these tradeoffs (Ratcliffe *et al*., 2011). Moreover, differences in the values elicited from alternative sources is well reported (Brazier *et al*., 2007), and has important implications on the estimated benefits attributed to competing interventions. Some studies have shown that patients who have firsthand experience of a given health state are more likely to place higher values on dysfunctional health states than members of the general public, who do not have a similar

experience (Dolan & Kahneman, 2008). The potential differences in the magnitude of the incremental gain of an individual moving from a dysfunctional health state to full health will be larger using general public, rather than patient preference values. This may have a significant effect on the resulting cost per QALY estimate and thus resource allocation decisions. Brazier *et al*. (2007) and Dolan & Kahneman (2008) present the reasons for the observed differences in patient and general population values. The key issues listed include: poor descriptions of hypothetical health states, response shift (where individuals use different internal standards or changes are observed in an individual's expectations) and adaptation (where individuals adjust to new or changed circumstances). Conventionally, preference values for health states are obtained from a representative sample of adult members of the general public, trying to imagine what a given health state will be like (NICE, 2013) and so the elicitation of preference values for weight specific QoL states will follow this convention. Though, it should be noted that, preference values obtained from adults may be inconsistent with those obtained from adolescents and this issue is discussed further in **Chapter 9**.

4.3 Conclusion

Although the reasoning behind the key choices that were made prior to commencement of data collection, regarding the development of the new weight specific preference based instrument have been discussed, it is important to keep in mind that there are limitations associated with every potential course of action. The potential limitations associated with each of the empirical studies undertaken in the thesis and the implications of the decisions made at this stage will be discussed in later sections in the thesis. The implications of these limitations on the overall findings are discussed in **Chapter 9**.

Chapter 5
Creation of the weight specific LLDS

5.1 Introduction

Following the discussion about the key stages involved in the creation of a new preference based instrument, this chapter reports on Part 1 of the three part process detailed in the previous chapter. In this first empirical study (Study 1) the creation of the LLDS is undertaken. The details regarding the development of the content of the new weight specific instrument are provided and discussed.

5.2 Methods

Study 1 consists of three stages. In Stage 1, qualitative interviews were used to elicit the views of adolescents regarding how different aspects of their lives were affected by their weight status. This information was used to craft the weight specific LLDS. Following this, in Stage 2, specialists in the field of adolescent obesity guided the refinement of the items. The last stage saw the evolution of the final item set through further qualitative interviews with adolescents, the mapping of items onto QoL dimensions identified from the literature and finally consultation with the thesis advisory group. **Table 5.1** presents an overview of the aims, methodology, and outputs for each of the aforementioned stages.

Table 5:1 Study summary

	Stage 1: Consultation with adolescents	**Stage 2: Consultation with specialists**	**Stage 3: Mapping* items to dimensions of QoL, consultations with adolescents and the thesis advisory group**
Aims	To identify aspects of QoL affected by weight status	To finalise the item set and identify appropriate response options	
Methods	Interviews with treatment and non treatment seeking adolescents Thematic framework analysis of interview transcripts to form a preliminary item set for the new questionnaire	Appraisal of preliminary item set by weight management staff and researchers with expertise in the field of adolescent obesity	Mapping items onto dimensions of QoL identified in the literature Interviews with treatment seeking adolescents to finalise item and response options wordings Consultation with the advisory group to produce the finalised item set and response options
Outcome	Preliminary item set	The weight specific LLDS with associated response options	

* Involving the grouping of items into dimensions of QoL that have been identified in the literature

5.2.1 Stage 1: Consultation with adolescents

5.2.1.1 Participants

One-to-one interviews were conducted with 16 adolescents (as sensitive issues are more likely to be raised within a confidential setting) enrolled in two Leeds based weight management programmes (henceforth referred to as services A and B) to elicit their views on how their weight status might impact upon different aspects of their lives. Both programmes offer a multifaceted weight management intervention addressing both health and lifestyle behaviours and are accessible by way of GP referrals. Staff identified families that could be approached to take part in the research and were given a pack containing: an information sheet and consent form for the adolescent (see **Appendix 5.1**) and, an information sheet and opt-out form for carers of adolescents (see **Appendix 5.2**). Adolescents were sampled purposively using the primary sampling criteria based on gender and age, the latter consisting of the younger adolescent group (11 - 14 year olds) and the older adolescent group (15 – 18 year olds). Purposeful sampling was undertaken to ensure that the views of the full range of the obese and overweight adolescent population were incorporated. It was assumed that all adolescents recruited from weight management services had a clinical indication of having above the normal BMI permitting referral to the services. Treatment seekers participating in the one-to-one interviews were recruited mainly from weight management service A as there were a limited number of adolescents within the age range that was required for the study from weight management service B (as most of the participants in this service were under 11 years old). Weight management staff identified potential participants that could be approached to take part in the study. The potential of selection bias is greater where participants were identified by staff, as opposed to utilising random selection, thus the limitations of this method of recruitment should be noted.

Views of a school general sample of five adolescents were elicited about the impact of weight status on QoL by way of focus group (FG) interviews. FGs extract a range of views within a group context, potentially allowing access to a larger amount of information, in comparison to one-to-one interviewing (Richie and Lewis, 2003). This general sample was used to incorporate issues that may not have been picked up by treatment seekers. Obtaining the views of non-treatment seekers was important as these individuals could be future consumers of weight management services. Also they may have strong opinions about weight status and how it affects their life. Participants were recruited from the school setting and were identified by school staff. The potential for selection bias, as potential participants were not randomly selected, should also be noted here. The school was based in West Yorkshire and

selected because of its willingness to participate in the research. Information sheets and consent forms were given to potential participants and signed consent forms were obtained from both parents/carers (see **Appendix 5.2**). Consent was obtained from both parties as weight related discussions taking place in the school setting could potentially stigmatise some adolescents resulting in a higher potential risk of distress. Purposive sampling was based again on age and gender. Adolescents' BMI was not measured, as the aim of this part of the study was to obtain the views of a general sample of adolescents regardless of their weight status.

Ethical approval was provided by the University of Leeds local research ethics committee (Ref - HSLTLM/10/008).

A total of 20 families from weight management service A (nineteen current users and one on the waiting list) were approached to take part in the study and interviews were carried out with 15 adolescents. Two families were approached from weight management service B and one adolescent agreed to take part in the study. In total 16 one-to-one interviews with adolescents were conducted. The interviews varied in length from 16 to 60 minutes (mean = 31 minutes (SD=13.21)). Five non-treatment seeking adolescents, three boys and two girls aged 11 to 14 years, took part in the FG interviews by school staff. Unfortunately the timing of interviewing meant that there was a clash with summer examinations and thus older adolescents, aged 15 to 18 years could not be recruited into the study. One FG interview was arranged with five of adolescents falling into the younger age group. The focus group interview with the school sample lasted for 47 minutes. **Table 5.2** provides a summary of the participants that were interviewed.

Table 5:2 Characteristics of respondents in the interviews

	Participant ID	Age	Duration of interview (mins)	BMI (baseline)	BMI-SDS (baseline)
Stage 1: 1-to-1 interviews (n=16)	**Girls – Younger age group (11-14 years)**				
	GL_Y_1	11	16.22	27.16	2.80
	GL_Y_2	12	20.37	27.10	2.43
	GL_Y_3	11	25.37	30.04	3.20
	GL_Y_4	12	25.5	NA (on waiting list)	NA (on waiting list)
	GL_Y_5	11	33.34	25.2	2.27
	GL_Y_6	11	35.06	42.2	3.56
	GL_Y_7	12	40.55	30.72	2.96
	Girls - Older age group (15-18 years)				
	GL_O_8	15	16.29	35.11	3.16
	GL_O_9	16	36.45	37.70	3.38
	GL_O_10	15	53.02	30.80	2.62
	Boys - Younger age group (11-14 years)				
	BO_Y_1	11	16.51	27.57	2.87
	BO_Y_2	14	31.34	30.60	2.86
	BO_Y_3	14	34.05	29.76	2.75
	BO_Y_4	11	60.1	20.77	1.53
	Boys - Older age group (15-18 years)				
	BO_O_5	15	17.27	33.48	3.08
	BO_O_6	16	42.29	40.81	3.67
Stage 2: FG interviews (n=1)	**Focus group (2 girls & 3 boys aged 11-14 years)**				
	BO_Y_FG		47.49	NA (perceived to be normal weight for age and gender)	NA (perceived to be normal weight for age and gender)

5.2.1.2 Interviews

Interviews with treatment seekers were offered in the most convenient setting to participants, a private room at the weight management venue, at home or at the University. A semi structured interview was carried out with adolescents, using a topic guide in order to identify the impact of weight status on their lives (see **Appendix 5.3**). To initiate the interview, participants were asked general questions about their perceived weight status. After this initial discussion, participants were then asked about the different things that they do in their life both when they are in school and in evenings, weekends and holidays. They were then asked to describe the effects of their weight on any of the activities that they carried out. The final section of the interview utilised the existing literature to: a) assess adolescents' views on the impact of weight status on aspects of QoL identified in the literature and b) confirm that all aspects of QoL potentially affected by weight status had been covered within the interview. Probing questions were used to identify how different activities were affected by weight status throughout the interview.

A similar approach was taken in the interviews with the non treatment seekers. One difference was the use of silhouettes to communicate different body sizes for boys and girls (see **Appendix 5.4**). Adolescents were asked to consider the body size synonymous to obese weight status and were asked to 'Imagine you know someone that is the size in example (4) – it could be a friend at school or at home, or it could be a relative. How do you think they would be affected by their size'? Free conversation was encouraged and a flip chart was used to note key discussion points.

All interviews were undertaken by the candidate. The use of closed questions was avoided where possible. Open questions were posed to obtain depth and explanation in respondents answers. For example questions such as 'did that affect you' were avoided instead questions like 'how did that affect you' or 'how did that make you feel' were used. Each interview was audio-recorded with permission of the interviewees and transcribed verbatim for data exploration and analysis.

5.2.1.3 Analysis

Data obtained from the interviews was subjected to qualitative thematic analysis to identify aspects of QoL affected by weight status as reported by participants. Data were sorted and managed using the Framework approach, developed by the National Centre for Social Research (Richie and Lewis, 2003). The approach has been successfully utilised in a similar study by Stevens (2009), who developed a generic preference based questionnaire for children aged 7 to 11 years. The framework method provided a systematic thematic way of summarizing and classifying data. For the current study, the aim of the analysis was to produce a theme and case based chart that summarised all the data into one matrix consisting of cases (represented by each row) and themes (represented by each column). The matrix summarised and synthesised the data generated from the interviews whilst retaining the terminology and language used by participants. A short summary of the steps involved for the Framework approach is provided below (see Richie and Lewis (2003) for a more detailed explanation):

- *Step 1:* Re-familiarisation of issues identified in the interviews by re-reading transcripts and re-listening to audio recordings
- *Step 2:* Identification of recurring themes and ideas from carrying out Step 1
- *Step 3:* Devising a thematic framework from the information generated in Step 2 by grouping together themes and ideas identified in Step 2 into 'Main themes' and 'Sub-themes'
- *Step 4:* Going through each interview transcript and code using the main themes and sub-themes identified in Step 3

The process of coding can be carried out using qualitative software or other non-qualitative software packages. In the current study both types of packages were used (NVivo9 QSR International and Microsoft Office - Word 2010). The qualitative analysis was carried out in two phases. In the initial phase of analysis, data focusing on adolescents own experiences were assessed and in the second phase the data generated from the interviews that was informed by the literature was analysed. This approach was taken in order to maintain a neutral stand point on the first phase analysis of the interviews. This meant that the findings from the literature were put to one side and the 'pure' data from adolescents themselves drove the themes and sub-themes. For this analysis, the themes and sub-themes were coded in Word 2010 and highlighted in different colours. All of the information that was highlighted was then charted by producing a case and theme based matrix in Excel. A case (row) in the matrix was either a one-to-one or FG interview and themes (columns) indicated a particular aspect of QoL identified from the interview data. In analyses carried out on the data generated from the second part of the interviews, informed by the literature, the transcripts were structured according to pre-identified themes identified from the literature. This made it easier to produce the case and theme based charts using the Nvivo9 qualitative software (NVivo9 QSR International). Transcripts were formatted and read into the programme to aid efficient analysis. The matrices were used to develop wording for the new weight specific questionnaire. The preliminary items that were crafted aimed to maintain the language and terminology used by participants.

Saturation

Data saturation (the point at which the collection of new data does not shed any further light on the issue under investigation, Richie and Lewis, 2003) was assessed with the aid of a saturation matrix (Brod *et al*., 2009). For the current study, this was the point at which no new aspects of QoL were identified as being affected by weight status, by carrying out more interviews with adolescents. Information collected from the interviews was summarised using

the saturation matrix. The rows of the matrix consist of each individual interviewed (in chronological order) whilst the columns represented general aspects or 'themes' of QoL. Cells in the matrix were highlighted in yellow if a particular respondent had identified a particular aspect of QoL being affected by their weight status. This point was reached on completion of the 13th one-to-one interview. Three further one-to-one interviews were carried out but no new information emerged from conducting these interviews. In the FG interview, the majority of issues and themes previously discussed in the preceding one-to-one interviews were discussed. The FG interview confirmed that data saturation had been reached as no new themes or issues emerged. The saturation matrix is presented in **Appendix 5.5**.

Validation

The data analysis was closely guided by input from supervisors and advisors experienced in the methodology and conduct of qualitative analysis. The coding of themes and sub-themes was an iterative process that was closely overseen by the primary supervisor. The final list of themes and sub-themes used in the coding of transcripts was also reviewed by the same supervisor.

5.2.2 Stage 2: Consultation with specialists

Staff members from WATCH IT (one of the weight management services used in the recruitment of adolescents detailed above), a Leeds based community weight management service, were invited to comment on the preliminary item set, within one of their weekly staff meetings when all members of staff were available to provide input. The Seattle Quality of Life Group (Sea QoL Group, based at the University of Washington), developers of the YQOL-W (Morales *et al*., 2011), a non-preference based weight specific QoL module for adolescents (see **Chapter 3** for a detailed account of the YQOL-W) were also invited to comment on preliminary item set via e-mail. To guide and coordinate the feedback from both groups of specialists, they were asked to:

- Identify items adolescents are unlikely to be willing to answer, or unlikely to answer truthfully
- Identify unclear or ambiguous questions
- Identify any important issues that have been missed out of the current item set and to give an explanation of why the issue is important

5.2.3 Stage 3: Mapping items to dimensions of QoL, consultation with adolescents and thesis advisory group

This stage involved two processes that were carried out simultaneously involving: a) the mapping of items to dimensions of QoL identified in the literature and b) additional qualitative work with adolescents to finalise the wording of items and generate response categories that were appropriate for the younger population. Both tasks were carried out to facilitate the collapsing of similar items to form a smaller pool of items. The thesis advisory group (a detailed description of the membership of the group is given in **Appendix 5.6**) was used as a sounding board to reflect on the overall results in order to finalise the item set and associated response options. The purpose of the advisory panel was to provide guidance on the proposed methods that were to be undertaken in the thesis and comment on the key findings. In addition to the guidance provided by the thesis supervisors, the group included members who could give guidance on: the overall methodological demands of the empirical studies, qualitative field work and analysis, the application of psychometric assessment and Rasch analysis in the development of a new instrument, and the practicalities of developing an instrument for economic evaluation for the younger population (see **Appendix 5.6** for further details).

5.2.3.1 Mapping

After the consultation process with the advisory group the refined item set was agreed. Further reduction of the item pool was necessary because the number of items forming a preference based instrument is constrained (ideally 5 to 9 items, see Brazier *et al*. (2007)). As such it was felt that mapping common items onto dimensions of QoL could further reduce the number of items, whilst still adequately capturing the full range of aspects of QoL affected by weight status. The mapping process was informed by the findings of Fitzpatrick *et al*. (1998). The review summarised the range of dimensions (or domains, the terminology is used interchangeably) of QoL assessed by patient reported outcome measures used in clinical trials. The item mapping was used to group each of the final agreed items onto suitable QoL dimensions. These QoL dimensions were also used to guide the interviews with adolescents in order to finalise appropriate item wordings.

5.2.3.2 Consultation with adolescents

Sampling technique

Two FG interviews with adolescents were undertaken to generate finalised item wordings and appropriate response options. FG interviews were conducted with treatment seeking adolescents. These study participants were recruited from a third Leeds based weight management programme (henceforth referred to as service C) accessible either by GP referral or self-referral. Adolescents taking part in the annual residential weight management summer camp were recruited for the FG interviews. The residential camp offers a six week programme where adolescents are free to choose whether they take up the full six week programme or only part of it. Adolescents were approached and recruited via staff delivering the weight management programme, information sheets and consent forms were provided to the FG participants. These forms were very similar to those used for the one-to-one interviews with adolescents recruited from weight management services A and B (see **Appendix 5.1 and 5.2**). Owing to the advice given by staff, the FGs were split by gender. Staff felt that adolescents would participate more fully in the FG if they were familiar with one another, as the camp activities and lifestyle classes were separated by gender. The purposive sampling criteria were based on gender and age, in line with Stage 1. Each FG was made up of five male and female participations aged 11 to 18 years and 11 to 16 years respectively.

Interviews

The interviews took place at the summer camp premises. The semi structured FG interviews were designed to formulate appropriate question wordings and ordering of response categories. At the beginning of the interview adolescents were shown groups of activities (as identified in the preceding mapping process) and asked to generate question wordings that would adequately summarise the different activities making up a particular item grouping. In order to allocate appropriate wordings for emotions, adolescents were asked to assess the different emotional descriptions generated from the interviews in Stage 1 and identify emotions they thought of as overlapping or identical, those thought 'most important' and the best wording that should be used to describe a particular feeling. Wordings for severity and frequency response options were obtained from previous investigations of the literature and interview data from Stage 1 (see **Table 5.3**). A randomly ordered set of cards containing different wordings for severity and frequency response levels was provided and participants were asked to rank options from best to worst. Ties between words perceived by adolescents to be of a similar level were allowed (here, adolescents chose their preferred wording). To conclude the FG interview, adolescents were asked if there was any aspect of QoL they felt

was affected by weight status that had been omitted from the previous discussions. Open-ended questions were posed to obtain the reasoning behind adolescents' responses and debate was encouraged so that consensus could be reached. For example, where it was suggested by a member of the FG to omit a particular response option, questions like 'why is that?' were directed to the individual who made the suggestion, or 'what do the rest of you think about that?' directed to the rest of the group.

Flipchart paper was used to record the feedback from adolescents as the FG interviews were taking place. Suggested question wordings were written down on the flipchart paper as well as the results of the ranking of the severity and frequency response options. As with the interviews that took place in Stage 1, the FGs were audio-recorded with permission of the interviewees and transcribed verbatim.

Adolescents taking part in the FG interviews were fully engaged throughout the interview process. Positive contribution from members of each FG was self-regulated. Adolescents took it upon themselves to flag up any irrelevant or divergent comments from others or chastised *clownish* behaviour. Where there was any confusion with regard to the tasks that were undertaken, adolescents took the responsibility of explaining what was required to each other. None of the adolescents asked to stop the interview early or asked to be excused from the FG. For both interviews one member of staff from the weight management programme sat as an observer in the interview room. The duration of the FGs ranged from 61 to 64 minutes.

Table 5:3 Severity and frequency response options ranked by FG participants

Severity		**Frequency**	
very	really	never	normally
quite a lot	a lot	rarely	quite often
a bit	quite a bit	Not often	very often
a little	a little bit	usually	often
very little	not at all	sometimes	always

Analysis

The exploration and assessment of the data generated from the interviews, in the most part, took place as the FGs were being carried out. The aim of the FG interviews was to generate appropriate item wordings and associated response options which was being formed and noted down on the flipchart paper throughout the interviews. Confirmation of outputs from the FGs was completed by listening to audio recordings and reading through transcripts.

5.2.3.3 Consultation with thesis advisory group

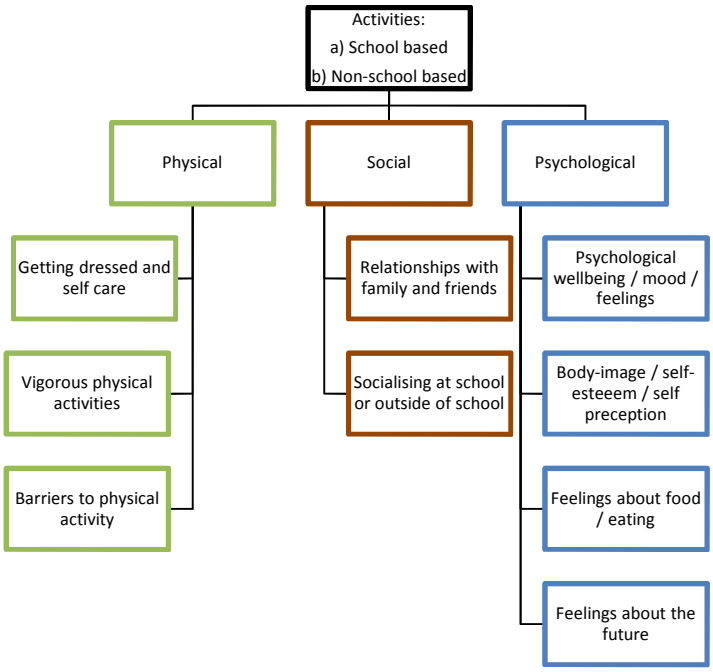
Reflecting on the results of the preceding stages and the generation of a finalised item set with appropriate response options was undertaken by feeding back findings to the advisory group. Members of the advisory group were presented the items and associated response options generated after the interviews in Stage 2 and were asked for their views on it. Feedback from the advisory group proved to be a very valuable contribution for finalising the item set and associated response options.

5.3 Results

5.3.1 Stage 1: Consultation with adolescents

Emergent themes, for the first section of the interviews, were agreed from the coding strategy developed by the candidate and reviewed by the primary supervisor. A thematic framework was identified and is presented in **Figure 5.1**. The major emergent themes are described below and illustrated with direct quotations from the qualitative interviews. Pre-identified themes from existing instruments were used to create the theme based matrix for the data generated from the second part of the interviews informed by the literature, themes included: Physical Activity, School, Psychological Health, Body Esteem, Relationships, Social Functioning, Eating, and Future. There was good overlap between the themes that emerged solely based on the interview data detailing adolescents own experiences of how weight status affects their lives and the pre-identified themes obtained from existing weight specific non-preference based measures. As demonstrated by **Figure 5.1**, above normal weight status impacts upon many different aspects of adolescents' lives. Treatment seeking participants differed in terms of how their weight impacted upon their day to day lives and the relative importance of different aspects on their overall QoL. The non-treatment seekers felt that weight status would have quite a major impact on all aspects of QoL. Data generated from treatment seekers and non treatment seekers complemented one another and the main emergent themes are described below.

Figure 5:1 Thematic framework



PHYSICAL ACTIVITY

Many participants spoke about the impact of weight status on physical activity, particularly the barriers it would impose. Difficulties related to performing physically demanding tasks such as running or playing sports (e.g. basketball, cycling or football) were reported by several participants. When performing these types of activities adolescents described having body pain or ache, being short of breath or getting tired quickly.

Example 1

Young person: ***"The odd times like my legs ache a little bit but that's only if like we've been walking quite fast, if we're going to be late for school we walk quite fast...."***

(Girl aged 11-14 yrs, Baseline BMI 27.1 (*BMI-SDS = 2.4*))

Example 2

Interviewer: *"When you walk to the bus stop do you ever think that your weight affects you when you're walking to the bus stop?"*

Young person: ***"It's like I get tired"***

(Girl aged 11-14 yrs, Baseline BMI - not available)

More general discussions also took place around physical activity. Some adolescents spoke about having difficulty keeping up with others when walking around or when bending over. Others identified having problems keeping up with others when performing physical activity.

Example 1

Interviewer: *"Do you ever think that your weight affects you when you're in games lessons?"*

Young person: ***"Yeah cause I can't do things as well as I can with others"***

Interviewer: *"What types of things can't you do as well as others?"*

Young person: ***"Run as fast as others, as long as others, and stuff like that"***

(Boy, aged 11-14 yrs, Baseline BMI 30.6 (*BMI-SDS = 2.9*))

Example 2

Interviewer: *"S*o thinking beyond school, when you think about your weekends or your normal kind of holidays, what are the types of things that you can think of that are affected by your weight?"

Young person: ***"Oh every time I bend down, I feel like that (pointing at belly) gets in the way.....and I'm like, and I'll bash it backwards and then when I'm reaching it'll come out and my back'll go and I'm like aargh***"

(Boy, aged 11-14 yrs, Baseline BMI 20.8 (*BMI-SDS = 1.5*))

SOCIAL FUNCTIONING

The impact of weight status on socialising manifested itself in the way adolescents were treated by others either in the school or non-school setting, whether or not they wanted to socialise because of the way they looked and also in terms of their relationships with friends and family.

Example 1

Young person: ***"...if like, I'm sitting next to, someone who I don't know as well, and we just start talking, I don't, I'm not as confident cos I don't feel confident, so I'm not gonna act it with them"***

Interviewer: *"Do you think that your weight ever affects your ability to make new friends then?"*

Young person: ***"Yeah, I kind of erm, stop myself sometimes"***

(Girl, aged 16-18 yrs, Baseline BMI 30.8 (*BMI-SDS = 2.6*))

PSYCHOLOGICAL HEALTH

There were many issues raised by adolescents around how they felt about the way they look as well as how they felt inside. Numerous variants of the same emotion were given, for example, feeling upset or self-conscious.

Example 1

Interviewer: *"So thinking through all the activities that you've mentioned on a typical school day, is there anything that you can think of that we've not talked about that you think is affected by your weight?"*

Young Person: ***"Getting dressed, could be one, but we've already covered that, cause I could have to get bigger clothes"***

Interviewer: *"Okay, so tell me a bit more about that"*

Young Person: ***"Well once I got this shirt that wasn't big enough so I had to take it back and get another one"***

Interviewer: *"How does that make you feel?"*

Young person: ***"Upset because the one didn't fit me and I wanted that one"***

Interviewer: *"And how do you find shopping in general for clothes?"*

Young Person: ***"(Feel) self-conscious...about what other people will think"***

(Boy, aged 11-14 yrs, Baseline BMI 30.6 (*BMI-SDS = 2.9*))

Example 2

Young person: ***"I sometimes feel upset, and feel like sometimes I feel upset that I can't do stuff like with my friends, school friends do"***

(Girl, aged 11-14 yrs, Baseline BMI 27.1 (*BMI-SDS = 2.4*))

The impact of weight status on school in terms of concentration or not being able to complete school work was mentioned by participants taking part in the school FG. These adolescents in particular thought that above normal weight status weight status would affect concentration on school work due to bullying. Treatment seeking participants pointed out that the main aspect of their school life that was affected by their weight was when they had to take part in physical activities.

Example 1

Interviewer: *"I want you to imagine that you had a friend or a relative that has this body shape, and I want you to tell me how you think that body shape would affect them in school"*

Young person (boy): ***"They might be focused too much on erm the bullies.....like their grades and everything, so they might slip down"***

(FG)

Example 2

Interviewer: *"Do you think your weight affects you at all, joining in at school?"*

Young person: ***"Sometimes, but not that much.....affects me in PE"***

(Girl, aged 11-14 yrs, Baseline BMI 30.0 (*BMI-SDS = 3.2*))

Finally some adolescents were very aware that their weight could have an impact on their future prospects. Others however saw their weight as something that they were doing something about now and thus would change in the immediate future.

Example 1

Young Person: ***"I agree with that because some people think, I think that some people won't accept you in jobs because of your size"***

(Boy, aged 11-14 yrs, Baseline BMI 30.6 ((*BMI-SDS = 2.9*))

Example 2

Young person: ***"When I think, when I grow up and want to be this want to be that, then I think well what's going to happen to my weight and everything, yeah"***

(Girl, aged 11-14 yrs, Baseline BMI 30.0 (*BMI-SDS = 3.2*))

Example 3

Young person: ***"No because you could like lose it and then it, you'd look back and you'd think oh I said it'd affect my future but it really hasn't."***
(Girl, aged 15-18 yrs, Baseline BMI 35.1 (*BMI-SDS = 3.2*))

Preliminary items, covering the range of the aspects of QoL affected by above normal weight status discussed by participants, were crafted from the quotations generated from the interviews. The wordings of items were closely guided by the phrasing used by adolescents. **Table 5.4** presents an example of how items were crafted from direct quotations.

Table 5:4 Example of crafting preliminary items from interview quotations

Adolescent's words (abbreviated)	Preliminary item wording
"Oh every time I bend down, I feel like that (pointing at belly) gets in the way..... when I'm reaching it'll come out and my back'll go and I'm like aargh" (Boy, aged 11-14 yrs, Baseline BMI 20.8 (*BMI-SDS = 1.5*))	It hurts when I bend down
"They might be focused too much on erm the bullies.....like their grades and everything, so they might slip down" *(FG)*	I can't concentrate on my school work because I get picked on

Once all of the coded interview quotes were assessed 50 preliminary items were crafted. These 50 items were then used in the consultation with individuals who have expertise and experience in the field of adolescent obesity. The result of the consultation process is given in the section below.

5.3.2 Stage 2: Consultation with specialists

Five members of staff from the WATCH IT community Leeds based weight management programme and three members of the Sea QoL group gave feedback on the preliminary item set. The comments from both groups complemented one another. Specialists agreed that adolescents would be likely to truthfully answer all of the proposed items. Some duplicated items were identified and it was suggested some of the items were re-worded. None of the specialists suggested that of any key aspect of QoL affected by weight status had been missed out, though the WATCH IT team leader recommended that the swapping of clothes between friends is quite an important thing, especially for young girls, and so advised adding an item to this effect to the item set. It was also suggested by the staff from WATCH IT that simplified alternative wordings should be used for the younger age group. For example, where emotions like 'frustrated' or 'annoyed' are used, these should be replaced with 'upset' or 'sad' for the

younger age group. Once all of the suggested amendments were incorporated, a revised 46 item set was produced.

5.3.3 Stage 3: Mapping items to dimensions of QoL, consultation with adolescents and thesis advisory group

As was discussed earlier, a preference based measure is limited in the number of items it can include and thus further refinement of the preliminary item set was necessary. As such, the collapsing of items covering a particular dimension of QoL was undertaken, the results are reported below.

5.3.3.1 Mapping

The mapping of the resultant forty-six items to dimensions of QoL generated six dimensions covering several sub-dimensions. **Appendix 5.7** displays the 46 item questionnaire mapped onto dimensions of QoL illustrated in the Fitzpatrick *et al*. (1998) review. For some items the mapping process was straightforward. For example, it was clear to see that items 1 to 7 were all concerned with movement, activities of daily living or physical activity and so they fit best into the 'Physical function' dimension. Two sub-dimensions, 'daily routine' and 'physical activity' were created for the seven items. Similarly, items 8 to 13 best fit into the 'Symptoms' dimension as they cover issues such as pain, shortness of breath and energy. Three sub-dimensions were created for these six items including 'Out of breath', 'Tired; Weak' and 'Pain; Hurt; Ache'. For the remaining items, the mapping process was not so clear cut. A number of the emotional items, such as items 19 and 20, could fit into two dimensions; the 'Psychological well-being' dimension as they are concerned with negative psychological constructs such as disliking bodily appearance or avoiding getting changed in front of others; or the 'Personal constructs' dimension as they are concerned with satisfaction with bodily appearance. As these two dimensions clearly overlapped, in this context, the two were combined into one dimension. It was difficult to find dimensions to combine Items 17 and 46 adequately. The items concerned future health and employment and none of the dimensions listed in the Fitzpatrick *et al*. (1998) review quite covered the concept of the 'future' or implications on the future. These items were mapped onto the 'Role activities' dimension as this was the most adequate choice out of all the other choices. The grouping of items to common QoL dimensions was used in the FG interviews with adolescents for producing finalised item wordings.

5.3.3.2 Consultation with adolescents and thesis advisory group

The results of the item wordings and response options generated during the two FG interviews, conducted with participants from weight management service C, were combined. There was good overlap in the findings that came from both interviews. The combined results from the FGs were presented to the thesis advisory panel for review and a finalised 29 item questionnaire with associated frequency and severity response options was developed (see **Tables 5.5 and 5.6)**.

Table 5:5 Finalised items after consultation with adolescents and thesis advisory group

Dimension	Sub-Dimension	Finalised item wordings N=29
		HOW DOES YOUR WEIGHT AFFECT YOU?
Symptoms	Pain	I have body pain / ache
	Tired	I get tired
		I get low energy
	Out of breath	I get out of breath
		HOW DOES YOUR WEIGHT AFFECT DAILY ROUTINES?
Physical function	Daily routine	I struggle to keep up when I am walking around with others
		I struggle when I am going up stairs
		I struggle to reach or bend down
		HOW DOES YOUR WEIGHT AFFECT PHYSICAL ACTIVITIES?
	Physical activity	I struggle to keep up with others when doing physical activity
		I struggle to keep up with others when I play sports
		I avoid doing things like running, cycling, swimming or playing sports
		HOW DOES YOUR WEIGHT AFFECT THE WAY YOU FEEL?
Psychological wellbeing / Personal constructs	Angry or Annoyed	I feel angry or annoyed because I am unable to do the same things as others
	Frustrated	I feel frustrated because I am unable to do the same things as others
	Uncomfortable or Embarrassed	I feel uncomfortable or embarrassed getting changed in front of others
		I feel uncomfortable or embarrassed shopping for clothes
		I feel uncomfortable or embarrassed meeting new people
		I feel uncomfortable or embarrassed eating in front of others
	Unhappy	I feel unhappy because I can't eat what I want
		I feel unhappy about the way I look
		I feel unhappy because I am unable to do the same things as others
	Disappointment	I feel disappointed because clothes aren't made in the size I need
	Self-control	I struggle to keep in control of what I eat
		HOW DOES YOUR WEIGHT AFFECT SCHOOL / COLLEGE?
Cognitive functioning	School / college work	I struggle to do as well as others at school
		I struggle to concentrate on school / college work
		HOW DOES YOUR WEIGHT AFFECT SOCIALISING?
Social wellbeing	Socialising	I get treated differently at school, such as being teased or picked-on or left out
		I get treated differently at home, such as being teased or picked-on or left out
		People treat me differently when I go out
		I avoid playing / hanging out or socialising with others
		HOW DOES YOUR WEIGHT AFFECT THE FUTURE?
Future prospects	Role activities	I worry about my health in the future
		I worry about the type of job/career I will be able to have

Table 5:6 Finalised response options

Response type	Levels				
Frequency	Never	Almost never	Sometimes	Often	Always
Severity	Not at all	A little	Quite a bit	A lot	Very much

5.4 Discussion

A weight specific LLDS has been created that is based on the views of adolescents aged 11 – 18 years living in the UK. A range of aspects of QoL that is affected by weight status was identified through consultation with adolescents. A single aspect of QoL an adolescent acknowledged as being affected by their weight status could affect more than one dimension of QoL. For example, an adolescent who finds it hard to keep up with their peers when doing physical activity could experience breathlessness and feelings of frustration. In this case, weight status impacted upon the adolescent's physical ability, which affected both the symptom and emotion dimensions of QoL.

The *bottom up* approach taken here obtained the views of the population of interest to inform the content of the LLDS. This approach was taken in constructing the new measure and is similar to that taken by Grewal *et al*. (2006) and Stevens (2009). Both studies directly identified dimensions of QoL through the narratives given by interview participants, in the development of a generic QoL measure for older and paediatric populations, respectively. The studies give further support to the qualitative approach in informing the content of the new measure in the current study. Furthermore, the involvement of adolescents in the development of the measure is in line with the Food and Drug Administration guidelines (FDA, 2006) on patient reported outcome measures, and endorsed the content and face validity of the measure. Carrying out interviews with the younger age group also ensured that the age-related vocabulary and language comprehension was appropriate for the age group under study (Stevens, 2009).

There are three limitations regarding the sample of participants interviewed in the study that need to be acknowledged. Firstly, the population of adolescents interviewed were based in Leeds, and it is not certain that the views of these individuals will reflect the views of adolescents throughout the UK. On the other hand Leeds is the third largest city in the UK, provides a diverse population in terms of ethnicity and socioeconomic status, and consists of both urban and rural areas. Leeds also provides a rich source of potential study participants, due to its long history of offering weight management programmes to the younger population. Some of the Leeds based services are recognised service and training providers by the Department of Health under the Framework Agreement for Child Weight Management (Cross-Government Obesity Unit, 2009). Primary Care Trusts (PCTs) throughout the UK utilise the services offered in Leeds, thus adolescents recruited from Leeds based organisations are likely to be representative of England and possibly the UK. Secondly, of the individuals interviewed

from the weight management services, the majority fell within the 11-14 year old age group. Related to this, the FG carried out in the school sample included adolescents in the younger age group, but none in the 15-18 year old age group. In addition to this, both the treatment seeking and school samples were identified by staff and teachers leading to the increased potential for selection bias. This might result in some key aspects of QoL affected by weight status being omitted. Thirdly, the majority of adolescents were recruited from the community based weight management services and this may differ from the views of adolescents who are very severely obese and require treatment in a hospital setting. These services offer guidance, support and education on both health and lifestyle behaviours. The views of adolescents undertaking other forms of weight management interventions have not been included in the study. The applicability of the new measure in the 'non-community based treatment seeking' population of adolescents is something that can be tested in future research.

Researcher bias and subjectivity are commonly understood as inevitable and important in qualitative research. Of particular relevance in this setting is the potential bias that could have arisen from the subjective nature of undertaking the analysis of interview data in generating the themes and subthemes that lead to the creation of the LLDS. The Framework approach was utilised in order to minimise the potential biases. Another potential course of action that may have minimised interpretation, bias on the part of the researcher, could have been through re-engaging with participants and asking them to review the interpretations of their transcripts.

In order to go some way in addressing the afore mentioned issues, in addition to the interviews with adolescents, consultation with individuals involved with the delivery of these services were carried out and added to the content of the LLDS. They also facilitated the identification of any issues that were omitted from the interviews. A comparison of the dimensions identified by the LLDS and the three existing weight specific QoL measures, specifically developed for the adolescent population and, for which measurement properties have been tested, is provided in **Table 5.7**. The details of these three measures were provided in **Chapter 3** and included: IWQOL – Kids (Kolotkin *et al*., 2006), Sizing Me Up (Zeller and Modi, 2009), and YQOL-W (Morales *et al*., 2011).

Table 5:7 Comparisons of the dimensions in the new measure with existing measures

IWQOL-Kids (27 items)	Sizing Me Up (22 items)	YQOL-W (21 items)	NEW LLDS (29 items)
Physical comfort	Physical functioning	Environment (challenge of physical activity)	Symptoms
			Physical function
Body esteem	Emotional functioning	Environment (finding clothes that fit, feeling comfortable in school/public)	Psychological wellbeing / personal constructs
	Social avoidance	Self (psychosocial impact, body dissatisfaction, coping strategies)	Role activities (future prospects)
Social life	Teasing/marginalisation	Social (fitting in, avoiding participation, feeling attractive, social acceptance)	Social wellbeing
Family relationships			
	Positive social attributes		
			Cognitive function

The content of the new measure and the existing measures show some similarities, as they all contain physical, emotional, and social dimensions. In the new instrument, current concern over the impact of weight status on future health is also included. This is also reflected in the YQOL-W, which includes the item 'I worry that my weight will prevent me from getting a good job'. In terms of the impact of weight status on family relationships, the new measure takes a more general standpoint – for example the IWQOL-Kids has six items about negative treatment by family members. The new measure includes one item that should cover this range of issues 'I get treated differently at home, such as being teased or picked-on or left out'. There are key differences in the content of the new measure with respect to the existing measures. Although physical functioning/comfort is addressed in existing measures, the *consequences* of weight status on physical activity, in terms of symptoms are not clearly defined. For example in the YQOL-W, one of the items relating to this dimension asks 'Because of my weight exercising is hard for me'. However, it is not possible to identify *what* exercising impacts upon i.e. breathing or low energy levels etc. The new measure allows respondents to identify limitations of performing different activities (in the physical function dimension) and the consequences resulting from performing these activities (in the symptoms dimension). Another key difference is that the impact of weight status on future work or health is only partially identified in one of the existing measures. The Sizing Me Up instrument contains a 'Positive social attributes' dimension. The new measure does not contain any items that fit this dimension. Perhaps this is because the interviews were designed to probe the limitations associated with weight status. It must however be noted that during each of the interviews

with adolescents that was carried out in the current study, positive feelings about weight were investigated, but none of the interviewed adolescents identified with this notion. Finally, one dimension that was identified in the new measure, but not in any of the existing instruments was the impact of weight status on cognitive function. The existing instruments discuss the social impact of weight status in school – but none of the items thus far have specifically addressed issues regarding academic work.

Another comparison of the new measure to existing ones can be made with regard to the wording of items. All three existing measures contain a reference to weight status, using terms like 'because of my weight' or 'because of my size' within each of the items. As such this issue was considered in the formulation of wording of items in the new measure. There are two options available for referencing items to weight status: a) either all items can have a reference to weight status by adding the statement 'because of my weight' before or after each item or b) none of the items contain a reference to weight status. It is not possible to have 'partial referencing', where some items include the reference to weight status, but others do not. Partial referencing may lead to respondents considering their weight status only for items where the reference is included; however, where the reference to weight status is not included, respondents may infer that their weight should not be taken into account. A benefit of adding a reference to weight status is that it helps to focus the respondent in terms of thinking about how weight status, as opposed to other issues, such as having a twisted ankle for example, affects different aspects of life. This is of benefit as there is disaggregation of the dysfunction caused by: a) the ailments not linked to weight status and b) the negative impact of above normal weight status. Conflating these separate issues may lead to an overestimation of the negative impact of weight status on QoL. The inclusion of a reference to weight status to items also brings up the issue of attribution. For example, for an item concerning *getting out of breath*, a respondent that has a co-morbidity, such as asthma, in addition to being above normal weight status will face difficulties in disaggregating whether their breathlessness is due to their asthma or whether it is due to their weight status. Evidence shows that people who are overweight have a higher risk of developing asthma. Indeed, some of the treatment seekers participating in the one-to-one interviews reported having asthma. If a reference to weight status is included, then this would mean respondents could face such difficulties. As a result of this, the decision was taken not to add the reference to weight status in the wording of items.

In order for a potential respondent to the new questionnaire to indicate the degree to which they have or don't have an issue with a particular aspect of their QoL, it was necessary to identify appropriate wording and ordering of response categories (or levels, terminology used interchangeably) for each item. The most recent existing measure, YQOL-W, employs an 11 point numerical scale from 0 (not at all) to 10 (very much). It is not common practice for numerical response scales to be used in existing preference based measures primarily because of the need for elicitation of preference values. Potential respondents are likely to sum the numbers across different items and come up with a total score rather than actually thinking about the different levels. Another limitation associated with the use of an 11 point numerical scale covering five to six dimensions of QoL is that it will be too much information for respondents involved in a preference elicitation task to process. Modelling limitations also preclude having a large number of response options as: a) a large sub-set of health states would need to be sampled for direct valuation b) a very large sample size would be necessary to cover all the plausible health states. These drawbacks led to the decision of not using a numerical scale for the new measure, instead the use severity or frequency response options were chosen. A decision was made to create two sets of response scales for severity and frequency, utilising both existing literature and the words used by adolescents in the interviews carried out in Stage 1. Both types of response options were considered as neither was identified as superior from the literature (or by adolescents).

A decision regarding the 'recall period' or 'time frame' that should be used in the new measure was needed. In terms of the existing measures and the other generic QoL assessments, it is commonplace to specify a period of time e.g. today, in the past seven days or in the past month. The IWQOL – Kids measure uses 'the past seven days' whilst the Sizing-me-up instrument uses 'the past month'. For the current measure, however, for some of the items, the question would not make sense if a particular activity has not been carried out say in the past week. For example 'I never feel uncomfortable or embarrassed shopping for clothes' over the past seven days - because I have not gone shopping. Furthermore, guidance provided by Food and Drug Administration (FDA, 2009) on patient reported outcome measures endorses the use of a maximum recall period of two weeks, which precludes the use of the 'over the past month' reference. The time frame of 'At the moment' will be used in the current setting as this overcomes the issue of respondents not necessarily carrying out a particular activity within a specific time period. The YQOL-W measure uses *right now* which may be acceptable for the US adolescent population, but in the UK the term may be misinterpreted.

5.5 Conclusion

The views of adolescents have driven the content of a new twenty-nine item weight specific QoL measure consisting of six dimensions and fourteen sub-dimensions. For the measure to be appropriate for the calculation of QALYs, preference values need to be elicited for health states as described by the new measure. The next step will be to reduce the number of items so that the new measure is appropriate for the elicitation of preference values.

Chapter 6
Refinement of the weight specific LLDS

6.1 Introduction

This chapter reports on the second empirical study (Study 2) undertaken in the thesis, the Part 2 of the three part process which aimed to identify a reduced item set, from the 29 item weight specific LLDS, developed in **Chapter 5**, and generate a reduced form weight specific descriptive system appropriate for preference elicitation. Psychometric assessments in conjunction with Rasch analysis were used to refine the LLDS. A detailed account of the three step process (adapted from the existing literature) undertaken is provided in the next section preceded by a general overview of Rasch analysis.

6.1.1 An overview of Rasch analysis

Rasch analysis (Rasch, 1960) has been widely used in numerous settings for the development and validation of outcome measures. Rasch analysis has recently been used in studies the health care setting for the development of visual functioning questionnaire (Gothwal and Bagga, 2012), in the validation of the World Health Organisation questionnaire (Krägeloh *et al*., 2012) and in the validation of the Hospital Anxiety and Depression Scale in spinal cord injury (Müller *et al*., 2012). It is a mathematical modelling technique that converts qualitative (categorical) responses to points on a continuous (unmeasured) latent scale using a logit model (Young *et al*., 2011). Data collected from ordinal instruments or scales that are intended to be summated into an overall score, are tested against the expectations of the Rasch measurement model. In the case of a categorical ordered scale with 5 response options for any given item, the Rasch model scores items 0, 1, 2, 3, 4. This indicates increasing levels of a response; the responses are then added across all the items in the instrument to give a total score that summarises the responses to all the items, and indicates a QoL score for any given individual who completes the instrument. Summing the scores of the items to give a single score for a person implies that the items are intended to measure a single construct; in other words, that the scale is uni-dimensional (for a more in-depth description of the Rasch measurement model, see the monograph by Hobart & Cano (2009)).

Rasch analysis fit statistics

Rasch analysis provides a method by which ordinal data, such as that generated by HRQoL instruments, can be converted to continuous data. The mathematical model is given primacy in the Rasch paradigm. In other words, the proponents of Rasch measurement prioritise the model (as opposed to trying to find a different model to fit the data). Therefore the case rests on the inherent properties of the model and essentially, it provides the optimum criterion for fundamental measurement. Any unidimensional measure captures an underlying trait (in this case, weight specific QoL or it could be a particular dimension of HRQoL), which is represented by a latent scale. There are two key components to the theory of Rasch measurement: a) Individual respondents are located along the latent scale according to their levels on the latent trait and b) Item response levels will be located along the same latent scale according to the level of QoL that they represent. Rasch analysis assumes that the probability of a respondent endorsing a particular item response is a logistic function of the relative distance between that individual's position on the latent scale, and the position of the item response on the latent scale (Hobart & Cano, 2009). Based on the aforementioned components, the Rasch model defines the ideal item response characteristics if measurement (at the interval level) is to be achieved. Observed data, collected from respondents completing the measure under investigation, is tested against this model, and a series of statistics are computed in order to evaluate whether the observed data and the modelled data are similar. The observed response patterns achieved are tested against expected patterns. The Rasch model shows what should be expected in responses to items if measurement at the metric level is to be achieved. If the invariance of responses across different groups of people does not hold, then taking the total score to characterise a person is not justified. The objective of Rasch analysis is to test how well the observed data fit the expectations of the measurement model, and in so doing a range of fit statistics are considered.

It is important to clarify that there is a difference in the application of Rasch models, which are dependent on the purpose for which the analysis is being undertaken. Rasch analysis can be used for a) the development or validation of HRQoL instruments or b) the development of a descriptive system, appropriate for the elicitation of preference values from respondents, with a minimum loss of information. In the case of the former it is assumed that the construct under investigation is uni-dimensional, that is, that the instrument under study, is only measuring one aspect of HRQL, for example physical functioning. However, the assumption of uni-dimensionality is not valid for the health-state classifications used by preference-based measures, like EQ-5D (EuroQol Group, 1990). These are based on assumptions related to multi-attribute utility theory and rely on the dimensions being orthogonal (independent of each other). Multi-attribute utility theory uses utility weighting to combine the independent dimensions into a single index (Young *et al*., 2009).

6.2 Methods

Rasch analysis was utilised in the current study in order to identify (exclude) the best (worst) performing items through the assessment of the discriminative ability of each item across different weight categories, as well as the assessment of the ordering of response categories. The analysis was conducted using the RUMM2030 (Andrich, *et al*., 2010) software package. Psychometric assessments were carried out in SPSS Version 18 (SPSS Inc, 2009).

6.2.1 The survey

6.2.1.1 Participants

In order to apply psychometric and Rasch analysis, the 29 item instrument was completed by adolescents. Eligible respondents included a) adolescents currently enrolled in weight management interventions, b) adolescents that may qualify for weight management intervention (owing to their current weight status) but not utilising the service and c) adolescents who are not obese or at risk of being obese (overweight) and are not eligible for enrolment into a weight management intervention. The participant sample (N=341) included treatment seekers (N=25) and non treatment seekers (N=316). Oversampling of adolescents who perceive themselves to be of above normal weight status was carried out, in order to better represent the population for which the measure targets. Perceived weight status was used as an indicator for BMI as literature shows that accurate weight perception in both overweight and obese adolescents can be observed (Khambalia *et al*., 2012).

Participants were recruited from two different sources: adolescents enrolled into weight management programmes (one being the same programme where adolescents were recruited from in Study 1 – see **Chapter 5**) were asked to complete the survey; adolescents not enrolled in a weight management intervention were recruited via an internet panel supplied by a market research company. The former group of participants were given the opportunity to be entered into a prize draw for completing the instrument. There was one first prize winner and seven runners up winning £30 or £10 respectively of retail vouchers. Parents of participants who were recruited via the survey company were paid by the survey company (a total sum of £2.75 or £4.85 depending on whether participants were recruited in first or second wave of data collection). It should also be noted that there was no way of knowing if adolescents recruited from the survey company were enrolled in a weight management intervention. Though not knowing this information is not necessary for the data analysis. Consent to take part in the study was required from all adolescents who completed the survey. Additional consent and weight and height data was collected from parents of participants aged 11-15 years recruited from the internet panel and the latter was self reported for participants 16 years old and over. Weight and height data for treatment seekers was provided by the management programmes. The survey was available in both electronic and paper formats. **Appendix 6.1** provides the paper based survey and was similar to the electronic survey (the prize draw in the last page was not included in the electronic survey). Ethical approval was provided by the University of Leeds local research ethics committee (Ref - HSLTLM/11/006).

6.2.2 Measures included in the survey

6.2.2.1 The weight specific QoL instrument

The survey contained the 29 item weight specific instrument. The two alternative frequency and severity response options were assessed in order to identify the best performing as per the findings of the psychometric and Rasch analysis. For the respondents recruited from the internet panel, the ordering of the two versions of the instrument was randomised so that half the respondents completed the frequency response options first and the other half completed the severity response option first. For the respondents recruited from the weight management service the survey was administered electronically (via an on-line survey hosted by Bristol Online Survey (BOS)) or as a paper based version when the survey was piloted.

6.2.2.2 Other measures

General background characteristics were assessed in the survey by including questions on: age, gender, and geographical location. Questions assessing self assessed health and life satisfaction were also included as these parameters provide an indication of the general health status and psychological wellbeing of study participants. BMI was calculated using either self-reported weight and height or those provided by the weight management programme. Additional background questions were obtained from the respondents recruited via the internet panel in order to assess the representativeness of the sample. These were taken from the Longitudinal Survey of Young People in England (LSYPE) (DCSF and National Centre for Social Research, 2009) a nationally representative sample of young people, consisting of around 15,500 13 to 14 year olds in the first wave of data collection, which took place in 2004. This comparison was used to identify similarities and differences (if present, in what respect) between the two datasets. The additional questions assessed the following: parental employment status, home ownership, use of sports facilities outside of school lessons and the hours of computer use for school work.

6.2.3 Sample size

No formal sample size estimation methods for undertaking Rasch analysis were found. Instead a rule of thumb sample size recommendation was followed. The sample size requirements for Rasch analysis are based upon the degree of precision required for estimating item difficulty and person ability. Sample size calculations for Rasch analysis aim to address the same issues as other types of statistical analysis. The optimal sample size aims to: a) predict more precise estimates (smaller standard errors) b) provide more powerful fit analysis and c) provide more robust estimates (less likely that type I or II errors will distort statistical findings). Generally, a sample size of 243 respondents[9] should adequately address the three points above. It has also been recommended that a sample size of around five to ten subjects per item should reduce the effect of chance (Linacre, 1994 and Linacre, 2002). Following this recommendation, for the longest potential summated scale, assessing psychological wellbeing / personal constructs which contains 11 items, a minimum sample size of 110 would be required.

[9] Thesis advisory group members with specialist knowledge of Rasch analysis provided guidance in terms of i) the identification of a suitable sample size and ii) the Rasch analysis methodology that was undertaken.

6.2.4 Pilot study

In preparation for the main study, a pilot study was carried out to examine the 29 item instrument and the other questions included in the survey for clarity and ease of comprehension. Participants were given the paper version of the survey to complete by themselves and were told to ask for help if they needed it. They were also asked to identify any questions that were either unclear or should be re-worded to aid understanding. Particularly, feedback was sought in terms of the identification of any ambiguous or confusing items in the 29 item instrument. On completion of the survey, adolescents were asked if they thought anything important was missing from the 29 item instrument that they perceived to be affected by their weight.

The pilot study consisted of 15 adolescents recruited from the same weight management service where the majority of the one-to-one interview participants were recruited from (service A, details given in **Chapter 5**). All but one of the adolescents self-completed the survey and did not report any problems. The participant who didn't was diagnosed with mental health problems and the instructions and questions needed to be read to them by their carer. However this participant completed the survey once this was done and did not seem to have problems understanding the questions. When all the participants were asked on completion of the survey, if any specific aspect of QoL was excluded or missing from the 29 item instrument, none of the participants reported any. Owing to this no changes to the survey were necessary. The data collected from the pilot study were added to the main study data.

6.2.5 Analysis using classical psychometric and Rasch analysis

Psychometric assessments in conjunction with Rasch analysis have been used in the development of preference based measures from pre-existing disease specific QoL instruments (Young *et al.*, 2009, 2010, and 2011). Previous studies have adopted a five step process for deriving a reduced health state classification from existing non-preference based HRQoL instruments; this informs the analysis used in the current study. In Step I factor analysis was used to establish instrument dimensions. Step II examined the dimensions identified from Step I and excluded items that did not meet the initial validation process, i.e., using the Rasch rating scale model to validate uni-dimensionality of the dimensions. In Step III the combined criteria based on Rasch analysis, classical psychometric testing, and data generated from the qualitative interviews carried out in the preceding study were used to select the final set of

items for the health state classification system. **Figure 6.1** provides a summary flow diagram of this process and a more detailed description of what each step entails is provided below. The fit statistics assessed in the psychometric and Rasch analyses have been summarised in **Appendix 6.2** and **6.3**.

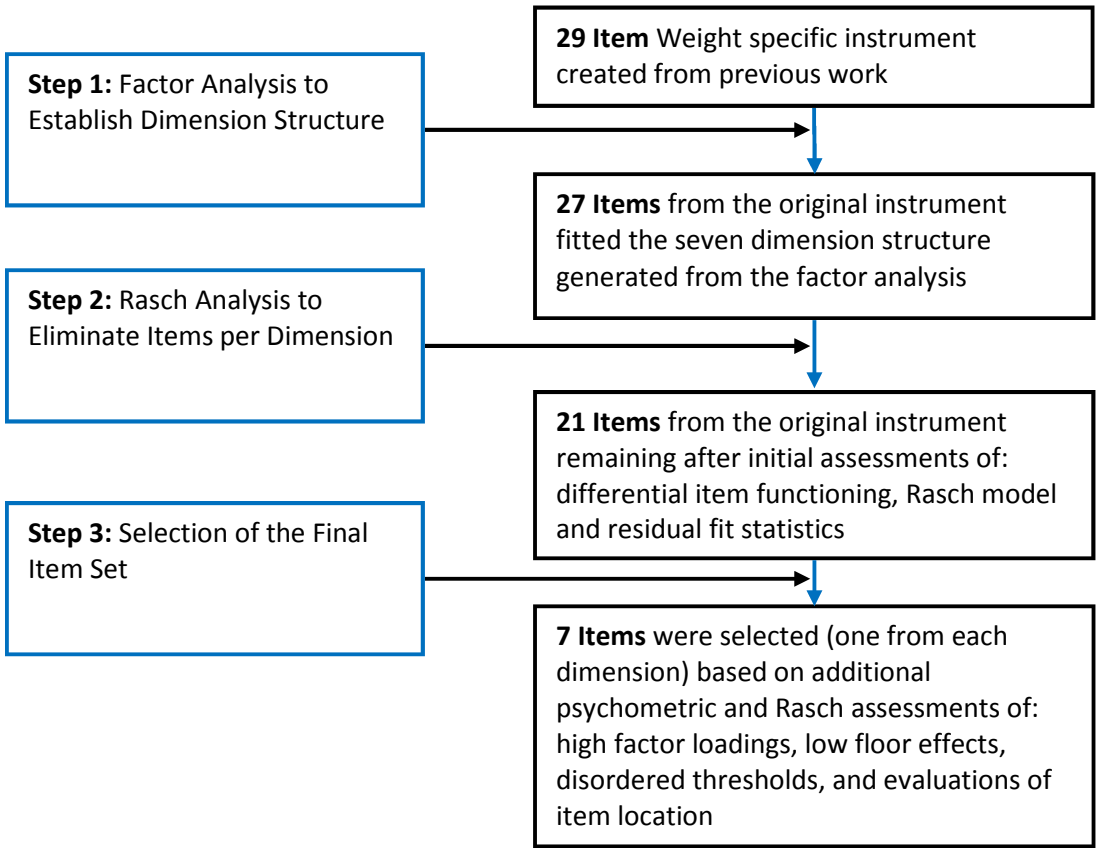
6.2.5.1 Step I: Factor analysis to establish dimension structure

In the first instance the optimal grouping of items measuring the same underlying construct or dimension of QoL was identified by the application of factor analysis. This was used to identify the underlying factors that explain patterns of correlation within a set of observed variables (Young *et al.,* 2009), thus facilitating the grouping of the 29 items into acceptable and justifiable categories.

Factor analysis methodology

Exploratory factor analysis (EFA) was utilised in the first instance to identify item groupings (Tabachnick and Fidell, 2007). Other methodological issues concerning the: a) extraction method, b) determination of the number of factors to include in the analysis and c) type of rotation to use, were also considered.

Figure 6:1 Flow diagram illustrating the Rasch and classical psychometric analysis process*



*adapted from Young *et al.*, 2011 (*p.* 199)

In terms of the factor extraction method, both the maximum likelihood estimation and the *Principal Axis Factors* method (PAF, as per SPSS) were implemented. The former assumes that the data are relatively normally distributed and in the latter, that the assumption of multivariate normality is *severely violated* (Fabrigar *et al*., 1999). With regard to the number of components that should be included in the analysis both *over*-extraction and *under*-extraction of factors retained for rotation can have detrimental effects on the results (Costello & Osborne, 2005). The default in most statistical software packages is to retain all factors with Eigenvalues greater than 1.0. There is broad consensus in the literature that this is among the least accurate methods for selecting the number of factors to retain (Velicer & Jackson, 1990). Thus, the selection of the number of components was informed by undertaking Parallel analysis. The Parallel analysis compared the Eigenvalues obtained using a Monte Carlo simulation of random numbers and the Eigenvalues generated from the *observed* dataset. The choice regarding the rotation method that should be used mainly depends on whether the factors in the analysis are correlated (oblique: direct oblimin & promax) or uncorrelated (orthogonal: varimax, quartimax, & equimax). The factor correlation matrix was assessed for correlations around 0.32 and above to inform the rotation method that was used and the lower threshold for factor loadings was set at 0.40 (Tabachnick and Fiddell 2007).

The factor structure of the scale estimated from the EFA was tested using confirmatory factor analysis (CFA). CFA allows an *a priori* specification from the EFA to be assessed by looking at the fit of the observed data to this factor structure. A close fit between the data and the hypothesised latent variables serves to confirm the factor structure. Indices of fit that were assessed included: 1) the Tucker Lewis Index (TLI), 2) the Comparative Fit Index (CFI), and 3) the Root Mean Square Error of Approximation (RMSEA) (Sim *et al*., 2011). For the TLI and the CFI, values approaching unity are desirable, with values greater than approximately 0.95 considered to indicate a good fit, whereas for the RMSEA values should approach zero, preferably below approximately 0.05 (Sim *et al*., 2011).

6.2.5.2 Step II: Using Rasch analysis to exclude Items

In Step 2, Rasch models were applied to each of the dimensions identified in Step 1 in order to exclude poorly performing items. The following assessments were made: a) Rasch model goodness of fit, b) differential-item functioning and c) item-level ordering. Explanations of each of these fit statistics are provided below. The process of fitting Rasch models was repeated until only well-fitting items remained and the overall item-trait goodness of fit of the model (see below) was non-significant. Any item that was removed from the Rasch models was excluded from further consideration in the final health-state classification.

Rasch model goodness of fit

An assessment of overall model fit was undertaken by examining the following: item–trait interactions, the person separation index, and person and item fit residuals.

The item–trait interaction **-** measures whether data fit the Rasch model for discrete groups of responders (based upon subgroups where each responder lies on the latent scale of the Rasch model i.e. individuals who tend to have similar QoL scores for a specific dimension will be grouped together). Observed and expected responses are compared across items and traits and the difference between these responses is summarised using the Chi-squared (X^2) test statistic, well-fitting models should have no deviation between the observed and expected responses and, thus the P-value for the overall model X^2 statistic should be >0.01 (Young *et al.,* 2009).

The person separation index (PSI) - measures the level of discrimination amongst different groups of respondents and the higher the PSI value, the better the level of discrimination, a value of 0.7 or more indicates a well-fitting Rasch model (Young *et al.,* 2009).

Fit residuals - provide estimates of the amount of divergence between the expected and observed responses for each respondent or item response; fit residuals are summed over all items (item fit residuals) or summed over all persons (person fit residuals). The mean item or person fit residual (Z-score) should be approximately zero with a standard deviation approximately equal to one (Young *et al.,* 2009). Items that did not meet the goodness of fit tests were not selected in the final item set.

DIF by age and gender

Differential-Item Functioning (DIF) was examined to establish whether responses to the HRQoL instrument systematically differ across patient characteristics (e.g. for an item asking about physical abilities, boys might select less severe item response options than girls). Items where it is necessary to adjust for systematic DIF across groups of responders are of limited value for making cross-population comparisons and, therefore, were excluded from further consideration. Gender (male/female) and age (younger adolescents (11-15 years)/older adolescents (16-18 years)) were examined for DIF using item-characteristic curves and item-by-characteristic ANOVA statistics (Young *et al.,* 2009). Evidence of DIF resulted in the exclusion of items from the final item set.

Item level ordering

In the construction of a health-state classification system that is appropriate for valuation, it is important to ensure that information relating to items and item levels are not ambiguous or unclear. Previous work with health state classifications has found that respondents sometimes have problems distinguishing between item levels (Young *et al.,* 2009). The identification of potentially problematic level orderings (i.e. identification of items where responders were unable to distinguish between item-response levels) is necessary in order to ensure the response ordering of health states being valued is robust. Item-threshold probability curves (a plot of the probability of being in each item level across the latent QoL scale) were examined to assess item-level ordering. For an ordered item, the thresholds between item levels are the points at which each item level is equally likely to occur. Disordered item levels highlight the inability of respondents to distinguish between item levels. Items with dis-ordered thresholds were not selected in the final item set *where possible* (see results section).

6.2.5.3 Step III: Selection of the final item set

In Step III further examination of item level Rasch statistics alongside other conventional psychometric analysis on the items that were not excluded from consideration was undertaken to assess the *best* items for the final version of the new weight specific instrument. The findings from the content of the qualitative interviews reported in **Chapter 5** were re-visited to inform the final selection of items. A key objective of the selection process for the reduced health state classification was to select items that span the full range of condition severity.

The performance of items across conventional psychometric criteria was taken into consideration when selecting items. The selection of the final set of items, which were not excluded in Steps I and II, was supported with the assessment of: feasibility, internal consistency, floor and ceiling effects, and the ability to distinguish between different weight categories. Item selection was predominantly based upon the spread of item levels across the latent space, as selected items should span the full range of condition severity, where the wider the spread the better the item. The threshold probability curves and item goodness of fit statistics used in Step II were re-examined. Item maps were assessed for each of the 7 dimensions to assess the spread of items across the latent scale. In addition to giving preference to the best overall performance of items across Rasch and psychometric tests, the interview transcripts from the qualitative study reported in **Chapter 5** were consulted again. Problems raised by the majority of the adolescents that were interviewed gave precedence in the final selection of items. Item wordings were also reviewed to retain the original expressions used by adolescents.

Rasch analysis was undertaken on the final set of items that were selected to make up the descriptive system. This allowed a preliminary assessment of the performance of the descriptive system and the identification of any potential problems that may arise. It should be noted that this analysis was mainly to assess the un-dimensionality of the final item set and the Rasch model goodness of fit statistics.

6.3 Results

6.3.1 Participants

A total of 341 adolescents completed the survey consisting of 25 adolescents recruited from three different community based weight management services and 316 from the internet panel. The characteristics of the sample are summarised in **Table 6.1**. The majority of the sample was in full time education. The ethnic background of around 80% of the sample was White and approximately 90 % lived in England. Based on the age and gender adjusted BMI calculations, there were 26 (8%), 152 (45%), 51 (15%) and 112 (33%) underweight, normal weight, overweight and obese adolescents in the survey (BMI calculations were based on the population thresholds discussed earlier). Just over half of the sample reported having health problems, the most common relating to having problems with eyesight or anxiety / depression. In terms of overall self-assessed health status, 212 (62%) participants report being in excellent or good health and 39 (11%) participants perceive their health to be poor or very poor. A summary table comparing the data from the 316 survey participants recruited from the internet panel with the LSYPE Wave 1 data is provided in **Table 6.2**. There are similarities across the two samples, however there some are some differences regarding type of home ownership and parental/carer employment.

6.3.2 Analysis of the 29 item instrument

Data from the 26 adolescents that fell into the underweight BMI category after adjusting for age and gender were excluded from the analysis (as the purpose of the measure is to assess the impact of above normal weight status on QoL), leaving a final sample size of 315.

6.3.2.1 Step I: Factor analysis to establish dimension structure

A summary of the findings of the psychometric analysis for the frequency and severity scales is given in **Table 6.3**. The results showed that the frequency and severity scales were close in their performance across all of the items, however, the frequency scale consistently had a lower percentage of responses answering *never* compared to *not at all* across all the 29 items. The results of the t-test to assess whether the items discriminate between weight categories showed that items were better at discriminating between normal weight and overweight (25 items when the frequency scale and 28 items when the severity scales were used) than between overweight and obese (5 items when either the frequency or severity scales were used). Three items displayed item total correlations above 0.8 and none displayed correlations lower than 0.32 for both scales.

Table 6:1 Participant characteristics N=341

Characteristics		N	%
Gender	Male	172	50.4
	Female	169	49.6
Age	10[a]	1	0.3
	11	34	10.0
	12	39	11.4
	13	28	8.2
	14	25	7.3
	15	35	10.3
	16	39	11.4
	17	62	18.2
	18	78	23.0
BMI	Underweight [b]	26	7.6
	Normal weight [b]	152	44.6
	Overweight [b]	51	15.0
	Obese [b]	112	32.8
In full-time education	Yes	323	94.7
	No	18	5.3
Ethnicity	White	274	80.4
	Mixed/dual heritage	11	3.2
	Asian or Asian British	28	8.2
	Black or Black British	17	5.0
	Chinese	5	1.5
	Other	3	0.9
	Preferred not to say	3	0.9
Geographical location	England	303	88.9
	Scotland	14	4.1
	Wales	16	4.7
	Northern Ireland	8	2.3
Self assessed health status	Excellent	84	24.6
	Good	128	37.5
	Fair	90	26.4
	Poor	35	10.3
	Very poor	4	1.2

[a] This individual was only a few weeks away from their 11th birthday when the survey was administered in the pilot survey, and thus it was decided to include them in the study

[b] Underweight <= 2nd centile, Normal weight >= 2nd centile or < than 85th centile, Overweight > = greater than 85th centile or < 95th centile and Obese >= 95th centile (Refs: NOO 2009/10, NOO 2011a and Cole *et al*., 2007)

Table 6:2 Comparisons between LSYPE and the study internet panel datasets*

Employment	LSYPE (N)	LSYPE (%)	Survey (N)	Survey (%)
Full-time paid employee (30 or more hours a week)	5430	35.2	175	58.5
Part-time paid employee (under 30 hours a week)	4239	27.5	37	12.4
Full-time self-employed	554	3.6	19	6.4
Part-time self-employed	242	1.6	8	2.7
Unemployed and seeking work	399	2.6	10	3.3
Full-time education	191	1.2	7	2.3
Temporarily sick/disabled	118	0.8	3	1.0
Permanently sick/disabled	412	2.7	6	2.0
Looking after home/family	3652	23.7	17	5.7
Retired from work altogether	162	1.1	6	2.0
I don't know/ I'd prefer not to say/refused	8	0.7	11	3.7
other answers	101	0.1		
missing/not applicable	262	1.7	17	5.4
Total	15770		316	
Valid total	15407	97.7	299	94.6
Accommodation				
Owned outright	1919	12.3	70	22.2
Being bought on a mortgage/bank loan	8629	55.2	117	37.0
Shared ownership (owns & rents property)	63	0.4	2	0.6
Rented from a Council or New Town	2489	15.9	44	13.9
Rented from a Housing Association	1392	8.9	24	7.6
Rented privately	911	5.8	35	11.1
Rent free	69	0.4	3	0.9
Some other arrangement	110	0.7	10	3.2
I don't know/ I'd prefer not to say/refused	63	0.4	11	3.5
Other answers			NA	NA
missing/not applicable	125	0.8	0	0.0
Total	15770		316	
Valid total	15645	99.2	316	100
Physical activity				
Most days / 5 times a week or more	5259	34.1	81	25.6
Once or twice a week	7052	45.7	106	33.5
Less than once a week / hardly ever	1992	12.9	65	20.6
Never	1111	7.2	58	18.4
I don't know/ I'd prefer not to say/refused	17	0.1	6	1.9
missing/not applicable	339	2.1	0	0.0
Total	15770		316	
Valid total	15431	97.9	316	100.0
Use computer for school work				
Yes	11888	91.5	292	94.8
No	1098	8.5	7	2.27
Other	0	0.0	9	2.92
I don't know/ I'd prefer not to say/refused	7	0.1	0	0.0
missing/not applicable	2777	21.4	8	2.53
Total	15770		316	
Valid total	12993	82.4	308	97.5
Average number of hours a day using a computer at home for school / college related work				
0	129	1.1	0	0.0
1	7508	63.2	90	28.5
2	3291	27.7	93	29.4
3	604	5.1	34	10.8

Employment	LSYPE (N)	LSYPE (%)	Survey (N)	Survey (%)
4	114	1.0	25	7.9
5	49	0.4	11	3.5
6	12	0.1	13	4.1
7	3	0.0	2	0.6
8	3	0.0	6	1.9
9+	6	0.1	18	6.2
I don't know/ I'd prefer not to say/refused	161	1.4	0	0.0
missing/not applicable	3890	24.7	24	7.6
Total	15770		316	
Valid total	11880	75.3	292	92.4

* LSYPE dataset n=15,000 and internet panel dataset n=315. The% is calculation is based on total where data is available (valid total)

Table 6.3 also shows the results of the factor analysis. This analysis showed that the 29 items could be grouped into seven factors. Factors were allowed to be correlated with the use of a direct Oblimin rotation (see **Appendix 6.4**). Again similarities between the factor loadings were observed in the PAF analysis when using either the frequency or severity scales. Only item 27 displayed disagreement regarding the factor loading between the two scales. The severity scale displayed more items with factor loadings lower than the 0.4 threshold (four items) than the frequency scale (two items). Confirmatory factor analysis supported the seven factor structure displaying better fit statistics overall than for a 6, 5, 4 or 3 factor solution (see **Appendix 6.5**).

Given that the frequency response scale systematically displayed a smaller proportion of responses in the *Never* category across all items and a higher number of items loading onto factors, the decision was made that this response scale should be used for the new instrument over and above the severity scale.

Table 6:3 Summary of psychometric analysis results by response category N=315 (Step I)

Item No.	FREQUENCY SCALE						SEVERITY SCALE					
	Check for floor effect (% *Never* responses)[a]	p-value (from t-test for discrimination between Norm_WT & Over_WT)	p-value (from t-test for discrimination between Over_WT & Obese)	High ITC[b]	Factor	Principal axis factors estimation (PAF) Factor loading[c]	Check for floor effect(% 'Not at all' responses)[a]	p-value (from t-test for discrimination between Norm_WT & Over_WT)	p-value (from t-test for discrimination between Over_WT & Obese)	High ITC [b]	Factor	Principal axis factors estimation (PAF) Factor loading[c]
17	41.3	0.009	NS	0.73	F1	**0.387**	44.4	0.006	NS	0.74	F1	**0.330**
21	41.9	0.01	NS	0.73	F1	0.586	46.3	0.001	NS	0.76	F1	0.635
22	58.1	0.001	NS	0.79	F1	0.624	61.6	0.001	NS	0.79	F1	0.670
23	54.3	0.001	NS	0.77	F1	**0.390**	59	0.002	NS	0.77	F1	0.458
24	60	0.019	NS	0.74	F1	0.594	62.9	0.016	NS	0.71	F1	0.724
25	64.1	0.008	NS	0.76	F1	0.640	67.6	0.007	NS	0.72	F1	0.719
26	41	0.013	NS	0.77	F1	0.766	42.9	0.002	NS	0.74	F1	0.769
28	64.8	0.009	0.054	0.75	F1	0.564	66.3	0.006	NS	0.75	F1	0.533
29	50.8	0.28	0.024	0.77	F1	0.692	54.9	0.004	NS	0.74	F1	0.670
8	47.9	0.007	NS	0.75	F2	0.657	59	0.001	NS	0.78	F2	-0.873
9	47.9	0.013	NS	0.75	F2	0.733	54.9	0.013	NS	0.79	F2	-0.850
10	56.5	0.05	NS	0.71	F2	0.577	63.5	0.013	NS	0.73	F2	-0.600
1	47.9	NS	NS	0.63	F3	0.617	60.6	NS	NS	0.66	F3	0.424
2	**37.5**	NS	NS	0.72	F3	0.814	45.7	0.011	NS	0.72	F3	0.747
3	**29.8**	NS	NS	0.70	F3	0.896	**37.1**	0.018	NS	0.72	F3	0.794
4	**37.5**	NS	0.028	0.69	F3	0.415	47.3	0.008	0.063	0.73	F3	**0.318**
13	69.2	0.018	NS	0.66	F4	0.599	71.7	0.007	NS	0.67	F4	0.640
14	78.4	NS	NS	0.63	F4	0.436	81	0.068	NS	0.66	F4	0.538
15	72.7	0.002	NS	0.68	F4	0.681	77.1	0.003	0.056	0.72	F4	0.817
16	66.3	0.002	NS	0.73	F4	0.665	69.8	0	NS	0.73	F4	0.581
11	60.3	0.028	NS	0.67	F5	-0.790	68.9	0.033	NS	0.65	F5	-1.000
12	52.7	0.025	NS	0.67	F5	-0.930	59.4	0.005	NS	0.65	F5	-0.777
18	50.5	0.07	NS	0.63	F5	-0.419	51.7	0.006	NS	0.65	F5	**-0.270**
19	61.3	0.001	NS	0.84	F6	-0.617	65.7	0	NS	0.86	F6	-0.526
20	60.3	0.001	NS	0.82	F6	-0.733	65.1	0	NS	0.86	F6	-0.558
5	69.8	0.008	0.043	0.73	F7	0.573	77.1	0.007	0.088	0.73	F7	-0.456
6	67	0.006	NS	0.71	F7	0.524	75.9	0.007	0.064	0.72	F7	-0.439

7	76.5	0.055	0.028	0.65	F7	0.700	81.3	0.087	0.029	0.63	F7	-0.544
27	64.8	0	NS	0.82	F6	-0.624	68.3	0	NS	0.86	F1	**0.328**

[a] Floor effect = *Never* or *Not at all* less than 40% of responses. [b] Item Total Correlations (ITC): <0.32 indicates lack of association, whilst ITC > 0.8 indicates duplication, none are below the 0.32 threshold. [c] factor loading less than 0.40 are highlighted in bold. NS = Not significant at 0.10% level (2-Tailed)

6.3.2.2 Steps II & III: Using Rasch and psychometric analysis to exclude items

The Rasch analysis was carried out using the frequency scale response option. Seven models were estimated independently (F1 to F7) in line with the item groupings identified in Step I. The goodness of fit statistics for each of these models is reported in **Table 6.4**. Each of the items included in models F2, F3, F6 and F7 displayed threshold disordering. Some of the items included in models F1, F4, and F5 displayed threshold disordering and were thus re-estimated once this was corrected. Model F1 was re-estimated once more due to the findings from the assessment of the characteristics of each individual item (see **Table 6.4**). Once all of the necessary adjustments were made, the fit of the majority of the Rasch models, with the exception of F5, was acceptable based on the thresholds discussed above. In model F5 the fit statistics were better when item threshold disordering was not corrected.

Table 6:4 Rasch model goodness if fit by dimension N=315* (Step II)

FREQUENCY SCALE							
Model (N)	**X^2 Goodness of fit**	**Degrees of freedom (X^2)**	**p-value (X^2)**	**Person Separation Index (PSI) with extm (No- extm)**	**Cronbach's Alpha with extm (No extm)**	**Item fit residual Mean (SD)**	**Person fit residual Mean (SD)**
F1 (N=238)	43.65	36	0.18	0.83 (0.86)	0.94 (0.92)	(-)0.04 (1.43)	(-)0.40 (1.34)
F1ordered (N=238) [a]	78.24	36	0.00	0.84 (0.87)	0.91 (0.90)	(-)0.16 1.81	(-)0.34 1.26
F1ordered & Item 17 deleted (N=223) [b]	41.61	32	0.12	0.83 (0.85)	0.93 (0.89)	(-)0.08 (1.21)	(-)0.33 (1.21)
F2 (N=180)	9.12	12	0.69	0.78 (0.77)	0.92 (0.79)	0.17 (1.73)	(-)0.64 (1.16)
F3 (N=236)	25.70	16	0.058	0.79 (0.75)	0.88 (0.77)	0.34 (2.22)	(-)0.37 (0.99)
F4 (N=138)	26.11	16	0.05	0.36 (0.54)	0.87 (0.72)	0.19 (0.72)	(-)0.32 (0.99)
F4ordered (N=138) [c]	20.96	16	0.18	0.33 (0.57)	0.85 (0.68)	0.30 (1.06)	(-)2.06 (1.25)
F5 (N=194)	22.09	12	0.04	0.56 (0.55)	0.83 (0.61)	0.44 (1.37)	(-)0.30 (0.93)
F5ordered (N=194) [d]	37.79	12	0.00	0.53 (0.50)	0.83 (0.62)	0.02 (2.47)	(-)0.35 (0.93)
F6 (N=130)	3.15	12	0.99	0.82 (0.83)	0.96 (0.85)	(-)0.21 (1.18)	(-)0.97 (1.55)
F7 (N=128)	5.54	12	0.94	0.42 (0.65)	0.89 (0.75)	0.27 (0.30)	(-)0.38 (1.00)

*Some of the tests statistics in Rasch analysis excludes individuals with extreme scores (individuals responding Never or Always to all items in the instrument). [a] Response levels for items 21, 24, and 28 were collapsed. [b] Response levels for items 21, 24, and 28 were collapsed and item 17 was deleted. [c] Response levels for items 13 and 14 were collapsed. [d] Response levels for item 18 were collapsed

Table 6.5 provides details of the Rasch analysis for each of the 29 items for the base-case models. The ordering of thresholds did not lead to any changes in the key item characteristics, with the exception of item 17 in model F1. Once all of the response options for items included in model F1 were ordered, the fit residual for item 17 changed to 3.708, which was above the threshold value of 2.5 (as discussed above) and a significant probability of model mis-fit was observed. Based on this, the F1 model was re-estimated by accounting for disordered thresholds and excluding item 17. **Table 6.5** also shows that 8 items (items 4, 13, 14, 16, 17, 18, 23 and 26 – in italics in the table) could be excluded from further analysis based the findings of the psychometric analysis in Step I and Rasch analysis in Step II. Items were excluded if any of the following was observed: low factor loadings, significant Rasch model mis-fit, fit residuals over +/- 2.5 the threshold value, disordered thresholds, or DIF by age or gender. None of the items included in model F7, came out as a strong candidate in its original five level form as each of these items displayed a problem either with low factor loadings, DIF or disordered thresholds. For this group only, items were excluded from further analysis only if DIF or low factor loadings were observed. Once the Rasch and psychometric assessments were completed, the selection of one item from each of the seven factors (F1 to F7) was undertaken. For factor F4, item 15 was the only one that did not breach any of the psychometric or Rasch conditions, and so this item was selected. For the remaining six factors, re-assessment of the qualitative interview transcripts was used to aid the selection of the key issues that were raised, in addition to the re-assessment of the remaining psychometric item characteristics (identification of floor effects and differentiation between weight groups). This lead to the selection of the following items from each of the six factors: Item 22 from F1, item 10 from F2, item 3 from F3, item 12 from F5, item 27 from F6 and item 5 from F7 (the seven items that were selected are in bold in **Table 6.5**).

Table 6.6 summarises the justification behind the final selection of each item from each of the factors.

Table 6:5 Summary of Rasch analysis results N= 315 (Step II cont.)

Factor *Item*	**FREQUENCY SCALE**						
	Factor loading	**Rasch misfit (p<0.01)**	**Rasch residual (+/- 2.5)**	**Disordered threshold**	**DIF Sex (p<0.01)**	**DIF Age (p<0.01)**	**Exclude item**
Factor 1 (F1)		**None**					
Item 17. I worry about my health in the future	*0.39*		*2.88*				*Yes*
Item 21. I feel uncomfortable or embarrassed getting changed in front of others	0.59			Yes			
***Item 22*. I feel uncomfortable or embarrassed shopping for clothes**	**0.62**			**Yes**			
Item 23. I feel uncomfortable or embarrassed meeting new people	*0.39*						*Yes*
Item 24. I feel uncomfortable or embarrassed eating in front of others	0.59			Yes			
Item 25. I feel unhappy because I can't eat what I want	0.64			Yes			
Item 26. I feel unhappy about the way I look	*0.77*					*Yes*	*Yes*
Item 28. I feel disappointed because clothes aren't made in the size I need	0.56			Yes			
Item 29. I struggle to keep in control of what I eat	0.69			Yes			
Factor 2 (F2)		**None**	**None**	**None**	**None**	**None**	
Item 8. I struggle to keep up with others when doing physical activity	0.66						
Item 9. I struggle to keep up with others when I play sports	0.73						
***Item 10*. I avoid doing things like running, cycling, swimming or playing sports**	**0.58**						
Factor (F3)		**None**		**None**	**None**		
Item 1. I have body pain / ache	0.62						
Item 2. I get low energy	0.81						
***Item 3*. I get tired**	**0.90**						
Item 4. I get out of breath	*0.42*		*2.84*			*Yes*	*Yes*
Factor (F4)		**None**	**None**		**None**		
Item 13. I get treated differently at school, such as being teased or picked-on or left out	*0.60*			*Yes*		*Yes*	*Yes*
Item 14. I get treated differently at home, such as being teased or picked-on or left out	*0.44*			*Yes*			*Yes*

Factor *Item*	**FREQUENCY SCALE**						
	Factor loading	**Rasch misfit (p<0.01)**	**Rasch residual (+/- 2.5)**	**Disordered threshold**	**DIF Sex (p<0.01)**	**DIF Age (p<0.01)**	**Exclude item**
***Item 15*. People treat me differently when I go out**	**0.68**						
Item 16. I avoid playing / hanging out or socialising with others	*0.665*					*Yes*	*Yes*
Factor (F5)		**None**	**None**		**None**	**None**	
Item 11. I struggle to do as well as others at school	-0.79						
***Item 12*. I struggle to concentrate on school / college work**	**-0.93**						
Item 18. I worry about the type of job/career I will be able to have	-0.42			Yes			*Yes*
Factor (F6)		**None**	**None**	**None**	**None**	**None**	
Item 19. I feel angry or annoyed because I am unable to do the same things as others	-0.62						
Item 20. I feel frustrated because I am unable to do the same things as others	-0.73						
***Item 27*. I feel unhappy because I am unable to do the same things as others**	**-0.62**						
Factor (F7)		**None**	**None**	**None**	**None**	**None**	
***Item 5*. I struggle to keep up when I am walking around with others**	**0.57**						
Item 6. I struggle when I am going up stairs	0.52						
Item 7. I struggle to reach or bend down	0.70						

Table 6:6 Justification of the selection of the final 7 items (Step III)

Factor	Item No.	Review of qualitative interview transcripts	Reasoning based on quantitative parameters	General justification	Original question wording	Finalised question wording
F1	22	The majority of interviewees mentioned feeling conscious of their body	Once the thresholds were re-coded, item 22 provided the best overall results as per the Rasch and SPSS analysis. It had one of the lower 'never' responses (floor effect)	Clothes and self-image were very strongly referred to in the interviews, more so than food, so it was felt that an 'image related' item should be included. Getting changed was thought to be too general as people generally may not feel comfortable changing in front of others. To make the item simpler 'embarrassed' was chosen over 'uncomfortable' as it could be misinterpreted. Also embarrassed taps into the *social* concepts of QoL as individuals are likely to get embarrassed when others are around	I feel uncomfortable or embarrassed shopping for clothes	I feel embarrassed shopping for clothes
F2	10	Avoiding or not being able to do activities because of weight brought up by respondents	This item shows a good performance in the Rasch analysis, although it had a higher floor effect and a poorer factor loading than other items in this factor grouping. It was, however, the only item that dealt with avoidance of physical activity and performed best at discriminating between weight categories	The item was shortened and examples taken out in order to make it less context specific. A choice was made to change 'playing sports' to 'doing sports' as 'playing' makes it sound more like fooling around with friends. Things like PE or Games lessons at school may fall out of this 'playing' context	I avoid doing things like running, cycling, swimming or playing sports	I avoid doing sports
F3	3	The phrase 'I get tired' was brought up in a number of the interviews especially in relation to running or sports or walking up hills	This was the best performing item in both the SPSS and Rasch analysis over all the 29 items. It also had the lowest number of 'Never' responses (least floor effect) of the 3 items included in this factor grouping	It was clear that this item should be included as it performed the best on all aspects and fully reflected adolescents own original wordings	I get tired	I get tired
F4	15	Interviewees mentioned being	In this factor grouping, item 15	There was no need for the inclusion of	People treat me	People treat

Factor	Item No.	Review of qualitative interview transcripts	Reasoning based on quantitative parameters	General justification	Original question wording	Finalised question wording
		treated differently however they do not specifically specify a setting, i.e. home or school. They just spoke about being treated differently in general or when going out. It was not so common to be treated differently by family members but this was mentioned in some of the interviews	performed best for the SPSS and Rasch analysis. Item 16 displayed DIF by age group. Items 13 and 14 displayed disordered thresholds and the former displayed DIF by age group also. Item 14 loads the lowest onto the factor would not have provided the best item to represent the factor	examples for this item. Examples might have created a potential problem if the example did not suit the individual respondent as they may think that the issue does not relate to them	differently when I go out	me differently when I go out
F5	12	The majority of school specific discussion related to being treated differently or unable to keep up in PE, the School FG and some of the Carnegie FG participants mentioned concentration in school	Item 12 had the highest factor loading and performed the best in the SPSS and Rasch analysis and so was chosen. Item 11 had a higher floor effect and item 18 displayed dis-ordered thresholds	Clearly this was the best performing item from this factor grouping. The item needed to be amended in order to work for adolescents who are working and no longer in fulltime education	I struggle to concentrate on school / college work	I struggle to concentrate on my studies / work
F6	27	Most respondents referred to feeling frustrated or annoyed or unhappy when they could not do the same things as others or just in terms of how they felt about themselves	Item 27 had a high factor loading, but the floor effect was not the best from all the factors in this group. Rasch analysis for all 3 items in this group was similar	This item was chosen because out of all the emotions, it was felt that 'unhappy' would be common to all ages. In terms of the other emotions 'frustrated' 11 year olds may struggle with this word. Item 19 refers to ‘angry or annoyed', for simplicity in the valuation study to follow, there was a preference for a single word to be used	I feel unhappy because I am unable to do the same things as others	I feel unhappy because I am unable to do the same things as others
F7	5	Difficulty with walking up hills was mentioned by a number of the respondents. Having difficulty with stairs was only mentioned by a few respondents	Both items 7 & 5 had the highest factor loadings and performed best in the Rasch analysis. Item 5 was chosen over item 7 because had a lower floor effect than item 7. Item 6 was not chosen because of low loading on the factor	The other 2 items were too context specific	I struggle to keep up when I am walking around with others	I struggle to keep up when I am walking around with others

6.3.3 Analysis of the final item set

6.3.3.1 Psychometric analysis

The results of the psychometric assessments on the reduced scale appeared to be adequate (see **Table 6.7**), displaying internal consistency, moderate floor effects, and the ability to discriminate between weight categories.

Table 6:7 Psychometric analysis of items from the reduced instrument N=315

Items	**Mean**	**SD**	**Scale mean if item deleted**	**Scale variance if item deleted**	**Corrected item-total correlation**	**Cronbach's Alpha if item deleted**
3. I get tired	2.54	1.234	10.54	27.829	0.624	0.874
5. I struggle to keep up when I am walking around with others	1.54	0.951	11.54	29.357	0.699	0.866
10. I avoid doing things like running, cycling, swimming or playing sports	1.96	1.271	11.13	26.901	0.679	0.867
12. I struggle to concentrate on school / college work	1.93	1.156	11.16	28.316	0.636	0.871
15. People treat me differently when I go out	1.48	0.900	11.60	30.348	0.636	0.873
22. I feel uncomfortable or embarrassed shopping for clothes	1.91	1.258	11.17	26.554	0.720	0.861
27. I feel unhappy because I am unable to do the same things as others	1.72	1.147	11.36	27.136	0.754	0.856
Total score	**Mean**	**SD**	**Cronbach's Alpha**	**P-value (from t-test for discrimination between Norm_WT & Over_WT)**	**P-value (from t-test for discrimination between Over_WT & Obese)**	**P-value (from t-test for discrimination between Normal_WT & Obese)**
All (N=315)	13.08	6.12	0.88	0.00	NS	0.00
Norm_WT (N=152)	10.68	4.13	NA	NA	NA	NA
Over_WT (N=51)	14.37	7.48	NA	NA	NA	NA
Obese (N=112)	15.75	6.47	NA	NA	NA	NA

6.3.3.2 Rasch analysis

Rasch analysis on the overall seven items showed that the scale fits the Rasch model after Bonferroni adjustment (p=0.00). None of the items had a residual greater than +/- 2.5 and only one item had a Chi Square probability of less than 0.01 (item 27). Overall the statistics indicated minimal significant individual item misfit (see **Table 6.8**). Combining the final set of seven items showed a number of items with dis-ordered thresholds.

Table 6:8 Rasch analysis item property summary for the reduced instrument N=247*

Assessments	**Item 3.** I get tired	**Item 5.** I struggle to keep up when I am walking around with others	**Item 10.** I avoid doing things like running, cycling, swimming or playing sports	**Item 12.** I struggle to concentrate on school / college work	**Item 15.** People treat me differently when I go out	**Item 22.** I feel uncomfortable or embarrassed shopping for clothes	**Item 27.** I feel unhappy because I am unable to do the same things as others
Total No. 'Never' responses	26	152	110	98	161	115	136
Factor loading	0.9	0.57	0.58	-0.93	0.68	0.62	-0.62
Item location	-0.9	0.73	-0.23	-0.07	0.62	-0.21	0.06
SE	0.08	0.08	0.07	0.07	0.09	0.07	0.07
FitResid	0.875	(-)1.76	(-)0.06	1.14	-0.56	-0.56	-2
DF	207.86	207.86	207.86	207.86	207.86	207.86	207.86
ChiSq	3.78	9.3	4.53	7.23	6.44	4.56	13.56
p-value	0.44	0.05	0.34	0.12	0.17	0.34	0.01
Disordered thresholds		Yes	Yes	Yes	Yes	Yes	Yes

*Sample size excluding individuals with extreme values N=247 out of the full sample N=315

The results of the Rasch analysis on the final seven items also supported the assumption of uni-dimensionality as a t-test statistic[10] of 0.024 was calculated. As this value was less than the 0.05 threshold value there was evidence to support the assumption of uni-dimensionality, i.e. scale scores could be aggregated across the seven items meaningfully. A test of local

[10] The t-test protocol is carried out on an individual person basis and is used to identify any differences in the estimates that have been generated. If the proportion of t-tests falls outside of the boundaries of acceptable significance, this indicates that the items within a scale display some degree of multidimensionality. The number of significant t-tests was counted to see if more than 5% of the tests fall outside of the acceptable significance range.

dependency was carried out to investigate whether the response to one item directly influenced the response to another by examining the correlations among the residuals (the difference between the observed and expected values) using a test statistic of + / - 0.349 (the mean residual correlation across the seven items, plus 0.2 (Hobart & Cano (2009)). The results showed that the seven items were independent (see **Appendix 6.6**). **Appendix 6.7** provides the person item distribution which indicates that the scale-to-sample targeting was adequate for making judgements about the performance of the scales and the measurement of people, as the person distribution closely matched item distribution when both calibrated on the same metric scale. This new seven item instrument was named the **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (WAItE).

6.4 Discussion

The aim of this study was to develop a weight specific outcome measure suitable for the elicitation of preference values. A three step approach was taken to develop a reduced instrument, from the original 29 items, that is brief, easy to use and practical to administer in any setting. The final seven items were tested against Rasch measurement principals showing that each of the seven items were independent and adequately met the various test requirements. Rasch analysis showed that there were a number of items with dis-ordered thresholds, but this was likely due to the high number of '0' scores leaving insufficient numbers to obtain robust values for the other response categories. This weight specific instrument is likely to be more sensitive to changes in BMI, by picking up changes in weight specific dimensions of QoL, than other generic preference based measures available, but this remains to be determined.

Some limitations in terms of the respondent sample and the performance of the final seven item instrument should be noted. An internet panel made up the majority of the sample. Though there were similarities when compared to the LSYPE, a national data set, in terms of physical activity and computer use there were differences between the samples with respect to the employment status of carers and type of accommodation. In terms of the final seven items that were chosen the following were observed: there were a large number of *Never* responses observed for some of the items; however, this may be because approximately 45% of the sample was classified as normal weight after adjusting for age and gender. Sensitivity of the measure to changes in BMI will be assessed in Study 3 (see **Chapter 7**) which will be an important determinant of the validity of the new measure.

When the seven items are combined in the Rasch analysis, the majority of the items displayed threshold dis-ordering (where item response categories that do not function as intended in the Rasch model – a more detailed description is given in **Section 6.2**). This may again be due to the high number of 'Never' responses. Ideally, for ease of completion and completeness, each item in the final set of seven should have the same number of response options. Though it is worth noting, when the Rasch analysis was undertaken on the seven dimensions independently, only item 22 - *I feel embarrassed shopping for clothes* (re-worded version) displayed disordered thresholds when the original five level response options were not collapsed, although the contrary was observed for the remaining six items. Given this finding, two options are available: a) collapsing all the other six items to four level response options per item and collapsing the same levels across all items or b) keeping the original five level response options for all of the items. The response options will be kept to five levels for the valuation study so that further assessment of threshold can be carried out as the behaviour of the regression coefficients will provide further indication of whether levels need to be collapsed. This decision was taken because the Rasch analysis thus far suggests that 'Almost never' and 'Sometimes' may be collapsible given their frequencies in the self reported data. Until the data from the valuation study is available, however, it is unknown whether those valuing these states value these levels differently. Thus it may be that the levels that need collapsing might not be the same in the two contexts.

Existing studies employing combined psychometric and Rasch analysis in the refinement of instruments have validated their findings by repeating their analysis on another independent sample, a so called *validation* sample, in order to test whether the same findings are reached (for example see Young *et al*., 2011). In the current study analysis on a validation sample was not undertaken due to time and resource constraints. It is, however, important to note that this current study is different in that the application of the psychometric assessment and Rasch analysis was carried out at the instrument development stage, whist other studies have applied this methodology to established instruments. The refinement of existing instruments, not originally developed for the purposes of preference elicitation, would be more likely to necessitate validation testing. From the outset, this has been the purpose of the instrument developed in the current study and may therefore imply that the testing of the findings using a validation sample may not be necessary. Nevertheless, in the final empirical study detailed in **Chapter 8**, Rasch analysis was undertaken on the data obtained from a sample of adults who completed the WAItE. This provided an opportunity to test the psychometric properties of the WAItE on an independent sample of respondents. Further validation could be investigated on an adolescent sample in future work.

6.5 Conclusion

In this chapter, the **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (WAItE) was created. It was possible to undertake the refinement of a long list of items and derive the WAItE from the combined information obtained from: interviews with adolescents, psychometric assessments, and Rasch analysis that were reported in the previous and current studies. The WAItE consists of seven items each with five corresponding frequency response options. In its current form the instrument cannot be used for the calculation of QALYs and a valuation study is necessary for this. The feasibility of undertaking this type of study, utilising the WAItE, is tested out in the final empirical study (**Chapter 8**). Before this, a preliminary assessment of the measurement properties of the WAItE is reported in **Chapter 7**.

Chapter 7
Psychometric properties of the WAItE: A preliminary assessment

7.1 Introduction

Once a new instrument has been created, there is a need to evaluate its psychometric properties. Undertaking the preliminary testing of the measurement properties of the newly created, seven item five level, WAItE is the focus of the study reported in this chapter (Study 3). Investigations of whether changes in weight status are related to the expected change in weight specific QoL scores, as described by the WAItE, are undertaken (Patrick *et al*., 2011). Assessment of the WAItE scores with respect to change in weight status can be assessed by the application of the QoL instrument before and after the implementation of a weight management intervention. Clinical interventions that are recommended for the management of overweight and obesity in the younger population, from the NICE guidance document (NICE, 2006) include: lifestyle, behavioural, physical, dietary, pharmaceutical, and surgical interventions. Lifestyle interventions incorporate multicomponent interventions provided by weight management programmes and strong evidence exists that weight loss camps are effective at producing weight loss in and improved QoL in obese and overweight adolescents (Gately *et al*., 2000 & 2005, Patrick *et al*., 2011, Griffiths *et al*., 2010). Testing of the sensitivity of the WAItE to a) change in weight status and b) weight loss, associated with weight management, was undertaken. The following hypotheses were evaluated, through the application of the WAItE in the context of a summer weight management camp (see Mokkink *et al*., 2010b for agreed definitions of the latter two concepts):

Practicality

- Are adolescents able to self complete the WAItE?

Internal consistency

- Are the seven items that make up the WAItE measuring the same underlying construct (i.e. is the WAItE uni-dimensional)?

Responsiveness

- Is the WAItE able to pick up changes in QoL associated with changes in weight status?
    - Are the WAItE total scores and the seven individual WAItE item scores responsive to change in QoL resulting from weight change?

7.2 Methods

7.2.1 Participants

The pilot study utilised a convenience sample of respondents enrolled in a six-week residential weight loss camp during the 2012 summer vacations, by *More Life* (formerly *Carnegie International Camp*)[11]. Although *More Life* is based in Leeds, the programme recruits from all over the United Kingdom (as well as a small number of participants from Europe) through a range of sources, including self/parental referral, medical referral, or referral from social services, primary care trusts (PCTs) or educational organizations. Acceptance into the program was contingent on having a BMI above cut-off values for above normal weight status, and health screening was performed by the family general practitioner (Gately *et al*., 2005).

All campers were eligible for inclusion in the study, unless the camp staff indicated otherwise. Study invitation letters were provided to campers and signed informed consent was required from each adolescent taking part in the study (passive parental consent was assumed) (see **Appendix 7.1**). Each participant was required to complete the WAItE at two time points: baseline (T1) and at the end of the programme (follow-up T2). Details regarding the weight and height of each study participant were obtained from the camp records (see **Section 7.2.3** for further details). Ethical approval was provided by the University of Leeds local research ethics committee (Ref: HSLTLM/11/049).

Data was obtained from all 30 adolescents that were eligible to take part in the study, 19 girls (63%) and 11 boys with a mean age of 14 years (SD: ±2.29) who attended the *More Life* camp. The study participation rate was 100% on the first day of camp; this dropped to 90% (n=27) at follow-up. At T2 the questionnaire was administered by the camp staff via telephone: one male participant (aged 10 years) did not give their consent to complete the WAItE, and another two male participants (both aged 14 years) could not be reached. No data was collected on ethnicity or parental employment status for campers, although the majority of participants were white. Some adolescents joined the programme after the official first day of camp; however, only data for individuals who were present on the first day of camp was collected. The majority of participants completed the full six week period of the camp (n= 24, 80%). The remaining participants either left after they completed week five (n=4), four (n=1) or three (n=1) in the camp. In **Table 7.1** a summary of the respondents who completed the questionnaire in the two time points is provided.

[11] This was one of the organisations used to recruit adolescents in Study 1 (**Chapter 5**)

Table 7:1 Characteristics of respondents who completed the WAItE at T1 and T2

		T1 (n=30)		T2 (n=27)	
		Frequency	**Per cent**	**Frequency**	**Per cent**
Gender	Male	11	36.7	8	29.6
	Female	19	63.3	19	70.4
Age (yrs)	9	1	3.3	1	3.7
	10	3	10.0	2	7.4
	11	1	3.3	1	3.7
	12	1	3.3	1	3.7
	13	5	16.7	5	18.5
	14	6	20.0	4	14.8
	15	6	20.0	6	22.2
	16	3	10.0	3	11.1
	17	3	10.0	3	11.1
	18	1	3.3	1	3.7
Weeks at camp	3	1	3.3	1	3.7
	4	1	3.3	1	3.7
	5	4	13.3	4	14.8
	6	24	80.0	21	77.8

7.2.2 Programme and procedure

The *More Life* summer camp runs annually in Leeds and was established in 1999 by Carnegie Weight Management, Leeds Metropolitan University, and it caters for young people aged 8 to 18 years old. A brief description of the *More Life* weight loss summer camp is provided below. A more detailed description can be found in the report by Gately *et al*. (2005). Evaluations of the camp programme have found that, in the short term, the programme is effective across a range of health outcomes, demonstrating significant reductions in BMI, fat mass, blood pressure, waist circumference, and improvements in aerobic fitness and self-esteem in participants.

The camp is based in the premises of a boarding school, which provided catering, residential, educational, and high-quality indoor and outdoor sports facilities. All children and staff members were housed on site throughout the summer period. The duration of stay, of adolescents, in the camp ranged from three to six weeks (mean: 5.7 (SD: ±0.70)). The daily programme consisted of diet, nutrition and culinary education as well as physical activity combined with a range of structured, fun-type, skill-based activities. There was moderate dietary restriction which was personalised to each individual. Energy intake was based on an approximation of basal metabolic rate (5460–13 860 kJ/day) and meals were designed to be similar to the food to which adolescents would be exposed in their home environment (i.e. foods such as salad, pasta, and pizza) (Gately *et al*., 2005).

7.2.3 Measures

Anthropometric measures - BMI

As part of the standard protocol, campers had weight and height measurements taken weekly by trained staff. Standardized body mass index (BMI-SDS scores) were computed according to age and gender. For each study participant, data on weight, height, age, and gender was provided by *More Life*.

Weight specific QoL – the WAItE

In addition to the standard measures collected by *More Life* the WAItE (see **Appendix 7.2**) was completed by adolescents in order to assess weight specific QoL. At baseline (T1) the WAItE was administered to all campers who were present on the first day of camp to self complete, by the candidate with the aid of camp staff. All campers that were present on the first day of the programme were eligible to complete the WAItE regardless of age. Some campers were under 11 years but were not excluded as they were willing and able to self complete the instrument (and the camp staff were happy for these participants to be included in the study). Additionally it would not have been feasible to exclude the campers that were under the age of 11 years, as self completing the questionnaire was undertaken as group exercise. The measure was administered, a second time (T2), a few weeks after the camp ended when follow-up phone calls to each of the adolescents who attended the camp were undertaken. Follow-up calls were carried out by camp staff and the WAItE measure was administered within the calls once agreement from adolescents was obtained. The use of different modes of administration over the two time periods was not ideal. Campers might have felt pressured to give positive responses when the WAItE was administered over the phone and so the limitations of this need to be borne in mind.

A weight specific QoL score was calculated for each individual completing the WAItE in the two time periods by adding up the score corresponding to each response option for each of the seven items. Response options were scored as follows: Never = 1, Almost Never = 2, Sometimes = 3, Often = 4 and Always = 5. A WAItE total score of 7 translated to the best level of QoL whilst a score of 35 was the worst level of QoL described by the instrument.

7.2.4 Analysis

Practicality (Brazier *et al*., 1999) was tested by looking at the response rate, the completion rate, the understanding of the instructions and the wording of the WAItE items and response options. *Internal consistency* (Mokkink *et al*., 2010b) was evaluated with the calculation of Cronbach's alpha coefficient that is commonly used as an indication of the degree to which a set of items measures a single uni-dimensional latent construct. *Responsiveness* (Mokkink *et al*., 2010b) was evaluated through a paired samples t-test to draw a comparison between T1 and T2 weight status and WAItE scores (Gately *et al*., 2005 and Patrick *et al*., 2011). Assessments of the relationship between the WAItE score and weight change was assessed through correlation analyses. The primary analysis involved correlation tests between change in BMI-SDS and the WAItE total scores between follow up and baseline (T2 – T1, a negative value from the subtraction equates weight loss (lower weight status in T2 than T1) or an improvement in weight specific QoL (lower WAItE score in T2 than T1, as a WAItE score of 7 = best QoL and a score of 35 = worst QoL). This statistical test provides insight on the relationship (if any) between the two variables. The Spearman correlation coefficient (a non-parametric test) was calculated, as this does not assume that the data are normally distributed (given the small sample size). Secondary analysis looked at the correlation between T1 and T2 scores between the two variables to assess whether there was a relationship between the absolute values. The SPSS statistics package (Version 18.0) was used to carry out the analyses (SPSS Inc, 2009). All the analysis was conducted only on the 27 individuals who completed questionnaire at both T1 and T2.

It was not possible to assess other relevant psychometric properties such as test-re-test reliability due to time and resource constraints. It was also not possible to conduct a multiple regression analysis because there were not enough participants to carry out this type of analysis.

7.3 Results

7.3.1 Testing the practicality of the WAItE

The completion rate was excellent as 100% (n=30) of participants completed the WAItE at T1 and, of those that could be contacted, only one participant refused to complete the questionnaire at T2. Participants self completed the questionnaire at T1, and as it was administered via telephone at T2. Each response level was picked at least once across each of the WAItE items at T1. Levels 5 and 4, indicating poorer levels of QoL were sometimes never chosen at T2, which would be expected if QoL is improving as a result of the intervention. **Appendix 7.3** displays the distribution of scores across response options. The majority of the participants were able to read the instructions and self completed the questionnaire without any help (camp staff read out the seven items of the WAItE and completed the questionnaire based on the verbal responses of two adolescents).

7.3.2 Testing the internal consistency of the WAItE

A Cronbach's Alpha statistic of 0.812 and 0.854 were calculated based on the data from the 27 participants who completed the questionnaire at T1 and T2, indicating that the seven items from the WAItE were uni-dimensional, and display good internal consistency, over both time periods. **Table 7.2** presents the results for the Cronbach's Alpha statistic assuming one of the WAItE items was deleted. For the analysis undertaken on the data at T2, each of the estimated alphas was less than 0.854 (where all seven items were included), thus suggesting that none of the seven WAItE items should be excluded. The analysis undertaken on the data from T1, however, suggests that perhaps the WAItE item *I avoid doing sports* may be excluded as the value of alpha is greater when this item is deleted (0.820 as opposed to 0.812). As this finding was not observed on the data obtained in T2 the item was not excluded. The inconsistent results suggest that furthest tests should be carried out on a larger dataset in future work.

Table 7:2 Cronbach's Alpha if one questionnaire item were to be deleted

WAItE item N=27	Cronbach's Alpha if Item Deleted (T1)	Cronbach's Alpha if Item Deleted (T2)
I get tired	0.793	0.828
I struggle to keep up when I am walking around with others	0.752	0.827
I avoid doing sports	**0.820**	0.844
I struggle to concentrate on my studies / work	0.806	0.841
I feel embarrassed shopping for clothes	0.794	0.828
I feel unhappy because I am unable to do the same things as others	0.764	0.820
People treat me differently when I go out	0.775	0.849

7.3.3 Testing the responsiveness of the WAItE – by total score

Based on the age and gender adjusted BMI-SDS calculations, there were 26 (96%) obese adolescents and one (4%) overweight adolescent in the study. **Table 7.3** displays the mean and standard deviation (SD) calculations for the weight status measures, the WAItE total score and any change that occurred between baseline and follow-up. Weight loss was detected between baseline and follow-up in all the weight measures (all data are presented as means (the summation of the total WAItE score across all 27 individuals divided by the total number of individuals) and ± SD): including decreases in mean BMI (from 34.13 ± 10.94 to 30.93 ± 10.38), and mean BMI-SDS (from 2.94 ± 0.80 to 2.50 ± 0.92). The mean change in BMI-SDS was (-0.44 ± 0.17). The mean total WAItE score at follow-up was better than the mean score at baseline (from 18.26 ± 6.04 to 15.67 ± 5.57 - lower scores indicate better QoL), with an improvement in the mean difference between the two time periods (-2.59 ± 4.65). The paired samples t-test detected a significant difference between baseline and follow-up BMI (n=27, t=18.81, p<0.01). When the t-test was applied on the WAItE total scores at T1 and T2, a significant difference between the scores in the two time periods also detected (n=27, t=2.90, p<0.01).

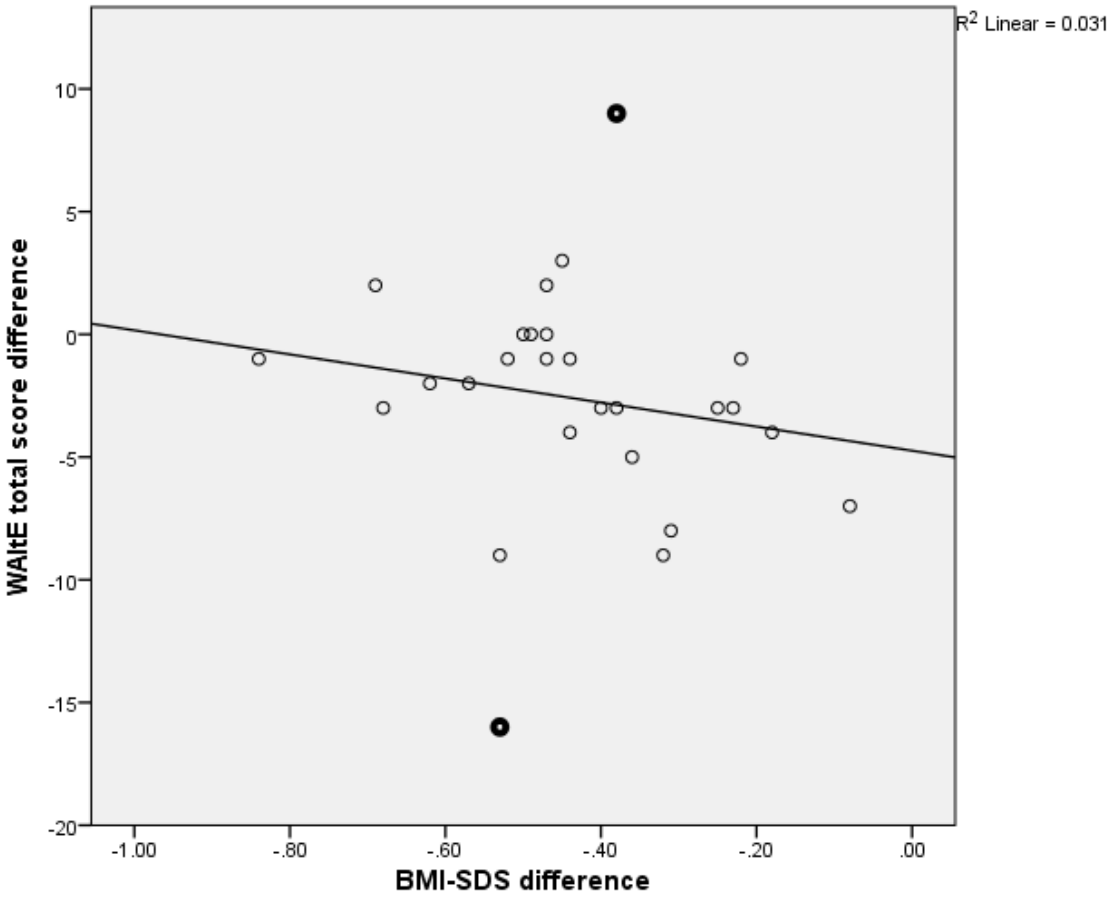
Table 7:3 Mean, and SD weight status measures and the WAItE total score

Weight status	**T1 (n=27)**		**T2 (n=27)**	
	Mean	**SD**	**Mean**	**SD**
Height (M)	1.66	0.12	1.67	0.13
Weight (KG)	95.39	37.50	87.96	35.19
BMI	34.13	10.94	30.93	10.38
BMI-SDS	2.94	0.80	2.50	0.92
WAItE total score	18.26	6.04	15.67	5.57
Weight status	**Mean**		**SD**	
BMI difference	3.19		0.88	
BMI-SDS difference	-0.44		0.17	
Percentage weight loss (KG)	8.98		2.05	
WAItE total score difference	-2.59		4.65	

7.3.3.1 Primary correlation analysis

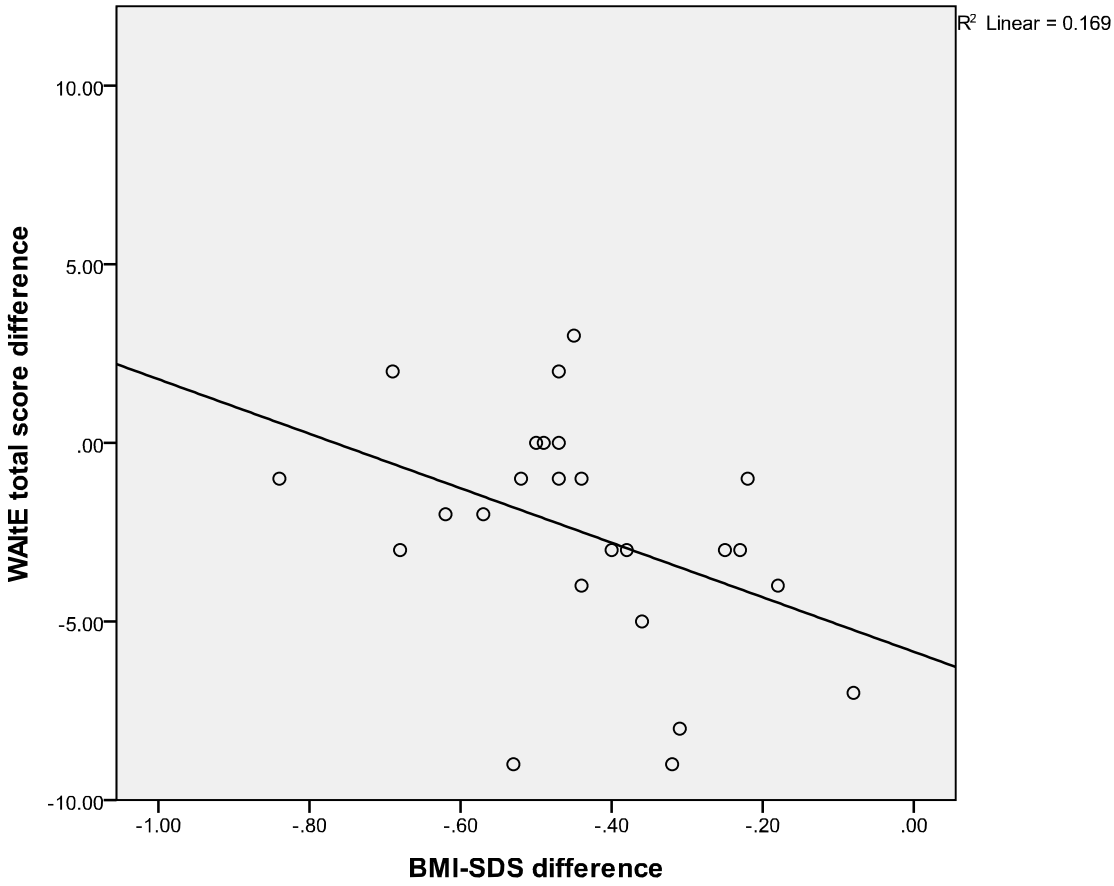
Figure 7.1 shows the scatter plot of the change in BMI-SDS against the change in total WAItE score (T2 – T1, negative values equate an improvement in either weight status or weight specific QoL between the two time periods). The Spearman's rho correlation coefficient was -0.31 (n=27, p=0.12). This indicated that over the two time points decreased BMI-SDS (improved weight status) was associated with decreased WAItE scores (improved weight specific QoL). The association between the two variables was not found to be statistically significant, but this may have been caused by the two outliers highlighted in bold. When these outliers were excluded from the analysis the relationship between the two parameters was statistically significant with a Spearman's rho correlation coefficient of -0.46 (n=25, p=0.02 – see **Figure 7.2**).

Figure 7:1 Scatter plot between change in WAItE score and BMI-SDS*



*Correlation is not significant at the 0.10 level (2-tailed)

Figure 7:2 Scatter plot between change in WAItE score and BMI-SDS excluding outliers*

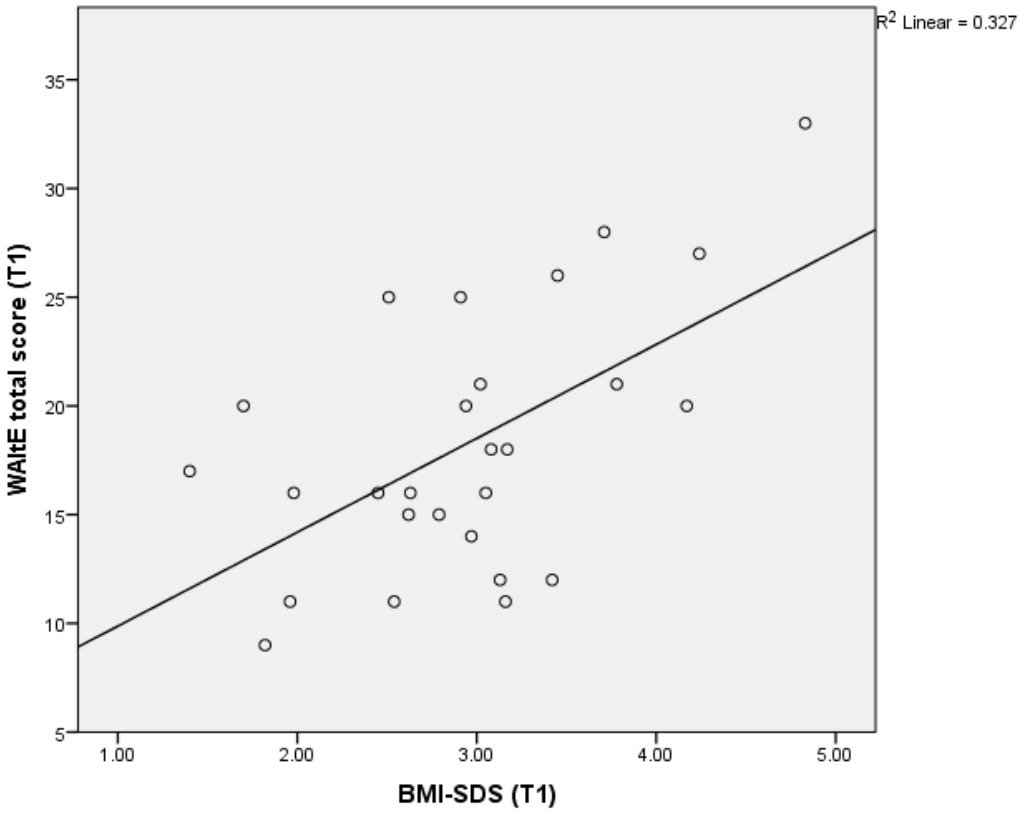


*Correlation is significant at the 0.05 level (2-tailed)

7.3.3.2 Secondary correlation analysis

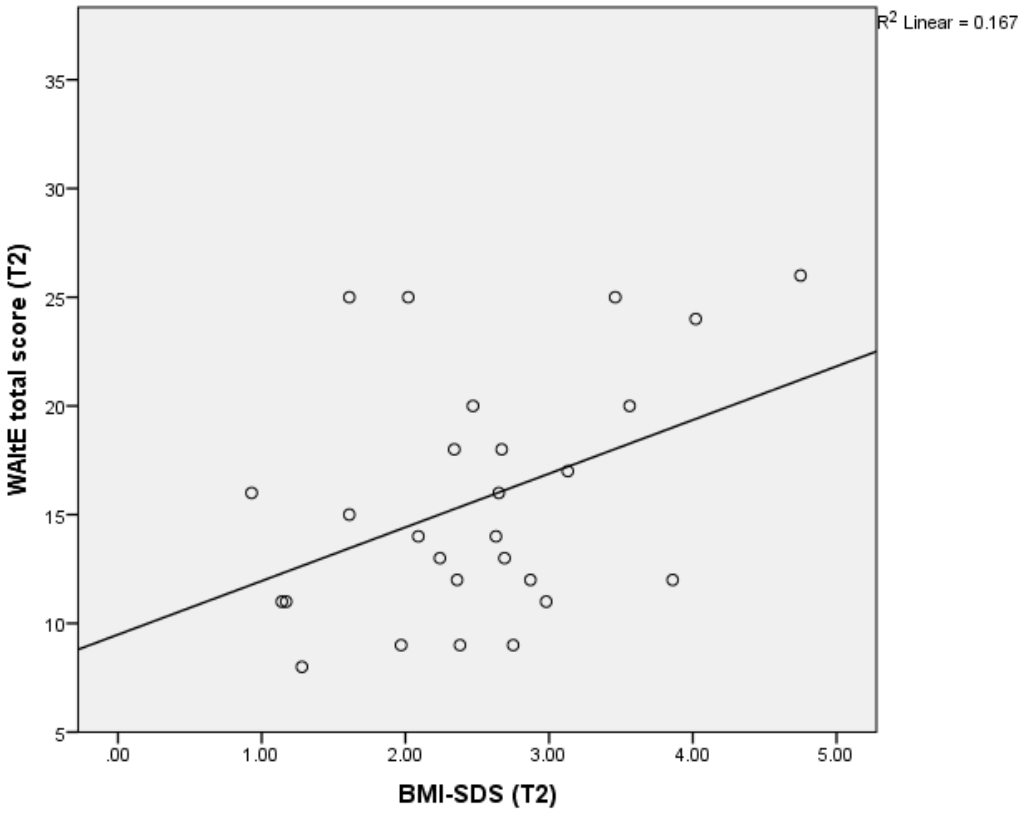
The relationship between the BMI-SDS and the total WAItE scores at T1 (**Figure 7.3**) and T2 (**Figure 7.4**) are shown in the scatter plots below. Spearman's rho correlation coefficient at T1 and T2 were +0.46 (n=27, p=0.02) and +0.31 (n=27, p=0.16) respectively, indicating that a higher WAItE score (lower weight specific QoL) was associated with higher weight status, in both time periods.

Figure 7:3 Scatter plot between T1 WAItE score and BMI-SDS*



*Correlation is significant at the 0.05 level (2-tailed)

Figure 7:4 Scatter plot between T2 WAItE score and BMI-SDS*



*Correlation is not significant at the 0.10 level (2-tailed)

7.3.4 Testing the responsiveness of the WAItE – by individual item scores

Table 7.4 describes the individual WAItE item scores and the changes in the scores between baseline and follow-up. For the majority of the individual item scores (with the exception of item 7), the follow-up weight specific QoL was better than the baseline scores. The items showing the biggest improvement between the two time periods are: item 1 *I get tired* and item 6 *I feel unhappy because I am unable to do the same things as others*. Item 3; *I avoid doing sports,* showed the least improvement over the two time points. Item 7; *People treat me differently when I go out*, seemed to get worse over the two time points. When the 2 outliers are excluded from the analysis, as shown in **Table 7.5,** apart from item 7, all the items show an improvement between the two time periods.

Table 7:4 Description of WAItE scores and changes

Item – absolute scores	**Mean**	**SD**	**Mean**	**SD**
	T1 (n=27)		**T2 (n=27)**	
1. I get tired	3.37	0.88	2.70	0.78
2. I struggle to keep up when I am walking around with others	2.41	1.31	1.81	0.88
3. I avoid doing sports	2.00	1.27	1.93	0.92
4. I struggle to concentrate on my studies / work	2.48	1.28	2.22	1.01
5. I feel embarrassed shopping for clothes	2.81	1.39	2.44	1.40
6. I feel unhappy because I am unable to do the same things as others	3.00	1.41	2.30	1.20
7. People treat me differently when I go out	2.19	1.18	2.26	1.29
Item – difference between scores(T2-T1)	**Mean**		**SD**	
1. I get tired	-0.67		0.83	
2. I struggle to keep up when I am walking around with others	-0.59		1.15	
3. I avoid doing sports	-0.07		1.14	
4. I struggle to concentrate on my studies / work	-0.26		1.32	
5. I feel embarrassed shopping for clothes	-0.37		1.42	
6. I feel unhappy because I am unable to do the same things as others	-0.70		1.14	
7. People treat me differently when I go out	0.07		1.11	

Table 7:5 Description of WAItE scores and changes – Outliers excluded

	Mean	SD	Mean	SD
Item – absolute scores	**T1 (n=25)**		**T2 (n=25)**	
1. I get tired	3.44	.87	2.76	.72
2. I struggle to keep up when I am walking around with others	2.32	1.31	1.76	.78
3. I avoid doing sports	2.04	1.32	1.92	.91
4. I struggle to concentrate on my studies / work	2.40	1.29	2.24	1.01
5. I feel embarrassed shopping for clothes	2.80	1.32	2.40	1.32
6. I feel unhappy because I am unable to do the same things as others	2.96	1.40	2.28	1.17
7. People treat me differently when I go out	2.12	1.17	2.20	1.32
Item – difference between scores(T2-T1)	**Mean**		**SD**	
1. I get tired	-.68		.85	
2. I struggle to keep up when I am walking around with others	-.56		1.04	
3. I avoid doing sports	-.12		1.17	
4. I struggle to concentrate on my studies / work	-.16		1.25	
5. I feel embarrassed shopping for clothes	-.40		.91	
6. I feel unhappy because I am unable to do the same things as others	-.68		.80	
7. People treat me differently when I go out	.08		1.12	

7.4 Discussion

This preliminary investigation of the practicality, internal consistency, and responsiveness of the WAItE used a convenience sample of adolescents enrolled in a residential summer weight loss camp to assess the practicality, internal consistency, and responsiveness of the new measure. There are two key limitations in this preliminary study that should be taken into consideration that need to be addressed in future research. Firstly the small sample size from only one study setting and secondly the different modes of administration of the WAItE between the two time periods. These issues increase the potential for obtaining unreliable findings. However the aim of this study was not to obtain definitive results about the measurement properties of the WAItE, instead, the aim was to undertake an initial preliminary investigation.

In terms of practicality of the WAItE, the overall response rate in the two time periods when the data collection for the WAItE were carried out were very high, with only one adolescent refusing to complete the questionnaire at follow-up. Cronbach's alpha coefficient was above 0.8 indicating good internal consistency between the seven WAItE items. The results of the responsiveness tests were also very encouraging. As expected, the *More Life* summer camp was successful in producing weight loss in youth attendees, providing a good opportunity to assess the association between changes in weight status and weight-specific quality of life,

using an intervention of known effectiveness. There was a significant difference between the total WAItE scores at baseline and follow-up. The mean total WAItE score at follow-up was lower than the baseline score indicating an improvement in weight specific QoL. It is worth noting that the observed improvement in weight specific QoL could be due to the process effects of intervention as opposed to weight loss *per se*, for example, being part of the camp environment might have led to feelings of greater self esteem as opposed to the effect of the weight loss in itself.

In the primary correlation analysis assessing the relationship in the difference between BMI-SDS and the total WAItE score (T2-T1), a negative Spearman's rho correlation coefficient was calculated, indicating that as weight status improves so does the weight specific QoL, as would be expected. The relationship was not however statistically significant when two outliers were included in the analysis. The review of the relationship between paediatric obesity and self-esteem or quality of life by Griffiths *et al*. (2010) (see **Chapter 3** for details about the review) showed that evidence of a significant relationship between decreased weight status and improvements in global self-esteem or total quality of life scores is weak. Of the five studies that were found to investigate these associations only one study found that weight loss was significantly correlated with self-esteem improvement. Explanations offered for this weak relationship were: a) the complexity of the causal process that operate between obesity and self-esteem and the fact that both are extremely resistant to change, b) the small sample sizes of the studies assessing this relationship and, c) the timing of when the assessments are undertaken, normally conducted at the start and end of an intervention. These observations could possibly explain the findings within this current study also.

In **Chapter 3** seven studies were identified that assessed the measurement properties of five existing weight specific QoL instruments. **Table 7.6** provides a comparison between the psychometric properties of the existing instruments and the WAItE. The Cronbach's alpha coefficient of all the instruments displayed good internal consistency, with the IWQOL-Kids and YQOL-W showing the highest alpha coefficients. When comparing the results of the construct validity, the WAItE performs better than the Sizing Me Up and Sizing Them Up instruments. Again the IWQOL-Kids and YQOL-W instruments display good construct validity. The responsiveness of the Sizing Them Up and the YQOL-W were assessed in two of the studies that were identified (Modi & Zeller, 2008 and Patrick *et al*., 2011). An inverse relationship between change in weight and QoL was observed for the YQOL-W instrument, as was observed for the WAItE. However, a statistically significant correlation between weight and change in YQOL-W score was not observed. Patrick *et al*. (2011) only reported that the

correlation between the two parameters was not significant but, they did not report the p-value. Modi & Zeller, 2008 did not provide an estimate for the correlation between change in BMI-SDS and the change in the Sizing Them Up score.

Table 7:6 Comparisons of the psychometric properties of weight specific instruments

Instrument	WAItE	IWQOL-Kids [a]	M-A QoL Q II [b]	Sizing Me Up [c]	Sizing Them Up [d]	YQOL-W [e]
Internal consistency - Cronbach's alpha coefficient	0.815	0.96	0.84	0.82	0.91	0.97
Construct validity - Correlation: absolute BMI-SDS and QoL score (inverse relationship)	Baseline: 0.40 (p=0.03) Follow-up: 0.31 (p=0.12)	-0.54 (p<0.0001)		0.14 (p=0.07)	Physical functioning scale -0.34 (p<0.0001) NS (all other scales)	-0.41 (p<0.01)
Responsiveness						
Sample size	27				220	133
Intervention	Weight loss camp				Bariatric surgery	Weight loss camp
Time period	8 weeks				6 months	4 weeks
Change in BMI (SD)	3.22 (0.86)				-20.3 (3.1)	-2.8 (1.0)
Change in BMI-SDS (SD)	0.44 (0.16)					0.2 (0.1)
Improvement in total QoL score	2.59 (t=2.90; P<0.01)				10.1 (t=13; P<0.01)	10.7 (p<0.001)
Correlation: change in BMI-SDS and QoL score (inverse relationship)	-0.31 (p=0.12)					1.0 (NS)

[a] Kolotkin *et al*. (2006). [b] Moorehead *et al*. (2003). [c] Zeller & Modi (2009). [d] Modi & Zeller (2008). [e] Morales *et al*. (2011) and Patrick *et al*. (2011).

7.5 Conclusion

The findings from this preliminary study are very encouraging and provide initial empirical evidence to support the reliability and validity of the WAItE. The purpose of the study was not to carry out a definitive validation study for the WAItE, but to begin an investigation into its psychometric properties.

Chapter 8
Valuation of the WAItE: A methodological exploration

8.1 Introduction

In order to facilitate the calculation of QALYs for CUA of weight management interventions using the WAItE, a valuation study needs to be carried out (NICE, 2013). This chapter addresses Part 3 of the three part process detailed in **Chapter 4**. A description of alternative methods available for the elicitation of preference values has been provided in **Chapter 2**, which included ranking, VAS, SG, TTO, and DCE. In this last empirical study (Study 4), a methodological study was undertaken in order to explore alternative applications of DCEs to obtain a scoring algorithm or population value set for the states described by the WAItE. Choice models were applied to the dataset to derive preference weights for all feasible states from the descriptive system. Briefly, DCEs involve the generation and analysis of choice, with a key underlying assumption that individuals adopt compensatory decision making. Meaning that the consideration of all the attributes included in a choice task (made up of multiple alternatives) is undertaken by individuals, and based on the levels of all attributes, and the preferred option is chosen (Ryan and Gerard, 2003). Having developed a new instrument that is suitable for the elicitation of preference values, this chapter provides a report of the methodological investigation to assign preference values to weight specific QoL states described by the WAItE. A brief overview of the intuition behind DCEs, and how it has been utilised in the context of health state valuation, is given in the next section.

8.1.1 An overview of the DCE technique

DCEs draw upon Lancaster's economic theory of value (Lancaster, 1966) and Random Utility Theory (RUT) (McFadden, 1973, *p.* 105-42), and involve three inter-related components: a) an experimental design used to implement a choice survey and generate choice data, b) discrete choice analysis that is used to estimate preferences from choice data and, c) the use of the resulting model to either derive welfare measures or construct other policy analyses (Lancsar & Louvier, 2008). RUT also referred to as the Random Utility Model (RUM), is derived from the economic assumption that a rational individual will select, from a set of alternatives, the option that provides them with the greatest expected utility. As the utility of the preferred choice increases, relative to the other alternatives available, there is an increased likelihood of the preferred option being chosen more often (if multiple choices are performed). Relative frequencies of choice can be used to produce numerical estimates of the utilities associated

with the preferred choice (and by implication the preferred health state from a given choice set). The first step in formulating the econometric model is to specify an indirect utility function that relates the level of utility enjoyed to the characteristics of attributes and levels in a given choice task. The decision making process within a DCE can be seen as a comparison of indirect utility functions. The indirect utility function used in analysis is merely an approximation of the individual's actual indirect utility function. It is assumed that a respondent makes a series of choices and for each choice an alternative that leads to the higher level of utility is picked by the respondent. An econometric representation of the decision is illustrated below (as per Ryan and Gerard (2003 p. 55)).

$$U_{qi}\,(A) = v_{qi}\,(A) + \varepsilon_{qi} \qquad (1)$$

Here $U_{qi}\ (A)$ = the indirect utility function of individual q for alternative i with attributes A. $v_{qi}\,(A)$ = the measurable component of utility estimated empirically. This simplified approximation of the indirect utility function also includes an error term to reflect the difference between actual observed and predicted behaviour. So, to complete the econometric model, a random element, ε_{qi} is added, and this reflects the unobservable factors affecting the estimation of the indirect utility function. This error component picks up the difference between the true utility, derived from choosing a particular option from a pre-defined set of alternatives, and that which is modelled (their predicted behaviour). The inclusion of this random element explains the derivation of the title RUM. The respondent will choose alternative i over alternative j if and only if:

$$\left(v_{qi} + \varepsilon_{qi}\right) > (v_{qj} + \varepsilon_{qj}) \qquad (2)$$

This can be presented as follows:

$$\left(v_{qi} - v_{qj}\right) > (\varepsilon_{qj} - \varepsilon_{qi}) \qquad (3)$$

Given the error term within the utility function, the analysis becomes a probabilistic choice analysis. The probability that any particular respondent prefers a particular option to any other alternative available to them can be expressed as the probability that the utility associated with that option, according to the model, exceeds the utility associated with all other options available to them. Given that the error terms are unknown, a probability model is estimated where:

$$P(i_q|A, C) = Pi_q = \left[(v_{qi} - v_{qj}) > (\varepsilon_{qj} - \varepsilon_{qi})\right] \quad (4)$$

The probability of choosing alternative i (over j) by individual q, given the set of attributes A and the choice set C, is given by the probability that the error difference is smaller than the difference of the *observable* utility component between i and j (Ryan and Gerard, 2003).

DCEs are regularly used in health economics to elicit preferences over healthcare products, programmes and in the valuation of preference for health states (Ryan *et al*., 2006, Viney *et al*., 2007, Coast *et al*., 2008 and Flynn *et al*., 2008). Several reviews have been conducted to compile the evidence of the use of DCEs in the health care setting (including Ryan *et al*., 2001, Ryan & Gerard 2003, and de Bekker-Grob *et al*., 2012). Studies have only recently endeavoured to anchor utility values, on the full health – dead scale, for the EQ-5D using the DCE_{TTO} technique, where QALY calculations can be made (Bansback *et al*., 2012 and Norman *et al*., 2012). This was done by incorporating an additional 'years of survival', namely a duration attribute, to the design of a DCE and then by creating a multiplicative design that allows for interactions between each of the QoL attribute levels and the life years attribute (which is treated as a continuous variable). The current study applied this novel DCE technique in the estimation of utility values for the health states described by the WAItE.

8.1.2 Hypotheses

In the current methodological study both the novel DCE_{TTO} valuation technique proposed by Bansback *et al*. (2012), including a duration attribute, and the DCE valuation with no duration ($DCE_{NoDuration}$) were assessed. Empirical studies utilising the TTO technique have observed *zero traders*, respondents refuse to trade any lifetime in exchange for improvements in health (Arnesen & Norheim, 2003). As DCE_{TTO} has never been tested utilising a mild descriptive system such as the WAItE, there is a risk that people might not be willing to trade duration in exchange for better QoL, therefore $DCE_{NoDuration}$ was also implemented.

In order to infer cardinal valuations (for states described by the WAItE) from the $DCE_{NoDuration}$ model, in the absence of a duration attribute, it was necessary to anchor the estimated valuations from the latent utility scale to the health utility scale using two strategies based on additional data external to the DCE exercise. Full descriptions of the DCE_{TTO} and $DCE_{NoDuration}$ methods are provided later in the chapter.

Assessments of the sign and significance of the parameter coefficients generated from the alternative models, through the application of DCE_{TTO} and $DCE_{NoDuration}$, were undertaken in order to identify the optimal valuation method and to test the following:

DCE_{TTO} & $DCE_{NoDuration}$	Can respondents distinguish between different levels of the WAItE dimensions?
	Do respondents self report understanding the DCE_{TTO} and $DCE_{NoDuration}$ tasks?
DCE_{TTO}	Do longer duration options have an impact on DCE_{TTO} coefficient estimates?
	Does respondent age have an impact on DCE_{TTO} coefficient estimates?

The empirical assessments of the aforementioned issues were intended to provide evidence to support or reject the use of DCE_{TTO} or $DCE_{NoDuration}$ in the valuation of the WAItE.

8.2 Methods

The methods employed in undertaking the DCE valuation study was informed by the DCE user's guide developed by Lancsar & Louvier (2008) and the Ryan & Gerard (2003) review of the applications of DCEs in health care. The methods can be broken down into four key steps: Step 1 – Identification of attributes and levels, Step 2 – Experimental design, Step 3 – Data collection and Step 4 – Data analysis and interpretation. The data analysis methods specifically employed in the evaluation of DCE_{TTO} was informed by Bansback *et al*. (2012). Detailed explanations of each of the steps are given below. The study utilised data that were generated from 3 independent samples, providing 3 separate datasets (Pilots A, B and C), further details are given in the section below under Step 3 – Data collection. The current study aimed to pilot alternative applications of DCE in the valuation of WAItE health states, hence should be viewed as the pre-testing undertaken in preparation for a full valuation study. Ethical approval for the study was provided by the University of Leeds local research ethics committee (Ref - HSLTLM/11/049).

8.2.1 Step 1 - Identification of attributes and levels

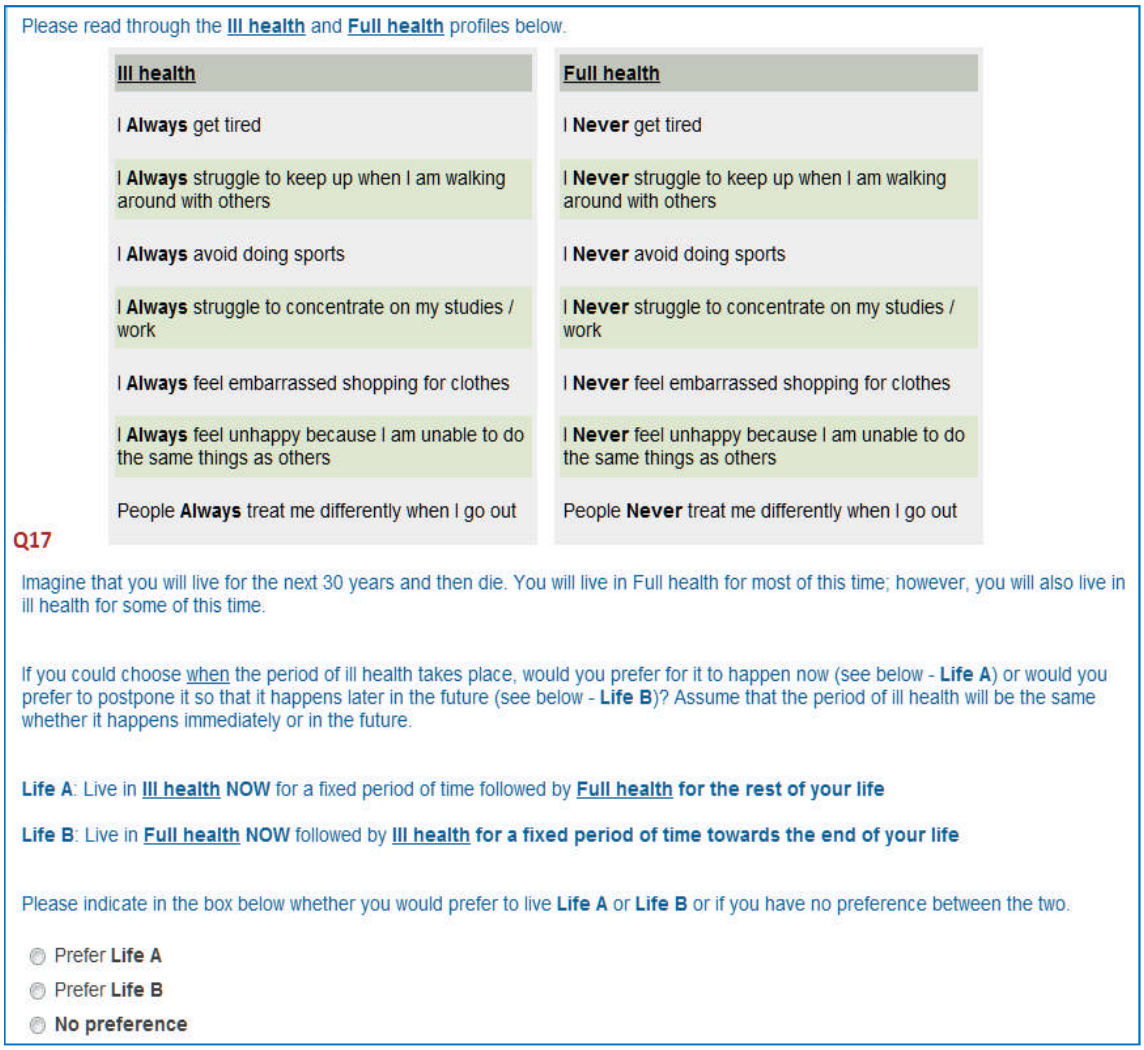
The first step of the valuation study required the identification of attributes and levels for the DCE task. Each of the DCE profiles were made up of a health state defined by the WAItE, containing seven attributes (henceforth referred to as the QoL attributes), with five possible levels for each attribute. In this case level one referred to the best level in each attribute *never* – so health state 1111111 is referred to full health – and 5555555 is referred to the worst health state possible in this descriptive system. For the DCE_{TTO}, in addition to the seven QoL attributes, an 8th attribute describing the number of years an individual would live in a particular state followed by immediate death, the *duration* attribute, was added making up a full eight attribute DCE_{TTO} profile. The duration attribute was also made up of five levels of life years. Three different duration versions were used: baseline: 10, 9, 8, 6 & 5 years (duration values similar to those used by Bansback *et al*., 2012), and then these were scaled up by a factor of two (i.e. 20, 18, 16, 12, and 10 years) and also by a factor of five (i.e. 50, 45, 40, 30 and 25 years) to assess the impact of two sets of longer duration options on coefficient estimates (**Table 8.1**).

An important issue that needs to be considered, which has been raised in the literature, is *time preference*: the relative value placed on an ill health state experienced at an earlier date compared with the value that would be placed on the same health state if it was experienced at a later date (Dolan & Gudex, 1995). As different duration levels were utilised, and may be influenced by time preference rates, a time preference question was also added to the survey (**Figure 8.1**). Respondents were asked whether they would prefer to have ill health now followed by good health in the future (negative time preference) or good health now followed by ill health in the future (positive time preference), or whether they had no preference (zero or neutral time preference). The worst WAItE health state (5555555) was given as the ill health state, and the duration of ill health was fixed over the two time periods.

Table 8:1 Attributes and levels used to design the DCE study

Summary of attributes			Levels				
			Never	Almost never	Some-times	Often	Always
QoL	1. Tired		1	2	3	4	5
	2. Walking with others		1	2	3	4	5
	3. Avoid sports		1	2	3	4	5
	4. Concentrate on studies / work		1	2	3	4	5
	5. Embarrassed clothes shopping		1	2	3	4	5
	6. Unhappy as unable to do the same things as others		1	2	3	4	5
	7. Treated differently		1	2	3	4	5
Duration	8. Life expectancy (three different variants)	D10	10yrs	9yrs	8yrs	6yrs	5yrs
		D20	20 yrs	18 yrs	16 yrs	12 yrs	10 yrs
		D50	50 yrs	45 yrs	40 yrs	30 yrs	25 yrs

Figure 8:1 Example of time preference question



8.2.2 Step 2 - Experimental design

For the DCE_{TTO} survey a factorial design (experimental designs in which every level of every attribute is paired with every level of every other attribute under consideration) (Lancsar & Louvier, 2008) made up of the seven QoL attributes (WAItE items), the duration attribute and associated levels for each attribute, summarised in **Table 8.1**, resulted in 390625 (5^8) possible profiles. For $DCE_{NoDuration}$ the total number of possible scenarios, made up of only the seven WAItE items, equates 78125 (5^7) QoL profiles. A full factorial design would result in some scenarios where a dominant alternative was present (i.e. where all attribute levels are better in profile A compared with profile B, for example, in a given pairwise choice set), and was thus factored into the design, so that dominant pairwise choice sets would be avoided. For the WAItE, there are no implausible scenarios (where the worst level in one dimension is incompatible with the best level in another) and so this did not need to be factored into the design. An efficient factorial experimental design was constructed. The designs for the DCE_{TTO} and for the $DCE_{NoDuration}$ were generated using the Ngene design software (ChoiceMetrics, 2012) and was guided by advisors with expertise in the area of the design and analysis of DCEs. The programme was asked to generate a design consisting of nine blocks each containing ten pairwise choice tasks for both the DCE_{TTO} and for the $DCE_{NoDuration}$. The DCE_{TTO} design accounted for interactions between the WAItE dimension levels and the duration attribute, where the latter was treated as a continuous variable. A row-swapping process was used based on constraints to avoid dominance. For each design, the Ngene programme was left to run for tens of thousands of iterations and the one that displayed the lowest D-error (a measure of the (in)efficiency of the experimental design) was chosen with the aim of minimising the standard errors (the reliability of the model parameters to be estimated could be quantified in terms of the asymptotic standard errors and covariances; thus improvements in reliability suggested a reduction in the asymptotic standard errors (Bliemer *et al*., 2008)). Based on 90 pairwise choice sets, each containing two profiles, the software generated the optimal sub-sample of choice sets to be included in the DCE survey. Each of the nine blocks of profiles was randomised in the survey so that an equal number of respondents were given each of the nine blocks, for both the DCE_{TTO} and $DCE_{NoDuration}$.

The pairwise alternatives for Pilots A and B were identical with the only difference being the duration options presented to individuals, as the levels presented in the former were shorter than those presented in the latter. Pilot C contained $DCE_{NoDuration}$, where there was no duration attribute, but instead two additional anchoring tasks that were external to the DCE tasks. The pairwise alternatives for Pilot C was different to the ones used in Pilots A and B as the design did not include a duration attribute. The two anchoring tasks utilised in Pilot C are briefly described below with further details given in **Section 8.2.4**. One of the anchoring tasks was a Visual Analogue Scale (VAS), where respondents were asked to value the WAItE PITS state (5555555) (**Figure 8.2**). The upper end point was described as the best imaginable health state and was set at a value of 100. The lower endpoint of the scale was set at a value of -50. Being dead was marked at zero. The upper anchor was informed by the VAS scale used for the EQ-5D and the lower anchor of -50 was informed by the lowest score for the worst EQ-5D health state (EuroQol Group, 1990). The second anchoring task was a binary choice version of the TTO between the PITS state for a fixed duration and full health (1111111) for a comparatively shorter duration using pre-set combinations of durations between the two states (as per **Table 8.2**) henceforth referred to as the *pairwise anchoring task*. The anchoring options presented to respondents were age dependent to make them realistic, and each individual could complete a maximum of two of these anchoring tasks. The shorter duration option in full health was first presented to respondents, and if they chose this over living in the PITS state they did not complete another anchoring task. On the other hand, if the PITS alternative was chosen, then a second anchoring task with a longer duration in full health was presented to respondents.

Figure 8:2 Example of VAS question (Anchoring 1: $DCE_{NoDuration}$)

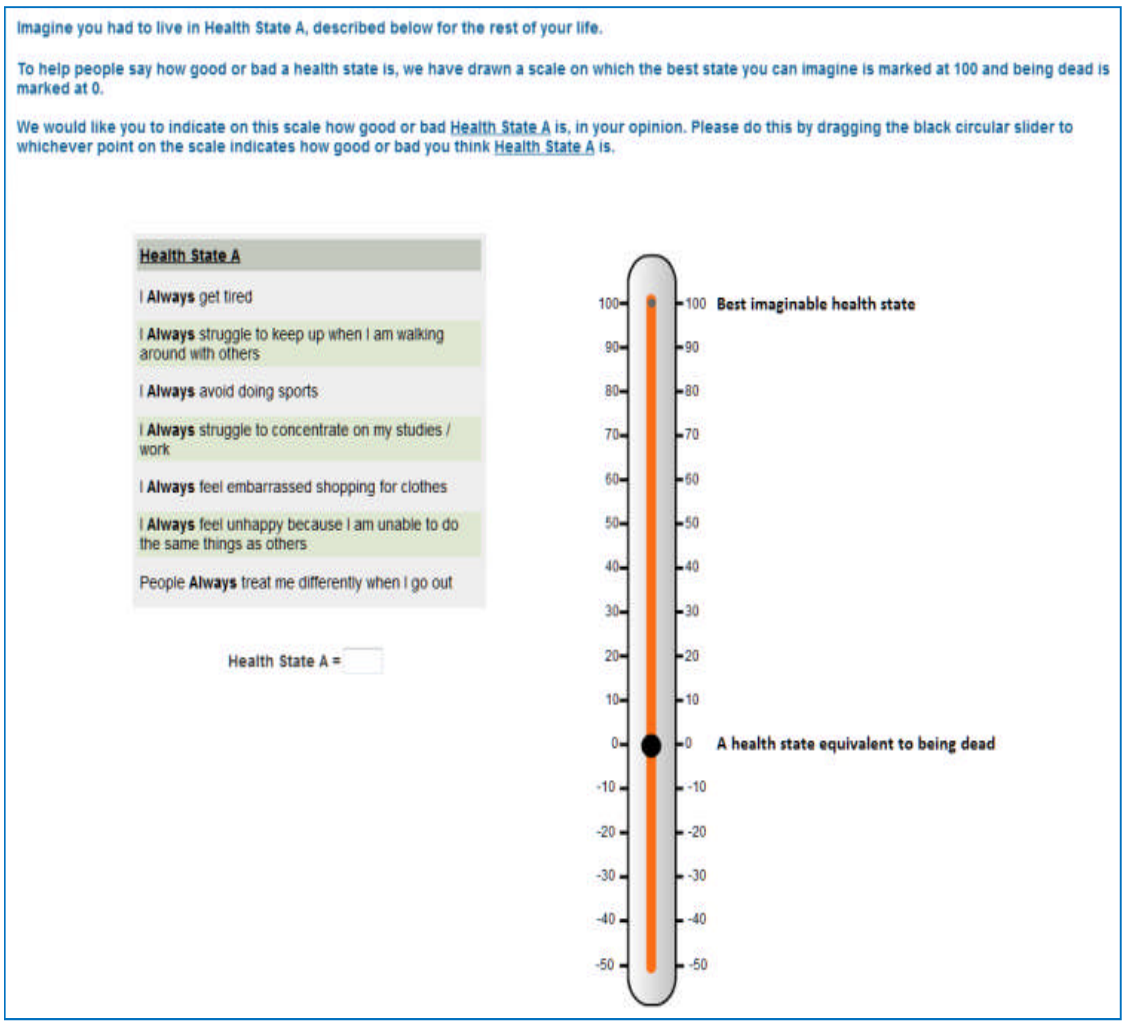


Table 8:2 WAItE PITS vs. full health states – (Anchoring 2: $DCE_{NoDuration}$)

Anchor 2 options: For 18 to 54 year olds		**Utility values**
WAItE PITS (5555555)	**WAItE Full health (1111111)**	
50 (Fixed)	10	0.20
50 (Fixed)	15	0.30
50 (Fixed)	37	0.74
50 (Fixed)	42	0.84
50 (Fixed)	45	0.90
50 (Fixed)	47	0.94
Anchor 2 options: For 55 Plus and 35– 54 year olds		
20(Fixed)	4	0.20
20(Fixed)	6	0.30
20(Fixed)	15	0.75
20(Fixed)	17	0.85
20(Fixed)	18	0.90
20(Fixed)	19	0.95

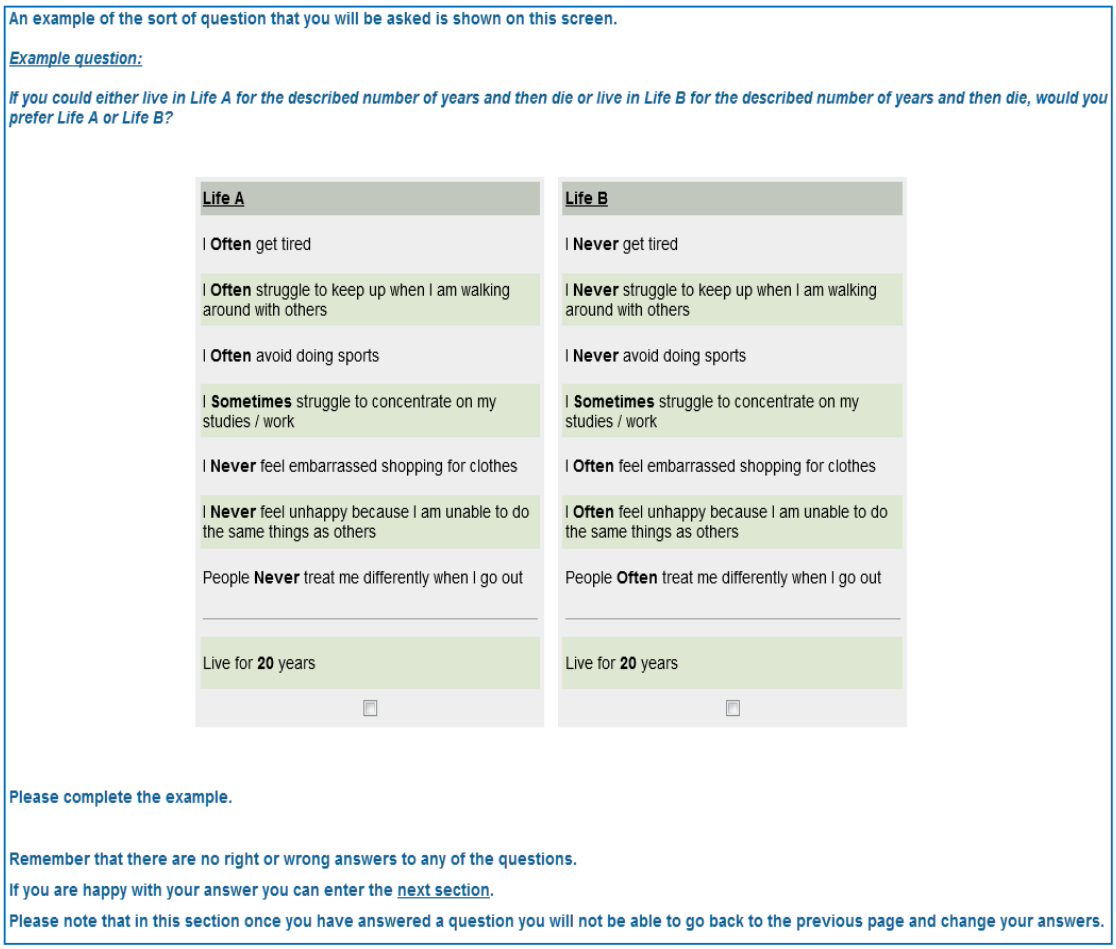
8.2.3 Step 3 – Data collection

A market research company, with an existing internet panel, managed and hosted the website were the web survey was held. Participants completing the survey were given a nominal payment of £1.75, by the survey company, if they completed the survey fully. The company recruited a sample of the UK population over 18 years from the panel based on quotas in terms of gender and age in order to obtain a balanced sample of respondents. A randomisation coding was applied so that there was an equal response to each of the profile blocks made up of 10 DCE tasks (as per the design explained in the previous step). Sample size calculations were not undertaken as the information generated from the study is to inform the main valuation study to be undertaken in future research.

On entering the survey, respondents were told: '*You are invited to take part in a research study in the area of weight status and how it can affect different aspects of life...We would like to find out which of the different aspects of life that can be affected by weight status, are most important to you'*, so that they knew the subject matter of the survey. After obtaining consent, the ordering of questions within the survey, for Pilots A and B, was as follows: 1) background characteristics, 2) socioeconomic status, 3) self-assessed health, 4) self report weight and height (so that BMI could be calculated), 5) the 7 item WAItE instrument and 6) the 10 DCE pairwise tasks each made up of two DCE profiles. For Pilot C ($DCE_{NoDuration}$ was implemented) one of the pairwise anchoring tasks was presented before the 10 $DCE_{NoDuration}$ tasks and the second anchoring task was presented on completion of the 10 $DCE_{NoDuration}$ tasks. After this the VAS anchoring question was presented. For both Pilots B and C a final section presented the time preference question. An introductory page providing information about the DCE valuation survey was presented to respondents before they were asked to complete the DCE tasks. Within the introductory section, before completing the main DCE survey, respondents completed a practice question. For all the DCE tasks, respondents were forced to make a choice between the two profiles presented to them, as an opt-out was not available. An example of the choice set included in the survey is given in **Figure 8.3**.

Conventionally, data from respondents that are observed to have failed to engage or understand the DCE task, as set by established criteria (Lancsar & Louviere, 2006), are excluded from the analysis. The time taken for each respondent to complete each section of the survey was measured and thus a speeder check was implemented, whereby participants were screened out of the survey if they completed the DCE section within a pre set minimum time limit. The minimum set limit was based on the assessment of how long it took the majority of participants (90%) from the first pilot study to complete the DCE section of the survey. Self assessed understanding of the DCE tasks was also measured, by asking respondents *how much* they understood the DCE questions, on completion of this section of the survey. For those respondents responding that they *did not understand*, it was assumed that these individuals did not understand what was being asked of them in the DCE tasks, and they were therefore excluded from the DCE analysis.

Figure 8:3 Example of DCE question (*with* duration)



Details of the three surveys which were completed by three independent groups of respondents can be summarised as follows:

Pilot A – the first participant sample was given the DCE_{TTO} tasks with duration options of 10, 9, 8, 6 and 5 years (D10), similar to those used in the Bansback *et al*. (2012) study (see **Table 8.1**).

Pilot B – the second sample completed the DCE_{TTO} tasks with the baseline duration options scaled up by a factor of two (D20) and a factor of five (D50) (see **Table 8.1**). Pilot B was split into three: Sample 1 consisted of 18-54 year olds who were given the duration options that were scaled up by five. Sample 2 consisted of participants aged 55 years or over given the duration options scaled up by two. Sample 3 consisted of 35-54 year olds with duration scaled up by two. Three samples were used for Pilot B in order to test the impact of different duration level options on different age groups.

Pilot C – the final participant sample was given the $DCE_{NoDuration}$ tasks, which did not have a duration attribute, with the two additional anchoring exercises.

8.2.4 Step 4 - Data analysis and interpretation

For DCE_{TTO}, this final step was informed by the Bansback *et al*. (2012) study, which allows the utility values estimated for WAItE states to be anchored on the full health - dead scale. The two part analysis undertaken for both DCE_{TTO} and $DCE_{NoDuration}$ are reported below.

8.2.4.1 Stage 1: Estimating the model coefficients

DCE_{TTO}

Stage 1 of the Bansback *et al*. (2012) methodology makes use of choice models to estimate a utility function based on the characteristics of the profiles contained in each pairwise choice task (i.e. modelling the choices as a function of the WAItE and duration level attributes). The multinominal logit (MNL) model, also known as the conditional logit model, has already been used to estimate the latent utility function (for example see Hakim & Pathak, 1999).

The MNL model aims to find the values of parameters such that the chosen alternatives always have the highest utility, thus maximizing the log-likelihood of the actual observed choices. Hakim & Pathak (1999, *p.* 106) summarise the MNL model as follows:

$$\mathrm{P\ (j|\ W)} = \ \mathrm{e^{V(j)}} | \sum \mathrm{e^{v(n)}} \qquad (5)$$

For all n in W where P (j| W) = probability of selecting a particular alternative j given choice set W. V(j) = systematic, observable component or mean utility value of choosing alternative z given the mean utility value that would have been derived from choosing another alternative from the other available alternatives n in choice set W. e = natural constant, the base of the natural logarithms = 2.7183.

To assess the sign and the magnitude of the DCE_{TTO} coefficients (significance was assessed at the 10% level in order to take into account the small sample size of each of the pilot studies), the analysis described in Bansback *et al*. (2012) is used to estimate the interaction effects between the WAItE dimension levels and the duration attribute. The utility function μ of each respondent q is defined as multiplicative function of interactions between the vector of levels for each WAItE attribute x and duration in each scenario so that:

$$\mu_{qj} = \propto + \beta t_{qj} + \gamma' \mathbf{X}_{qj}.t_{qj} + \varepsilon_{qj} \qquad (6)$$

α = the constant term. βt_{qj} = the duration attribute effects, the value of living in full health for the specified duration. $\gamma' \mathbf{X}_{qj}.t_{qj}$ = the interaction term between a give health state and the duration attributes. Here t is treated as a continuous variable for the majority of the models. The model estimated 28 dummy explanatory variables for each level of each dimension of the WAItE interacted with the duration attribute, as it provides the weight associated with the time t_j lived in each health state. Level one acts as a baseline for each WAItE dimension. Because of this, there are four interaction coefficient estimates for each of the seven WAItE attributes. Finally, ε_{qj} = the random term (assumed to be independent and identically distributed).

In order to enable the assessment of the sign and the magnitude of the DCE_{TTO} and $DCE_{NoDuration}$ coefficient estimates, in the main effects models, the duration attribute is excluded from the model specification. For DCE_{TTO} this meant that the dimension level dummies and duration were not interacted with each other. These models were estimated purely in order to assess model performance in terms of investigating whether the un-interacted parameter estimates behaved as we would expect and to enable comparison of the DCE_{TTO} and $DCE_{NoDuration}$ parameter estimates. Consequently, both the main effects and interaction models were estimated for DCE_{TTO} where different duration variants were used. For $DCE_{NoDuration}$ only the main effects model was estimated as there was no duration parameter (a description of the $DCE_{NoDuration}$ model estimation is provided below).

$DCE_{NoDuration}$

The functional form of μ is what differs between the estimation of the $DCE_{NoDuration}$ and DCE_{TTO} models. A simplified version of equation 6 is used in the estimation of the latent utility function for $DCE_{NoDuration}$ (see equation 7). As there is no duration attribute included, only main effects parameters for the 28 WAItE level dummy variables are estimated, so that:

$$\mu_{qj} = \propto + \lambda' \mathbf{X}_{qj} + \varepsilon_{qj} \quad (7)$$

8.2.4.2 Stage 2: Calculation of utility values

DCE_{TTO}

The application of Stage 2 of the Bansback *et al*. (2012) approach, to DCE_{TTO}, involves the modelling/derivation of the health state values using the DCE_{TTO} coefficients obtained from Stage 1. Bansback *et al*. (2012) show how the value for each health state can be anchored on the full health - dead utility scale using an anchoring utility function as shown below in equation 8. Here health state values are derived based on the marginal rate of substitution. The assumption is that for each profile made up from living in a given state x_j for the maximum amount of years, 10, 20 or 50 years (depending on the duration level options utilised), there is a comparatively shorter number of years of living in full health (t < 10, 20 or 50) which generates the same level of utility. So the calculation of the utility value of a particular health state is based on the life expectancy that the *sample* is willing to sacrifice: the point at which the mean utility value of living in health state x_j for 10, 20 or 50 years is equivalent to the mean utility value of living in full health (1111111) for a comparatively shorter period of time.

$$V_j = 1 + \frac{\hat{\gamma}'}{\hat{\beta}} \mathbf{X}_j \quad (8)$$

Bansback *et al*. (2012) explain that the dysfunctional health states can be anchored on the health state utility scale (V). When $\hat{\gamma}$ = 0, by implication V = 1 representing the value of full health (WAItE state 1111111). Deviations from full health, the dysfunctional WAItE states, are associated with negative coefficient estimates thus the expression $\frac{\hat{\gamma}}{\hat{\beta}}$ should be negative. If the absolute value of $\hat{\gamma}$ = $\hat{\beta}$, then $\frac{\hat{\gamma}}{\hat{\beta}}$ = 1 and V = 0. If the WAItE state is very severe, then the absolute value of $\hat{\gamma}$ will be greater than the value of $\hat{\beta}$ ($\hat{\gamma}$ > $\hat{\beta}$), meaning that the disutility associated with the dysfunctional state may be greater than the difference between full health and being dead. This would result in a negative V which suggests a state valued worse than being dead.

Bansback *et al*. (2012) state that it is important to note that the anchoring of the utility function to dead at 0 is achieved from equation 6 through the relative size of the regression coefficients $\hat{\gamma}$ and $\hat{\beta}$. Whilst anchoring the utility function to one as the value for full health is obtained by the value given to the first argument in equation 8, as $\hat{\gamma}$ = 0 for the full health state.

As Stage 2 is contingent on the coefficients from Stage 1 having the correct sign and consistent ordering, the primary focus of the analysis in this study is on the coefficient estimates derived from the models in Stage 1. For illustrative purposes, using the estimated coefficients from DCE_{TTO}, the health state value for selected states were also calculated and presented in the results section. The Bansback *et al*. (2012) approach was specifically designed for the inclusion of a duration attribute within a DCE task. Where there is no duration attribute, equations 6 cannot be derived, this methodology cannot be applied to $DCE_{NoDuration}$ thus an alternative technique was utilised for the calculation of utility values.

$DCE_{NoDuration}$

In order to infer cardinal values from the $DCE_{NoDuration}$ model (equation 7) it was necessary to transform the derived estimates onto the health utility scale anchored on full health and dead (Ratcliffe *et al*., 2009). One method of re-scaling, that has been used previously, would be to assume that the value for the best WAItE state is equal to one, and that the bottom anchor, the value for the WAItE PITS state, is equal to zero - the same as being dead (Ryan *et al*., 2006). This method has its limitations as it suggests that different descriptive systems assign the same utility value (zero) to the PITS state (the worst state). In addition to this, evidence shows that individual's assign different values to the EQ-5D PITS and the state of immediate death (Gudex *et al*., 1996). Applying this method in the current context would mean that the value of the WAItE PITS state (see health state A in **Figure 8.2**) is assumed to be equal to the value for PITS state described by, for example, the EQ-5D (being confined to bed, unable to wash or dress, unable to perform usual activities, in extreme pain or discomfort and extremely anxious or depressed). The WAItE PITS state is comparatively milder than the worst EQ-5D state therefore it is not convincing that the two states should be assigned the same utility value. Because of this, a different approach was taken. VAS and pairwise anchoring, external to the DCE exercise, were used to assign utility values to the WAItE PITS state. These two anchoring tasks generated two alternative median sample values for the WAItE PITS state. These values were used as the lower anchor for transforming the estimated utility values for WAItE states from the latent scale, generated from the $DCE_{NoDuration}$ model coefficients, to the health utility scale.

In the VAS task individuals were asked to directly value the PITS state, given fixed values for full health (100) and being dead (0). By dividing the raw VAS score by 100, the value for the WAItE PITS state was anchored on the health utility scale. Using i) the median sample value generated by aggregating each of the individual valuations of the PITS state and ii) the $DCE_{NoDuration}$ model coefficients, the remaining WAItE health states could be re-scaled relative to the PITS state, and by implication generating re-scaled values anchored on the full health - dead scale (Ratcliffe *et al*., 2009). To calculate the re-scaled value of a given WAItE state, say 2222222 ($V_{2222222}$), it is assumed that the relative position of this state between full health ($V_{1111111}$) and PITS ($V_{5555555}$) on the VAS scale (divided by 100) is equal to its relative position between full health ($\hat{\beta}X_{1111111}$) and PITS ($\hat{\beta}X_{5555555}$) on the $DCE_{NoDuration}$ latent scale. This is can be formally written as:

$$\frac{V_{1111111} - V_{2222222}}{V_{1111111} - V_{5555555}} = \frac{\hat{\beta}X_{1111111} - \hat{\beta}X_{222222}}{\hat{\beta}X_{1111111} - \hat{\beta}X_{555555}} \quad (9)$$

When the relevant values for full health ($V_{1111111}$ = 1 and $\hat{\beta}X_{1111111}$ = 0) are imposed, this equation can be expressed as:

$$\frac{1 - V_{2222222}}{1 - V_{5555555}} = \frac{0 - \hat{\beta}X_{222222}}{0 - \hat{\beta}X_{555555}} \quad (10)$$

Equation 10 can then be re-arranged so that it is solved for $V_{2222222}$ to enable the calculation of the re-scaled value for state 2222222 on the VAS scale (which was anchored on the health utility scale):

$$1 - \frac{(1 - Q_{5555555}) * (-\hat{\beta}X_{222222})}{-\hat{\beta}X_{555555}} = V_{2222222} \quad (11)^{12}$$

Here $Q_{5555555}$ = the sample median value for the WAItE PITS state from the VAS. Equation 11 can also be used to estimate the re-anchored values for WAItE states using the pairwise anchoring task.

As the pairwise anchoring task involved a trade-off between living in the WAItE PITS state for a fixed period of time compared with living in full health for a comparatively shorter period of time, the utility value of the PITS state was calculated by dividing the shorter time period by the fixed longer alternative. The sample value for the PITS state was inferred by identifying the sample median utility value for the PITS state. The median sample value for the WAItE PITS state was assumed to be a close approximation to the true sample value (if a full TTO task was undertaken, the point at which each individual was indifferent between the two states would be obtained and aggregated to generate a sample value for the PITS state). It would not have been meaningful to obtain the mean utility value for the sample, using pairwise anchoring method, as this would have been based on the pre-specified utility values presented to respondents. As with the VAS anchoring, the lower anchor was assumed to be the worst health state and full health was assumed to be equal to one. In order to have a meaningful comparison between the two anchoring tasks the median sample VAS and pairwise anchoring values were used to estimate the value for the WAItE PITS state.

[12] The generalised form of this equation would be as follows:

$$1 - \frac{(1 - Q_{5555555}) * (-\hat{\beta}X_{\mathrm{j}})}{-\hat{\beta}X_{555555}} = V_{\mathrm{j}}$$

Comparing DCE_{TTO} and $DCE_{NoDuration}$ results

To be able to directly compare DCE_{TTO} and $DCE_{NoDuration}$, an assessment of the main effects models for the WAItE dimension level parameter estimates was necessary. For DCE_{TTO}, this was done by estimating only the WAItE attribute level parameters with no duration parameter. Choice sets were included in the model where the duration levels were the same between the two profiles making up a pairwise choice set. The duration attribute was not included in the $DCE_{NoDuration}$ tasks and thus there was no duration parameter to be estimated using this method. There are limitations associated with making this comparison, as this now means that the $DCE_{NoDuration}$ model is based on an optimal design, whilst the set of observations that have gone into the DCE_{TTO} model is not, and thus the DCE_{TTO} design will be comparatively less efficient.

8.3 Results

8.3.1 The Sample

Between the three pilot studies, a sample of 757 members of the market research panel started to complete the survey. Of these 93 (12%) were screened out because they completed the DCE section of the survey within the minimum time limit of 75 seconds (this was based on data from Pilot A - 90% of participants completed the DCE section in 75 seconds or more). One hundred and fifty five (20%) individuals dropped-out whilst answering the non DCE sections of the survey, whereas 17 (2%) dropped out whilst answering the DCE questions. Overall data from a total of 489 (65%) of the participants who initially started the survey were available for inclusion in the DCE analysis (and received the £1.75 payment). This was made up of: 112 individuals in Pilot A, 206 individuals in Pilot B and 171 individuals in Pilot C (**Table 8.3**). The key characteristics of the respondents are shown in **Table 8.4** (additional information is presented in **Appendix 8.1**). Most respondents were in excellent to fair health 441 (90%). Over half of participants were either overweight or obese (having a BMI of 25 and over) 275 (56%). In terms of self assessed understanding of the DCE tasks only seven (1% of the total number of participants) felt that they did not understand what they were being asked to do. These individuals were excluded from the DCE analysis. In the speeder check, minimum amount of time that was set for the completion of the DCE section of the survey was set at 75 seconds as this was how long it took 90% of the participants from Pilot A to complete the DCE section of the survey (see **Appendix 8.2**).

Table 8:3 Survey completions across the three pilot studies

Respondents	Pilot A		Pilot B		Pilot C	
	N	%	N	%	N	%
Eligible starts	128	100.00	277	100.00	352	100.00
Screen-outs (completed the survey within the minimum time limit)	8	6.25	33	11.91	52	14.77
Dropouts (Outside of DCE exercise)	5	3.91	31	11.19	119	33.81
Dropouts (During DCE exercise)	3	2.34	4	1.44	10	2.84
Total number who completed the full survey but failed to formally sign out from the survey and to be included	0	0.00	3	1.08	0	0.00
Total number excluded from the final sample	16	12.50	71	25.63	181*	51.42
Final sample available for analysis	112	87.50	206	74.37	171	48.58

*There were a high number of individuals in Pilot C, relative to the other two studies, who either completed the DCE section of the survey in less than the minimum time or dropped out of the survey in the non DCE section. As this pilot study contained more questions than the other two pilot studies (including: the two additional anchoring questions, the VAS question as well as a time preference question) there was more of an opportunity for drop-outs.

Table 8:4 Characteristics of respondents in the surveys

Characteristic		Pilot A (N=112)		Pilot B (N=206)		Pilot C (N=171)	
		N	%	N	%	N	%
Block / version	1	15	13.39	19	9.22	17	9.94
	2	16	14.29	19	9.22	18	10.53
	3	12	10.71	29	14.08	19	11.11
	4	10	8.93	30	14.56	20	11.70
	5	12	10.71	20	9.71	19	11.11
	6	13	11.61	27	13.11	19	11.11
	7	11	9.82	19	9.22	22	12.87
	8	13	11.61	24	11.65	18	10.53
	9	10	8.93	19	9.22	19	11.11
Gender	Male	57	50.89	102	49.51	85	49.71
	Female	55	49.11	104	50.49	86	50.29
Age group	18-24	16	14.29	24	11.65	26	15.20
	25-34	35	31.25	38	18.45	31	18.13
	35-44	14	12.50	48	23.30	41	23.98
	45-54	23	20.54	55	26.70	40	23.39
	55-64	16	14.29	29	14.08	20	11.70
	65-90	8	7.14	12	5.83	13	7.60
Weight status	1=Under (BMI 18.5 or less)	6	5.36	9	4.37	6	3.51
	2=Normal (BMI 18.6 to 24.9	41	36.61	83	40.29	69	40.35
	3=Overweight (BMI 25 to 24.9)	41	36.61	67	32.52	50	29.24
	4=Obese (BMI 30 plus)	24	21.43	47	22.82	46	26.90
Difficulty with DCE task	Very difficult	12	10.71	25	12.14	23	13.45
	Quite difficult	43	38.39	98	47.57	95	55.56
	Neither difficult or easy	31	27.68	58	28.16	37	21.64
	Fairly easy	26	23.21	25	12.14	16	9.36
Understan-ding DCE task	Fully understand	84	75.00	153	74.27	116	67.84
	Partially understand	26	23.21	49	23.79	54	31.58
	Did not understand	2	1.79	4	1.94	1	0.58
WAItE	Mean (SD)	17.24 (5.2)		16.89 (5.2)		17.45 (5.6)	

8.3.2 Stage 1 - Estimating the model coefficients

8.3.2.1 DCE_{TTO} long vs. short durations – Main effects models

Table 8.5 presents the results for the main effects models. The results for Pilot B, where different duration levels were utilised, are presented independently for Samples 1 to 3 and also where the three samples are combined. For the latter, the duration attribute was estimated as a categorical variable. In the majority of the models, the estimated WAItE dimension level dummy parameter coefficients did not have the expected sign (the expectation was that all QoL attribute coefficients would be negative and increasing whilst the duration attribute would be positive) and are not statistically significant at the 10% level. For Pilot A, only 4 (14%) out of the 28 WAItE dimension level dummy variables had the expected negative sign. Pilot B seemed to generate the better performing models, having a higher number of estimated WAItE dimension level parameters with a negative sign. The best model used data from participants in Sample 3 and was composed of 35-54 year olds who were presented with the duration levels scaled up by two (i.e. 20, 18, 16, 12 or 10 years). Here 26 (93%) of the estimated WAItE dimension level dummy coefficients had the expected negative sign and the majority of the parameters were statistically significant at the 10% level. A significant coefficient indicates that the relevant response option is distinguished from baseline, level one. The ordering of coefficients across levels within each of the dimensions also seemed to be as would be expected in most instances. The coefficient for the constant term was only significant in this model. A significant constant term suggests that there is specification error in the analyses, as this indicates that something is influencing the choice other than the attributes (Bansback *et al*., 2012), and thus the coefficient for the constant term would not be expected to be significant. The combined Pilot B data was the next best performing model, estimating negative coefficients for 16 (57%) of the WAItE level parameters (only one of these parameters was significant at the 10% level). The duration coefficients for both the models were positive and statistically significant, indicating respondents preferred more years of life, as would be expected.

Table 8:5 Parameter estimates for Main effects models*

Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B_ Samples 1 to 3	
	18 + yrs (D10)[a]	p-value	18-54yrs (D50)[a]	p-value	54+yrs (D20)[a]	p-value	35-54yrs (D20)[a]	p-value	18+yrs (D50) (D20)[a]	p-value
Tired_L2	0.232	(0.24)	0.244	(0.23)	0.644	(0.12)	0.042	(0.90)	**0.284**	**(0.06)**
Tired_L3	0.082	(0.79)	0.061	(0.84)	0.666	(0.35)	**-1.050**	**(0.05)**	-0.007	(0.98)
Tired_L4	0.146	(0.74)	-0.129	(0.76)	0.82	(0.36)	-0.911	(0.19)	-0.063	(0.83)
Tired_L5	-0.102	(0.88)	0.067	(0.91)	0.717	(0.62)	**-2.150**	**(0.05)**	-0.060	(0.90)
Walking_L2	0.321	(0.1)	0.300	(0.11)	-0.146	(0.66)	**-0.541**	**(0.03**	0.129	(0.34)
Walking_L3	0.355	(0.26)	0.294	(0.32)	0.663	(0.31)	**-1.000**	**(0.06)**	0.161	(0.46)
Walking_L4	0.433	(0.34)	0.257	(0.53)	0.097	(0.92)	**-1.580**	**(0.05)**	-0.035	(0.91)
Walking_L5	0.543	(0.40)	0.399	(0.50)	0.578	(0.70)	**-1.940**	**(0.07)**	0.169	(0.72)
Sports_L2	0.225	(0.23)	0.233	(0.20)	0.243	(0.50)	0.220	(0.45)	0.181	(0.18
Sports_L3	0.445	(0.18)	0.272	(0.35)	-0.147	(0.85)	-0.263	(0.60)	0.153	(0.50)
Sports_L4	0.384	(0.39)	0.281	(0.50)	0.709	(0.49)	-0.753	(0.28)	0.159	(0.61)
Sports_L5	0.331	(0.62)	0.385	(0.53)	0.739	(0.64)	-1.390	(0.20)	0.200	(0.67)
Concentrate_L2	0.201	(0.31)	0.029	(0.88)	0.099	(0.80)	**-0.991**	**(0.00)**	-0.130	(0.36)
Concentrate_L3	0.184	(0.56)	0.012	(0.97)	0.632	(0.29)	**-0.851**	**(0.10)**	-0.016	(0.94)
Concentrate_L4	0.35	(0.42)	0.095	(0.81)	-0.097	(0.91)	**-2.100**	**(0.00)**	-0.309	(0.30)
Concentrate_L5	0.215	(0.75)	0.029	(0.96)	0.417	(0.77)	**-2.660**	**(0.01)**	-0.307	(0.51)
Embarrased_L2	0.080	(0.71)	0.211	(0.26)	0.445	(0.44)	-0.485	(0.18)	0.114	(0.46)
Embarrased_L3	0.030	(0.93)	0.097	(0.72)	0.694	(0.39)	**-1.210**	**(0.02)**	0.051	(0.82)
Embarrased_L4	0.107	(0.83)	0.251	(0.53)	0.472	(0.67)	**-1.740**	**(0.02)**	-0.015	(0.96)
Embarrased_L5	-0.077	(0.91)	0.213	(0.73)	0.532	(0.75)	**-2.210**	**(0.04)**	-0.061	(0.90)
Unhappy_L2	0.419	(0.02)	**0.328**	**(0.09)**	-0.045	(0.91)	-0.077	(0.82)	0.208	(0.16)
Unhappy_L3	0.203	(0.51)	0.013	(0.96)	0.670	(0.40)	**-1.300**	**(0.01)**	-0.092	(0.67)
Unhappy_L4	0.044	(0.92)	-0.140	(0.73)	0.939	(0.37)	**-1.760**	**(0.02)**	-0.204	(0.52)
Unhappy_L5	-0.069	(0.92)	-0.106	(0.86)	0.516	(0.74)	**-2.920**	**(0.01)**	-0.335	(0.48)

TreatDif_L2	0.0989	(0.65)	-0.0129	(0.95)	0.467	(0.3)	**-0.847**	**(0.01)**	-0.0381	(0.80)
TreatDif_L3	0.148	(0.65)	0.194	(0.5)	1.14	(0.16)	**-1.17**	**(0.03)**	0.065	(0.78)
TreatDif_L4	0.0946	(0.83)	0.0816	(0.83)	0.475	(0.68)	**-1.6**	**(0.02)**	-0.122	(0.69)
TreatDif_L5	-0.0237	(0.97)	-0.0582	(0.92)	0.56	(0.72)	**-2.19**	**(0.05)**	-0.293	(0.53)
Duration	0.197	(0.26)	0.0342	(0.29)	0.116	(0.58)	**0.472**	**(0)**	-	-
Duration_L2	-	-	-	-	-	-	-	-	**0.449**	**(0.01)**
Duration_L3	-	-	-	-	-	-	-	-	**0.861**	**(0.03)**
Duration_L4	-	-	-	-	-	-	-	-	**1.08**	**(0.04)**
Duration_L5	-	-	-	-	-	-	-	-	**1.52**	**(0.01)**
Constant	-0.00658	(0.91)	-0.0421	(0.5)	-0.176	(0.2)	**0.253**	**(0.05)**	-0.0287	(0.56)

Main effects models	**Pilot A**	**Pilot B Sample 1**	**Pilot B Sample 2**	**Pilot B Sample 3**	**Pilot B Sample 1 to 3**
	18 + yrs (D10)[a]	**18-54yrs (D50)[a]**	**55+yrs (D20)[a]**	**35-54yrs (D20)[a]**	**18+yrs (D50)[a] (D20)[a]**
No. observations	1100	1130	410	480	2020
No. individuals	110	113	41	48	202
Likelihood ratio test	125.896	99.9	123.29	102.412	228.549
Adjusted rho-square	0.043	0.025	0.111	0.064	0.058
No.-ve QoL coefficients (%)	4 (14)	5 (18)	4 (14)	26 (93)	16 (57)

*The seven individuals who self reported not understanding the DCE tasks were excluded from the analysis. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by two and D50 = standard duration levels scaled up by five

The results from the main effects model also suggest that presenting longer durations scaled up by more than two do not generate the parameter estimates with the expected sign. This is illustrated in the estimated parameter coefficients of Pilot B Sample 1, based on data obtained from 18 – 54 years olds given duration options scaled up by five (duration levels of 50, 45, 40, 30 and 25 years). Here only five (18%) of the estimated WAItE dimension level coefficients were negative. Overall, the results from the main effects models suggest that only the 35-54 year olds could on average consistently distinguish between different levels of the WAItE dimensions, and 55 plus year olds were on average not able to distinguish between the different levels, as only four (14%) of the parameters estimated for the WAItE level dummy variables were negative. The duration coefficient was positive for the model based on the data from Pilot B Sample 3, though in this case it was not significant (p=0.10)

8.3.2.2 DCE_{TTO} long vs. short durations – Interaction models

Table 8.6 presents the models estimating the interaction effects between the WAItE dimension levels and the duration attribute. The expectation would be that the interaction and duration coefficients have opposite signs. If the interaction coefficients between the WAItE and duration attributes were negative it would mean on average: respondents prefer to live in longer health profiles *and* prefer to live in less severe levels of each of the WAItE attributes. For the majority of the models, however, the attribute coefficients did not have the expected sign. Again, where data from Pilot B Sample 3 was utilised, the majority of the parameter estimates for the WAItE and duration interaction dummies had the correct sign (21 (75%)) and ordering. The next best performing model was generated when all of the data from Pilot B was combined (17 (61%)), but of these the majority were not statistically significant at the 10 % level.

When the results from Pilot B Samples 1 and 2 are considered, care needs to be taken when interpreting these results. The majority of the coefficients for Pilot B Sample 2 (generated from data composed of respondents aged 55 plus years old, presented with duration options scaled up by two) are statistically significant at the 10% level. However, contrary to what would be expected, the interaction parameter estimates have a positive sign whilst the duration attribute has a negative sign. This implies that on average, respondents preferred to live in shorter health profiles and in more severe levels of each of the seven WAItE attribute levels. Though counterintuitive to expectations, the calculation of $\frac{\hat{\gamma}'}{\hat{\beta}}$ from equation 7 can result in sensible results, as long as the magnitudes of the two parameters have the opposite signs.

Table 8:6 Choice models for Interaction effects models*

Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B - Samples 1 to3	
	18 + yrs (D10)[a]	p-value	18-54yrs (D50)[a]	p-value	54+yrs (D20)[a]	p-value	35-54yrs (D20)[a]	p-value	18+yrs (D50)[a] (D20)[a]	p-value
Tired_L2xD	0.036	0.13	**0.009**	**0.06**	**0.077**	**0.00**	0.026	0.18	**0.007**	**0.08**
Tired_L3xD	0.019	0.52	0.003	0.64	**0.096**	**0.00**	-0.022	0.29	-0.003	0.52
Tired_L4xD	0.027	0.51	0.0003	0.97	**0.123**	**0.00**	0.012	0.74	-0.007	0.18
Tired_L5 xD	0.002	0.98	0.006	0.58	**0.158**	**0.00**	-0.040	0.45	-0.007	0.21
Walking_L2 xD	**0.044**	**0.04**	**0.008**	**0.04**	0.032	0.12	-0.019	0.3	0.004	0.19
Walking_L3 xD	**0.052**	**0.06**	0.009	0.10	**0.097**	**0.00**	-0.021	0.48	0.003	0.46
Walking_L4 xD	**0.067**	**0.09**	0.007	0.35	**0.072**	**0.04**	-0.035	0.30	-0.003	0.43
Walking_L5 xD	0.082	0.14	0.013	0.22	**0.138**	**0.00**	-0.035	0.50	0.001	0.80
Sports_L2 xD	0.030	0.16	0.006	0.16	**0.038**	**0.03**	0.034	0.12	0.003	0.33
Sports_L3 xD	**0.059**	**0.03**	0.007	0.18	**0.043**	**0.20**	0.029	0.26	0.002	0.58
Sports_L4 xD	0.054	0.16	0.009	0.17	**0.116**	**0.00**	0.019	0.57	0.002	0.68
Sports_L5 xD	0.047	0.41	0.012	0.27	**0.152**	**0.01**	0.004	0.94	5.22E-05	0.99
Concentrate_L2xD	0.026	0.24	0.001	0.82	**0.036**	**0.08**	**-0.052**	**0.01**	-0.003	0.36
Concentrate_L3xD	0.027	0.40	0.003	0.63	**0.094**	**0.00**	-0.024	0.43	-0.002	0.60
Concentrate_L4xD	0.052	0.19	0.005	0.49	**0.065**	**0.10**	**-0.086**	**0.01**	**-0.008**	**0.09**
Concentrate_L5xD	0.039	0.46	0.004	0.68	**0.121**	**0.01**	-0.081	0.10	**-0.012**	**0.03**
Embarrased_L2xD	0.014	0.53	0.006	0.14	0.0293	0.27	-0.002	0.93	0.002	0.54
Embarrased_L3xD	0.005	0.87	0.003	0.60	**0.068**	**0.02**	-0.038	0.15	-0.001	0.74
Embarrased_L4xD	0.015	0.72	0.008	0.27	**0.082**	**0.06**	-0.051	0.11	-0.001	0.92
Embarrased_L5xD	-0.007	0.90	0.009	0.39	**0.131**	**0.04**	-0.040	0.44	-0.005	0.27
Unhappy_L2xD	**0.051**	**0.01**	**0.010**	**0.01**	**0.025**	**0.15**	0.028	0.12	0.006	0.15
Unhappy_L3xD	0.030	0.26	0.0012	0.81	**0.090**	**0.00**	-0.036	0.18	**-0.006**	**0.08**
Unhappy_L4xD	0.015	0.72	-0.001	0.90	**0.127**	**0.00**	-0.052	0.18	**-0.011**	**0.01**
Unhappy_L5xD	-0.005	0.93	0.001	0.94	**0.142**	**0.01**	-0.083	0.11	**-0.014**	**0.01**

TreatDif_L2xD	0.021	0.33	0.001	0.80	**0.049**	**0.01**	-0.024	0.10	-0.002	0.50
TreatDif_L3xD	0.021	0.47	0.008	0.16	**0.123**	**0.00**	-0.023	0.33	0.001	0.73
TreatDif_L4xD	0.016	0.67	0.005	0.49	**0.111**	**0.01**	-0.049	0.10	-0.005	0.20
TreatDif_L5xD	0.016	0.78	0.004	0.74	**0.150**	**0.02**	-0.046	0.38	**-0.012**	**0.02**
Duration	0.003	0.99	-0.003	0.95	**-0.615**	**0.03**	0.423	0.10	-	-
Duration_L2	-	-	-	-	-	-	-	-	**0.511**	**0.00**
Duration_L3	-	-	-	-	-	-	-	-	**1.040**	**0.00**
Duration_L4	-	-	-	-	-	-	-	-	**1.370**	**0.00**
Duration_L5	-	-	-	-	-	-	-	-	**1.930**	**0.00**
Constant	0.0195	0.75	-0.0294	0.66	-0.104	0.47	0.178	0.11	-0.031	0.53

Interaction effects models	**Pilot A**	**Pilot B Sample 1**	**Pilot B Sample 2**	**Pilot B Sample 3**	**Pilot B Sample 1 to 3**
	18 + yrs (D10)[a]	18-54yrs (D50)[a]	55+yrs (D20)[a]	35-54yrs (D20)[a]	18+yrs (D50)[a] (D20)[a]
No. observations	1100	1130	410	480	2020
No. individuals	110	113	41	48	202
Likelihood ratio test	127.757	103.547	127.233	99.534	224.071
Adjusted rho-square	0.044	0.028	0.118	0.059	0.056
No.-ve QoL coefficients (%)	2 (7)	**NA**	**NA**	21 (75)	17 (61)

* The seven individuals who self reported not understanding the DCE tasks were excluded from the analysis. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by two and D50 = standard duration levels scaled up by five

8.3.2.3 DCE_{TTO} vs $DCE_{NoDuration}$ – Main effects models

Table 8.7 shows the results of comparing the DCE_{TTO} and $DCE_{NoDuration}$ methods. The data used for the DCE_{TTO} models was based on the choice tasks where the duration level option was the same across the two alternatives (although the DCE_{TTO} design will now be inefficient as this was not its original purpose – the implications of this were highlighted by the observed negative estimates for the rho square parameters for the DCE_{TTO} models). The results in **Table 8.7** show, where there is no duration attribute, that the model utilising data from Pilot C performs the best in terms of the number of estimated negative WAItE level parameters. However, the design for this study only was based on not having a duration attribute, making it more efficient compared to the design of the other pilot studies. Bearing this in mind, Pilot B Sample 3, as with the two models previously discussed, generated a high number of negative WAItE coefficients. Overall there were very few statistically significant parameter estimates for the models utilising data from Pilots A and B. The rho squared parameter for these models were negative, and P-values of 1 were observed for model parameters utilising Pilot B Sample 2 and 3 data.

8.3.2.4 $DCE_{NoDuration}$

The results for the anchoring tasks are shown in **Figures 8.4 and 8.5** (frequency tables are provided in **Appendix 8.3** and **8.4**). When the VAS was used, the values participants assigned to the PITS state covered the full length of the thermometer. The mean value for the PITS state was 0.23 and the median value was 0.20 when VAS was utilised. Comparatively, when the pairwise anchoring question was used the median value for the PITS state was 0.75.

8.3.2.5 Time preference

Responses to the time preference question included in Pilot B showed that 49% of the respondents had positive time preference 41% negative time preference and 10% had no preference.

Table 8:7 Parameter estimates for Main effects models (No Duration attribute)*

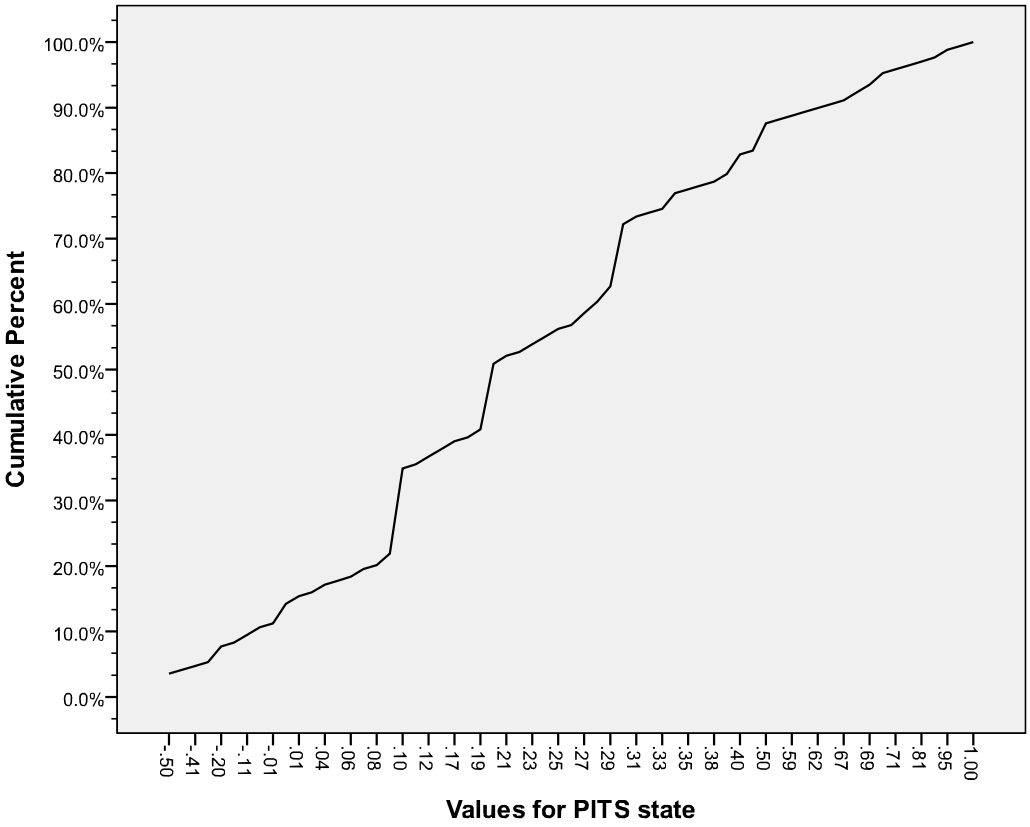
Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B - Sample 1 to 3		Pilot C	
	18 + yrs (D10)[a]	p-value	18-54yrs (D50)[a]	p-value	54+yrs (D20)[a]	p-value	35-54yrs (D20)[a]	p-value	18+yrs (D50)[a] (D20)[a]	P-value	18 + yrs (No D)[a]	p-value
Tired_L2	0.408	0.53	0.617	0.49	2.500	1	-9.490	1	0.175	0.58	0.022	0.88
Tired_L3	0.597	0.85	-0.259	0.70	4.370	1	-7.770	1	-0.066	0.92	-0.112	0.62
Tired_L4	0.286	0.84	**-1.010**	**0.01**	-2.290	1	0.285	1	-0.153	0.91	-0.392	0.23
Tired_L5	0.22	0.69	-0.314	0.84	-0.194	1	-3.310	1	-0.014	1	-0.689	0.11
Walking_L2	-0.103	0.97	0.610	0.49	5.260	1	3.060	1	-0.090	0.89	0.066	0.59
Walking_L3	0.049	0.97	0.100	0.94	-3.110	1	1.200	1	0.179	0.80	-0.254	0.26
Walking_L4	0.446	0.83	-0.736	0.19	-1.400	1	-3.770	1	-0.269	0.79	-0.210	0.52
Walking_L5	0.166	0.94	-0.504	0.71	5.670	1	-1.190	1	-0.178	0.27	-0.490	0.25
Sports_L2	-0.470	0.77	0.719	0.71	5.010	1	1.610	1	0.108	0.68	0.147	0.29
Sports_L3	-0.461	0.87	0.790	1	0.559	1	2.440	1	0.191	0.64	-0.035	0.87
Sports_L4	-0.126	0.90	0.896	1	0.244	1	-6.330	1	0.242	0.84	-0.041	0.90
Sports_L5	-0.160	0.95	0.463	0.77	1.070	1	-2.380	1	0.226	0.57	-0.286	0.51
Concentrate_L2	-0.228	0.92	0.248	0.88	-0.858	1	-2.110	1	0.055	0.93	-0.073	0.60
Concentrate_L3	0.200	0.93	-0.224	0.89	1.910	1	4.330	1	0.142	0.83	-0.023	0.92
Concentrate_L4	0.312	0.92	-0.082	0.97	-3.290	1	-2.470	1	-0.181	0.01	-0.220	0.51
Concentrate_L5	0.437	0.80	-0.575	0.40	0.408	1	4.390	1	-0.154	0.73	-0.563	0.21
Embarrased_L2	0.662	0.58	-0.32	0.5	1.430	1	-3.330	1	0.043	0.97	0.130	0.32
Embarrased_L3	-0.300	0.91	0.167	0.94	3.310	1	6.170	1	0.084	0.92	-0.237	0.30
Embarrased_L4	-0.269	0.82	0.261	0.80	2.830	1	-6.140	1	-0.016	0.98	-0.184	0.57
Embarrased_L5	-0.511	0.48	0.885	0.60	-4.030	1	-1.700	1	0.117	0.72	-0.360	0.40
Unhappy_L2	0.640	0.62	0.059	0.99	-8.040	1	-1.750	1	0.149	0.91	0.002	0.99
Unhappy_L3	0.031	0.99	0.149	0.67	6.110	1	-1.010	1	-0.046	0.94	-0.090	0.70
Unhappy_L4	-0.375	0.67	-0.150	0.90	10.00	1	-1.640	1	-0.179	0.90	-0.282	0.39
Unhappy_L5	-0.080	0.93	0.203	0.71	-5.010	1	0.940	1	-0.110	0.94	-0.637	0.15

TreatDif_L2	0.703	0.27	0.145	0.89	-0.691	1	-2.09	1	0.204	0.92	-0.010	0.49
TreatDif_L3	0.185	0.91	-0.187	0.91	4.160	1	-0.462	1	-0.013	0.99	**-0.382**	**0.07**
TreatDif_L4	-0.188	0.95	-0.172	0.83	-2.500	1	8.820	1	-0.156	0.81	-0.510	0.12
TreatDif_L5	-0.569	0.77	0.119	0.69	-6.820	1	10.000	0.00	0.109	0.89	**-0.864**	**0.04**
Constant	-0.039	0.97	0.0282	0.98	-3.050	1	4.610	1	0.044	0.94	-0.089	0.13

Main effects models	**Pilot A**	**Pilot B Sample 1**	**Pilot B Sample 2**	**Pilot B Sample 3**	**Pilot B Sample 1 to 3**	**Pilot C**
	18 + yrs (D10)[a]	18-54yrs (D50)[a]	55+yrs (D20)[a]	35-54yrs (D20)[a]	18+yrs (D50)[a] (D20) [a]	18+yrs (No D)
No. observations	237	239	106	112	457	1700
No. individuals	110	113	41	48	202	170
Likelihood ratio test	32.911	32.236	26.278	33.992	21.306	90.905
Adjusted rho-square	-0.076	-0.078	-0.216	-0.155	-0.058	0.014
No.-ve QoL coefficients (%)	13 (46)	12 (43)	12 (43)	17 (61)	14 (50)	23 (82)

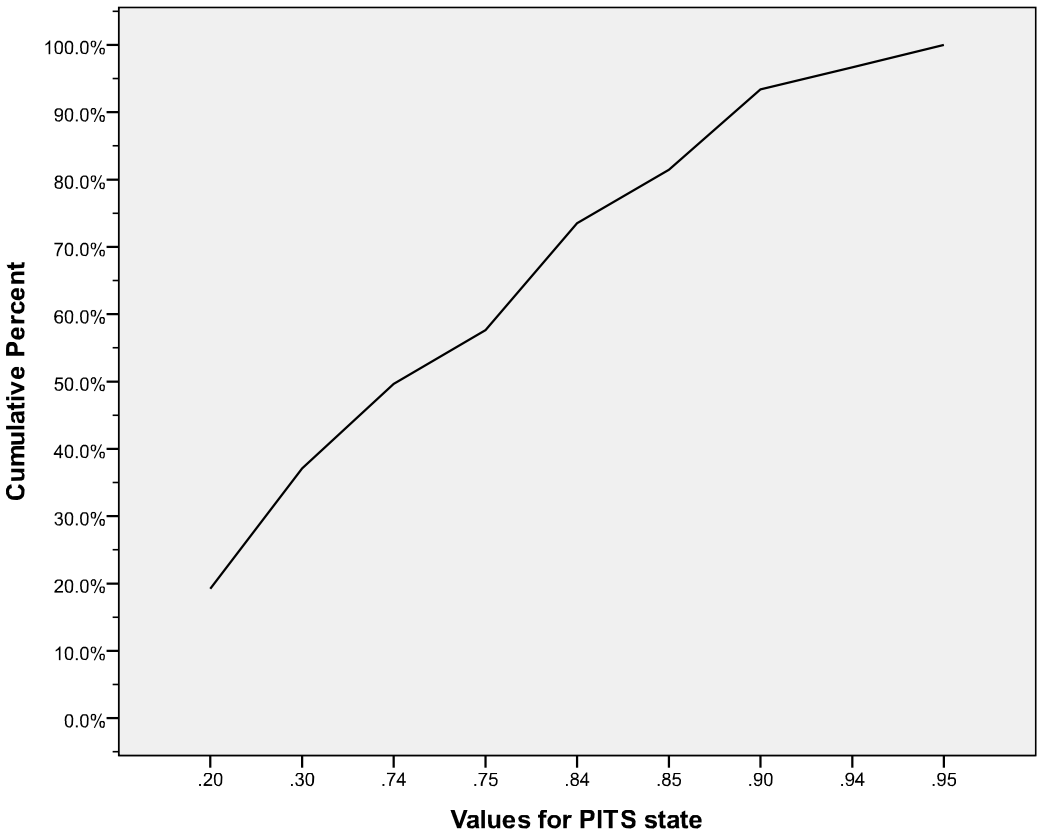
* The seven individuals who self reported not understanding the DCE tasks were excluded from the analysis. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by two and D50 = standard duration levels scaled up by five. **NOTE:** Negative adjusted rho-square values were calculated for all models apart from the model based on Pilot C data and are in red. The implications of this are provided in the discussion section.

Figure 8:4 Valuation of the PITS state using VAS Anchoring*



* There were two individuals whose responses were excluded. These individuals gave a response of 'good' and 'B' and so it was assumed that they did not understand the VAS task.

Figure 8:5 Valuation of the PITS state using pairwise Anchoring



8.3.3 Stage 2 - Calculation of utility values

The main objective of the current study was to evaluate the applicability of the DCE_{TTO} and $DCE_{NoDuration}$ approaches in the elicitation of preference values for WAItE health states by assessing the estimated coefficients generated in models presented in the previous section. This section provides an illustrative example the calculation of utility values (Stage 2) for DCE_{TTO} and $DCE_{NoDuration}$.

DCE_{TTO}

The DCE_{TTO} approach proposed by Bansback *et al*. (2012) is used to calculate utility values by utilising the parameter estimates obtained from the coefficients estimated from the interaction effects in models utilising data from Pilot B. The utility values for WAItE states 5555555 and 2222222 are presented in **Table 8.8** using the Bansback *et al*. (2012) approach. The coefficient estimates from **Table 8.6** are used to populate the parameters from equation 8 above. The estimated value for health state 5555555, using the coefficients estimated for Sample 3, would be calculated as: 1.00 + (- 0.096 - 0.083 + 0.009 - 0.192 - 0.094 - 0.197 - 0.110) = 0.238. Estimated utility values that were valid (i.e. values not exceeding one) ranged from -0.613 to 0.238 for health state 5555555 and 0.535 to 0.980 for health state 2222222. These estimates were generated from data obtained from Sample 2 and Sample 3.

Table 8:8 Utility values generated from application of the DCE_{TTO}

Parameters*	**DCE_{TTO} Pilot B by sample**		
	Sample 1	**Sample 2**	**Sample 3**
Tired_L5 xD / Duration	-1.913	-0.257	-0.096
Walking_L5 xD / Duration	-3.988	-0.224	-0.083
Sports_L5 xD / Duration	-3.801	-0.247	0.009
Concentrate_L5 xD / Duration	-1.287	-0.197	-0.192
Embarrased_L5 xD / Duration	-2.670	-0.213	-0.094
Unhappy_L5 xD / Duration	-0.243	-0.231	-0.197
TreatDiff_L5 xD / Duration	-1.100	-0.244	-0.110
Value HS 5555555	-14.000	**-0.613**	**0.238**
Tired_L2 xD / Duration	-2.648	-0.125	0.062
Walking_L2 xD / Duration	-2.555	-0.052	-0.045
Sports_L2 xD / Duration	-1.838	-0.062	0.080
Concentrate_L2 xD / Duration	-0.306	-0.058	-0.123
Embarrased_L2 xD / Duration	-1.847	-0.048	-0.004
Unhappy_L2 xD / Duration	-3.000	-0.041	0.066
TreatDiff_L2 xD / Duration	-0.295	-0.079	-0.057
Value 2222222	-11.489	**0.535**	**0.980**

*The parameters were generated through the application of coefficient estimates obtained from **Table 8.6** to equation 8

$DCE_{NoDuration}$

Two things were necessary in order to use equation 11 to calculate the re-anchored value of state 2222222 using the $DCE_{NoDuration}$ method: 1) the two median values for the PITS WAItE state generated from Pilot C data and 2) the coefficient estimates provided in **Table 8.7** for the $DCE_{NoDuration}$ model results. The estimated value for health state 2222222 ranged from **1.058** to **1.018** respectively when VAS and pairwise anchoring was utilised. Both these estimated values exceed the value for full health.

8.4 Discussion

This study has presented and assessed alternative DCE methodologies for estimating health state values for a new weight specific measure – the WAItE. Although none of the model results generated consistently ordered and statistically significant estimates for the WAItE dimension level coefficients, the small sample sizes used in the pilot studies were unlikely to produce robust and definitive parameter estimates. DCE sample size estimates are generally based on rules-of-thumb and budget constraints. The minimum sample size for a DCE study normally ranges between 150 and 1200 respondents (Orme, 1998). Nonetheless, as the parameter estimates that were generated from the pilot studies provide a good indication of what to expect if a larger DCE valuation study is undertaken, it is worth discussing why the expected results were not found i.e. why it was that a) DCE_{TTO} did not generate negative consistently ordered significant parameters and, b) $DCE_{NoDuration}$ generated more negative parameters but mostly not statistically significant. A critical evaluation of the application of the four steps undertaken in carrying out the DCE valuation study, detailed within the methods section of this chapter, is used to structure the discussion. The final section provides some reflections on the general limitations of the valuation study and, were the study to be repeated, some suggestions on what could be done differently.

8.4.1 Identifying attributes and levels and data collection

The seven QoL attributes were generated from the new WAItE instrument: the reasoning behind its creation and the discussion about its development has been addressed in earlier chapters. The valuation study was undertaken with the adults aged 18 years and over in line with the NICE guidance (NICE, 2013). However, the elicitation of utility values for states described by the WAItE from an adult population, instead of the population for which the instrument was intended, namely adolescents, could raise potential problems. The WAItE items could be inappropriate for the adult population. Looking at item 7 of the WAItE descriptive system, for example, *People treat me differently when I go out*: for an adolescent

this could be a very important issue and have a large negative impact on their overall QoL as where, when, and whom they go out with may be outside of their control. On the other hand an adult may see this as a trivial issue and something that is either easily avoided. Owing to this, supplementary assessment of the suitability of the WAItE in the adult population was undertaken by evaluating its psychometric properties in this population (see **Appendix 8.5 and 8.6**). A good Cronbach's Alpha statistic of 0.826 was calculated based on the data from the 489 adult participants who completed the WAItE (see **Table 8.4 above**), and there was a statistically significant inverse relationship between the BMI and the total WAItE score (Pearson's rho correlation coefficient = +0.170, p<0.00). This suggests that the WAItE, though designed for adolescents, displayed good psychometric performance when used with adults. The ordering of the 5 level response options could also potentially differ between adults and adolescent. Again, supplementary Rasch analysis was used to empirically test this issue and the results showed that there was agreement over the response option ordering across the two populations (see **Appendix 8.7**).

The focus of discussion in the remainder of this section will be around the duration attribute and the choice of levels for this attribute. The DCE_{TTO} methodology had been shown to be successfully applied to the EQ-5D. As the EQ-5D describes more severe health states relative to the states described by the WAItE, the use of longer time horizons were also assessed in the current study. The shorter time horizons offered to respondents were based on the duration levels used in the Bansback *et al*. (2012) study, which valued EQ-5D health states, could be perceived as potentially very short when combined with a comparatively mild descriptive system such as the WAItE. Two sets of longer duration options were offered in order to test whether short survival prospects *combined with* the relatively mild health states would lead respondents not to trade years of life for better health. In order to test the effect of having scaled up durations, different age groups were given different duration options. The principal findings from adopting the DCE_{TTO} in this study were as follows: If the DCE_{TTO} approach developed by Bansback *et al*. (2012) is to be adopted for the valuation of the WAItE, the duration levels need to be carefully considered. The shorter duration levels used in conjunction with the valuation of the EQ-5D do not seem to be compatible with the milder health states described by the WAItE. On the other hand, the use of extremely long duration level options of 50, 45, 40, 30 and 25 years for WAItE DCE profiles also did not produce parameter estimates that would be expected. The best performing model using DCE_{TTO} was generated from data obtained from 35-54 year olds valuing DCE profiles combined with duration levels of 20, 18, 16, 12 or 10 years. When the same duration levels are implemented for individuals aged 55 plus years, this finding was not replicated. The implications of this are

that only 35-54 year olds seem to be able to distinguish between the dimension levels of the WAItE. A supplementary analysis to empirically test this was undertaken (see **Appendix 8.8 and 8.9**). Models were estimated based on the data obtained from only 35 to 54 year olds from Pilot A (utilising duration levels of 10, 9, 8, 6 or 5 years) and Pilot B Sample 1 (utilising duration levels of 50, 45, 40, 30 and 25 years) to assess whether the expected parameter coefficients are estimated. If so, this would imply that 35 to 54 year olds can distinguish between the WAItE dimension levels regardless of the duration levels that are used. The results of this analysis showed that the estimated parameters for 35 to 54 year olds did not generate consistently ordered and negative parameters when the main effects and interaction models were estimated for the alternative duration options. It is worth noting, however that these results were based on very small samples (data was available from approximately 50 participants in each of the three studies). The implications associated with having too low a ratio of observations to regressors is further discussed in the **Section 8.4.3**.

In order to address the situation where individuals may have been unwilling to trade any life to live in a better health state as described by the WAItE, $DCE_{NoDuration}$ was also considered. This situation would come about if making the duration levels longer was not enough to encourage trading. Here a different design with just the standard DCE consisting of only WAItE states was generated and two forms of external anchoring were considered. Although the model estimates for this design performed the best in terms of the number of WAItE parameter estimates that were negative, this was the only design that was generated for this type of model. The anchoring results showed that the VAS value for the PITS state was a lot lower than the value for the PITS state generated from the pairwise anchoring. This may have been because there was no trading opportunity in the VAS task. Given the large difference in the estimated values of the PITS state between the two anchoring strategies, it is difficult to identify which is the true estimate without further empirical investigation in future work.

For the DCE_{TTO} and the pairwise anchoring tasks individuals may have primarily focused on the duration attribute, leading to the problem of attribute non-attendance: the tendency for respondents to ignore one or more of the attributes in DCE tasks. Respondents may have been solely focused on the duration attribute and ignored the WAItE QoL attributes, in the DCE_{TTO} profiles, leading to biased coefficient estimates (Hole at al., 2012).

8.4.2 Experimental design

It is important to consider the role of the experimental design in generating the unexpected parameter estimates. The Ngene programme was asked to generate 90 pairwise choice sets for the estimation of both the main effects ($DCE_{NoDuration}$) and interaction (DCE_{TTO}) models. Intuition suggests that the design estimating interactions between the QoL and duration attributes might need a larger number of choice sets to be directly valued, than the design based on the estimation of main effects. This is something that needs to be investigated further in future work. To the knowledge of the candidate no guidance exists for selection of number of choice sets when interactions are involved. Nevertheless, the number of profiles that are directly valued, in a DCE survey, should take into account the attributes and complexity of choice sets. Another design issue that may have affected the results was that the order in which choice sets were presented to respondents may have influenced their choices. Future work looking into so called *ordering effects* may be beneficial.

In terms of the number of choice tasks that respondents should be asked to complete in a DCE survey, consensus as to the optimal number is lacking. In the studies evaluated by the Ryan & Gerard (2003) review, respondents were asked to complete between nine and 16 choice tasks each. In the recent Bansback *et al*. (2012) study each respondent answered a total of 13 tasks comprising five TTO and eight DCE_{TTO} exercises. Based on the lack of guidance in this aspect of DCE design, the decision was made to set this number at 10 in order to avoid overburdening respondents.

In a DCE study, it is important that respondents are taking the tasks seriously. This requires the alternatives being presented to respondents to be realistic. The model estimates may have been affected because some of the alternatives making up the choice sets were not sufficiently realistic to respondents. For example in Pilot B, individuals in Sample 1 aged 50 to 54 were given duration options of 50, 45, 40, 30 and 25; these individuals may not have thought it realistic for them to live for upwards of another 40 years. Related to this, respondents could also have been asked to indicate the health state they thought was *worse*, rather than which they *preferred*. Alternative wording in this manner may have simplified the choice task for respondents.

8.4.3 Data analysis and interpretation

In terms of Stage 1 of the estimation of the main effects and interaction effects models, in the estimation of the utility function, the MNL model was used in this study. In his thesis, Bansback (2010) compared the model results generated from the application of the conditional logit (synonymous with the MNL model), random effects logit and mixed logit models and found that the results of the conditional logit and random effects logit models were similar. However there was a marginal improvement in the log likelihood and prediction success when the latter was used. Owing to this, the random effects logit was chosen to be the final model for the application of DCE_{TTO}. Another study in a related area, where the DCE_{TTO} approach was employed, utilised random-effects probit models (Norman *et al*., 2012). The advantage of using random effects models is that they account for the panel nature of data (i.e. the estimation method allows for the fact that multiple observations come from one respondent)[13]. A limitation associated with the use of these models, particularly relevant in the current study setting, is that more complex models require more observations, and thus, more data would be necessary.

One of the key limitations associated with making direct comparisons between the model results generated from the $DCE_{NoDuration}$ and DCE_{TTO} datasets was that the design of the latter was made inefficient by the removal of the duration parameter. The adjusted rho squared estimates, a measure of how well observed outcomes are replicated by the model, were negative for Pilot A and B from **Table 8.7**, indicating very poor model fit (a rho squared value close to one indicates good model fit). In addition to this, the P-values for the models based on data from Pilot B Samples 2 and 3 were extremely high (equal to 1), which is an indication that the model did not converge. This could have been observed because of the problem of near micronumerosity, christened by Arthur Goldberger (Goldberger, 1991). *According to Goldberger; micronumerosity, arises when the number of observations barely exceeds the number of parameters to be estimated* (Gujarati, 2003, p. 348). Thus the parameter estimates from these models should be treated with caution.

[13] In order to better account for the panel nature of the data binary random effects models were run in the Stage 1 analysis (the estimation of the interaction effects models) but there was no difference in the results.

Further assessments of the application of random-effects models could be undertaken in future work (this may also address this issue of model misspecification when the constant term was observed to be statistically significant (**Table 8.5** – Pilot B Sample 3). In order to achieve good model fit, additional data collection would be necessary as well as additional data analysis. It was not possible to undertake this additional work within the time frame of the thesis but it can be investigated in future work.

The results from Pilot B Samples 1 and 2 generated interaction parameter estimates (**Table 8.6**) with a positive sign whilst the duration attribute had a negative sign. This could be interpreted that on average, respondents preferred to live in shorter health profiles and in more severe levels of each of the seven WAItE attribute levels. Intuitively this is questionable. When we consider Stage 2 of the Bansback *et al*. (2012) methodology, however, the anchoring utility function utilises the coefficient estimates in Stage 1 through the calculation of the ratio between interaction parameter estimates and the duration parameter estimate. In order for the function to work correctly the ratio between the duration and WAItE coefficients should be negative, meaning that *either* the numerator (the interaction coefficient) or the denominator (the duration coefficient) has to be negative. In terms of equation 6 though, the duration parameter should be positive and WAItE dimension levels should be negative. In the application of Stage 2 of the Bansback *et al*. (2012) analysis methodology, the utility function presented in equation 8 was used to calculate utility values based on the parameters generated from the interaction effects models in Stage 1. Using the results from Pilot B, for illustrative purposes, the QoL weights obtained by dividing the interaction coefficient with the duration coefficient would be expected to be between 0 and say, -1.0. However, the utility calculation based on parameters generated from Sample 1 data were not as would be expected (for example some of the calculated values exceeded the value for full health). Based on the parameter estimates from Samples 2 and 3, the utility values for the PITS health state 5555555, ranged from -0.613 to -0.238. These values are very low in comparison to the valuation of the same health state using the two anchoring methods (median values were 0.20 for VAS and 0.75 when the pairwise anchoring). Also these values are very low when compared to -1.133 calculated by Bansback *et al*. (2012) for the worst EQ-5D health state (33333 – the worst level on all five dimensions of the descriptive system).

Time preference is an important issue when asking individuals to consider hypothetical health states further in the future rather than in the immediate future, having implications on the interpretation of the estimated coefficient. In the current analysis, it is assumed that a DCE_{TTO} choice involving 10 years in a slightly dysfunctional WAItE state vs. eight years in a very slightly

dysfunctional WAItE state should be the same as 40 years in the former state vs. 32 years in the latter state (i.e. interpreted as identical choice set). If people pay more attention to the eight year difference in the latter pair (as opposed to the two year difference in the former pair), ignoring that it is over 40 years instead of 10, then it means the assumption of proportionality may not hold. Also if people have a positive time preference (i.e. prefer benefits now rather than in the future), then a benefit (health improvement) of eight years not starting for 32 years may be valued less than a benefit of eight years starting next year, say. Therefore it could be argued that if longer duration options are being used, then time preference rates should also be elicited. A counter argument to this is that, as DCEs aim to elicit aggregate rather than individual level values, at the aggregate level, time preference could be assumed to be zero or neutral. This may result from having approximately an equal number of individuals within a given population who have positive or negative rate of time preference. The latter argument was empirically tested with the inclusion of a time preference question. The results for Pilot B showed that there was approximately an even split between the sample with those individuals with a positive time preference and those with a negative time preference. This suggests that, at the aggregate level, it may not be necessary to apply a discount rate for time preference when longer levels of duration are offered in a DCE survey.

8.4.4 Overall reflections

In its current from the WAItE can be used to undertake CEA of weight management interventions (more detailed discussion of this is provided in **Chapter 9**), however QALYs cannot be calculated. If the valuation study had been successful then it would have been possible to calculate weight specific utility values and it would have been possible to undertake CUA of weight management interventions. If it was possible to calculate QALYs from the WAItE, economic evaluations of weight management interventions would need to look into how best to utilise QALY calculations into their analysis and whether a short or long term analysis is undertaken. If a long term analysis is undertaken and the impact of the intervention assesses a life time horizon then the analysis might also need to incorporate utilities from other sources such as the EQ-5D or HUI to obtain adult utility values reflecting benefits in later life. Consensus in terms of the optimal method of undertaking economic evaluations of weight management interventions is yet to be reached and is a fast developing research area**.**

In **Chapter 4** a discussion of the key stages involved in the development of a new preference based instrument was provided. It is worth re-visiting the last two issues concerning a) *What technique should be used for valuing health* and b) *who should provide preference values*. In terms of the former DCE was chosen due to the benefits of applying a new method, DCE_{TTO}, to a different study setting and also due to time and resource constraints. The benefit of applying DCE_{TTO} was that the WAItE could be anchored on the full health dead scale with the addition of a duration attribute. In addition to this two alternative anchoring strategies where tested when $DCE_{NoDuration}$ was utilised. As nether method generated the expected parameter coefficients, it is worth re-assessing the applicability of the valuation method that was chosen. Firstly, if states describes by the WAItE were mild, was it feasible or realistic to have a bottom anchor of dead. Perhaps it would have been beneficial for the bottom anchor to have been the worst state describes by the WAItE 5555555. This is of particular relevance as when the duration attribute was excluded the estimated coefficients were more in line with expectations (negative with increasing magnitude as the levels got worse). Additionally if the WAItE is to be used in combination with instruments like the EQ-5D or HUI then the need to anchor to dead is less obvious. In future work it is worth re-evaluating the need to have a duration attribute if DCE is used to value WAItE health states. Alternative valuation methods such as TTO or SG should also be considered if a new valuation study is undertaken to provide further empirical evidence of the suitability of anchoring to dead when the WAItE is used.

In terms of the issue around whose values to obtain - An internet panel was used to recruit members of the adult population in valuing states described by the WAItE. A key limitation of this could have been the fact that the respondent sample could not have been representative of the UK general population of adults. For example, selection bias might result from recruiting participants from only one internet panel. It might be possible in future work to recruit individuals from more than one internet panel. The decision to use adults in the valuation study was based on the recommendations from the NICE guidance (NICE, 2013), although as discussed above and again in **Chapter 9**, the potential for preference values obtained from adults could be inconsistent with those obtained from adolescents. A key reason for the discrepancy may be that a descriptive system describing how adolescents' lives are affected by weight status may differ a great deal compared to how weight affects different aspects of the lives of adults. In future work it might be beneficial to obtain the values of WAItE states from a sample of adolescents aged 16-18 years. This would provide a basis for evaluating whether adults and adolescents valuations differ.

8.5 Conclusion

This study has shown that results from the application of DCE_{TTO} did not generate the expected coefficients. $DCE_{NoDuration}$ generated more negative parameters, but most of these were not statistically significant. Moreover the alternative anchoring techniques that were utilised in the $DCE_{NoDuration}$ valuation, to enable the anchoring of WAItE states on the full health - dead scale, proved to be problematic. The results of this piloting work, in preparation for a future full valuation study of the WAItE, have revealed that DCE_{TTO} and $DCE_{NoDuration}$ employed in the current study may not be valid. Further work needs to be undertaken to identify the problems causing the unexpected results and perhaps also look into the feasibility of the application of alternative preference elicitation techniques.

Chapter 9
Overall discussion and directions for future research

9.1 Introduction

This chapter provides a summary of the thesis, the limitations of the empirical studies, and an assessment of the contribution of this research to the wider literature. Suggestions for the potential directions for future research are also presented.

9.2 Summary of the thesis

The premise of this thesis was outlined in **Chapter 1**: the need for a valid and reliable instrument to measure weight specific QoL to undertake fully informed economic evaluations of weight management interventions aimed at the adolescent population. **Chapter 2** provided the background literature for this research context. A more detailed report of the existing evidence was given in **Chapter 3** where a literature review was undertaken. Details of the existing generic and weight specific QoL instruments that have been used with obese and overweight adolescents and their limitations for providing valid and reliable QALY calculations particularly for the adolescent population was discussed. **Chapter 4** presented the key steps involved in the development of a new preference based measure and the associated methodological challenges. The findings from **Chapters 3 and 4** informed the methods undertaken in the empirical studies carried out in the thesis leading to the creation of the new Weight-specific Adolescent Instrument for Economic-evaluation, the WAItE. **Chapter 5** provided an account of qualitative interviews with adolescents and consultations with specialists in the field of adolescent obesity, which informed the construction of the weight specific LLDS. The psychometric assessments and Rasch analysis, that were used to identify the final seven item WAItE, were presented in **Chapter 6**. The preliminary assessment of the psychometric performance of the WAItE was considered in **Chapter 7**. The final study was described in **Chapter 8**, where a methodological assessment of the feasibility of the valuation of the WAItE was undertaken. Several important findings have been observed by the multiple methodological approaches that were undertaken in this thesis. The implications of the study findings are discussed below in relation to their consistency with previously published literature in relevant research areas.

9.2.1 The impact of weight status on adolescents' QoL

The qualitative interview work undertaken with adolescents demonstrated that weight status impacts upon a wide spectrum of aspects of their QoL. This is in line with existing weight specific instruments developed: a) where the content was informed by clinicians and specialists in the field of adolescent obesity (IWQOL – Kids (Kolotkin *et al*., 2006), Sizing Me Up (Zeller and Modi, 2009) and Sizing Them Up (Zeller and Modi, 2008)) and b) where qualitative interviews with adolescents were used to inform the content (OPOI (Doyle, 2011) and YQOL-W (Morales *et al*., 2011)). The effect of weight status on areas of physical functioning/comfort, emotional functioning/body esteem and social functioning were all assessed in existing measures. The IWQOl-Kids included items on family relationships and the Sizing Me-Up instrument included measures of positive social attributes – neither of which were covered by the WAItE. The issue of positive attributes associated with weight status is not applicable in this context as the purpose of the WAItE is to measure improvements in QoL that can be attributed to weight loss. Another finding was that similar issues were raised by both the younger (11-14 years) and older (15-18 years) adolescent age groups. This justifies the suitability of a single questionnaire for 11 to 18 year olds, given that item wordings are understandable for the younger age groups.

9.2.2 Constructing the WAItE

Classical psychometric assessments in conjunction with Rasch analysis have previously been used for generating a reduced item set from large condition-specific instruments, for preference valuation (Young *et al.*, 2009, 2010, and 2011). It was possible to identify dimensions of weight specific QoL, select seven items, the frequency response options from the LLDS, and create the final WAItE instrument, by utilising a web based survey and recruiting respondents via an internet panel. In previous studies, data were obtained from patient populations. However, within the current study the majority of the data were obtained from a general sample of adolescents. Initially, a treatment seeking sample was sought. However the response rate was very poor, and thus a general sample of adolescents was also recruited. The selection quota for the general sample of adolescents was set so that overweight or obese adolescents (calculated using self reported height and weight), who could be potential consumers of weight management interventions, were over sampled. This study was successful in identifying a descriptive system, from the 29 item LLDS, that was appropriate for the elicitation of preference values by selecting one item from the seven dimensions identified in the factor analysis (information from the original qualitative data was also utilised). The final descriptive system contained seven dimensions, each with five levels.

The analysis with the final seven selected items showed that the response options within dimensions might benefit from being collapsed from five levels to four levels. The collapsing of the response options was not undertaken at this stage and the original five response levels were used for the valuation study, which would provide further empirical evidence on the acceptability of the ordering of the original levels. The collapsing of response options after carrying out the valuation study is accepted practice (Stevens, 2012).

9.2.3 The psychometric performance of the WAItE

The preliminary findings of the psychometric performance of the WAItE were promising. The WAItE showed properties of internal consistency and the majority of adolescents completed the instrument at two time periods. A review of the measurement properties of existing weight specific instruments was reported in **Chapter 3**, the psychometric performance of existing instruments has been assessed in similar populations: adolescents undertaking weight management interventions. Empirical investigations of the psychometric properties of five of the existing weight specific measures were found: 1) IWQOL-Kids - Kolotkin *et al*. (2006) and Modi & Zeller (2011), 2) M-A QoL Q II - Moorehead *et al*. (2003), 3) Sizing Me Up – Zeller & Modi (2009), 4) Sizing Them Up - Modi & Zeller (2008) and 5) YQOL-W - Morales *et al*. (2011) and Patrick *et al*. (2011). The review showed that there was very little empirical evidence of the responsiveness of existing instruments, with only two of the aforementioned studies assessing the responsiveness of the Sizing Then Up and YQOL-W instruments. The review by Griffiths *et al*. (2010) assessed the relationship between weight status and QoL (further details of the review is provided in **Chapter 3**), and found that evidence of a significant relationship between decreased weight status and improvements in global self-esteem or total quality of life scores is weak. When the responsiveness of the WAItE was assessed a significant inverse relationship between weight loss and QoL was found when two outliers were excluded from the analysis. Further testing of the WAItE is a key requirement for future research and the alternative directions this could take are discussed later in the chapter.

9.2.4 Using DCE to value weight specific states

The methodological study assessing the feasibility of different applications of DCE to value states described by the WAItE, and showed that it is not feasible to generate utility values, that can be used in the calculation of QALYs, using the DCE variants employed within the study. The majority of estimated coefficient parameters were not as expected. Existing studies that have utilised similar methods have been successful in applying the DCE method, to generic descriptive systems such as the EQ-5D, with much larger sample sizes ranging from 1220 to 932 (Bansback *et al*., 2012 and Norman *et al*., 2012).

Some suggestions as to why the DCE method was unsuccessful in the valuation of a weight specific descriptive system are listed below:

- Adults are valuing a descriptive system describing how adolescents' lives are affected by weight status. Adults may value aspects of their own lives, that are not included in the descriptive system, more highly than the items included in the WAItE and thus the descriptive system may omit important aspects of QoL that apply to them
- Successful application of DCE to anchor on the full health - dead scale have utilised a generic descriptive system describing severe health states relative to the states described by the WAItE. Respondents may have felt that the states described by the WAItE were not severe enough to encourage trading life years

An unexpected finding from the DCE study was that *only* the 35-54 age group were able to distinguish between the WAItE dimension levels when the 20, 18, 16, 12 or 10 years duration levels were used. However, these findings were not replicated when 10, 9, 8, 6 or 5 and 50, 45, 40, 30 or 25 year duration levels were used with the same age group (though all three model estimates were based on relatively small sample sizes). These modelling results may have been observed because:

- Of a fallacy in the model estimates for the 20, 18, 16, 12 or 10 year duration levels. Had the DCE tasks been repeated on a different sample of 35 – 54 year olds using the same duration levels, it may be that the estimated coefficients would not comply with the expected results
- Out of the general adult population, 35-54 age groups are collectively most likely to understand the DCE task. However, this assumption cannot be verified unless a think aloud protocol (a method that allows researchers to understand, at least in part, the thought process of a study respondent by observing them while they attempt to complete a defined task and asking them to talk through their thought process (Newell

& Simon, 1972) is implemented as respondents complete the DCE tasks in further research, in order to understand the thought process of the different age groups.

Future work undertaking the valuation of the WAItE is also a fundamental requirement for further research and different avenues for this type of research are suggested later in the chapter.

9.3 Contribution of the research to the literature

The development of a new weight specific measure for adolescents addresses the following gaps in the literature:

- The measure was explicitly designed for use in economic evaluation as a weight specific preference based measure for adolescents
- The measure used qualitative interviews with adolescents living in the UK to involve them in its development from the beginning in the creation of item wordings and response options
- The development assessed whether a non-patient population could be used in the selection of items and response options to make the new measure amenable for preference elicitation
- The feasibility of utilising the DCE elicitation method for the valuation of a disease specific health states was tested

This new measure has great potential to expand the use of cost effectiveness analysis in economic evaluation of weight management interventions. One of the key strengths of the WAItE, which make it stand out from existing weight specific measures, is that its content was informed by adolescents living in the UK. Hence, it should be more applicable for the assessment of interventions targeted at adolescents based in the UK as it should be more culturally appropriate. The measure focuses on aspects of life affected by weight that are important to adolescents and thus it also stands out from existing generic preference based measures that may omit these important issues.

This research has taken a multiple methodological approach in that it has combined the use of qualitative (interviews for generating content) and quantitative (assessment of measurement properties involving psychometric assessments and Rasch analysis) methods to create the WAItE. The use of this approach has worked well in this study and has been found to be particularly useful for developing a new condition specific measure that has the required characteristics that facilitates the elicitation of preference values.

The preliminary findings of this research has shown that the DCE_{TTO} and $DCE_{NoDuration}$ variants of the DCE elicitation method, applied to a representative sample of the UK adult population, in the valuation of weight specific health states described by the WAItE, were unsuccessful.

There are several key advantages to this new measure:

- It is based on the views of adolescents and has been developed using *bottom up* methods
- It is more appropriate for the UK population of adolescents as the content was informed by adolescents living in the UK
- It has been developed with adolescents who are treatment seekers and also a general sample of adolescents
- The views of specialists in the field of adolescent obesity were consulted in generating the content
- It is very short and the wording of items and response options are in the language that adolescents can understand
- Adolescents who do not have learning difficulties can self complete it
- The results of the initial psychometric assessments are very promising and it seems to perform well in terms of being sensitive to change in BMI
- It has the potential, with further empirical investigation utilising alternative valuation techniques such as TTO or SG, to generate utility values for the calculation of QALYs

9.4 Limitations of the research

This thesis needs to be viewed in light of the limitations of each of the empirical studies that have been undertaken. Initial discussions of the study limitations were provided in each of the relevant chapters. Here these are considered in terms of the validity and generalisability of the overall results.

The participants included in qualitative study that informed the content of the new weight specific measure were recruited from two settings: treatment seekers accessing community weight management interventions and a general school sample. Other potential avenues for recruitment could have been pursued and thus, broadening the diversity of the sampling frame. However, as the second section of the interview schedule utilised themes from the literature to elicit the views of adolescents regarding the impact of weight on QoL, it is unlikely that any important issues were missed. Furthermore, due to the requirement of eliciting preference values for health states, preference based measures are limited in their length, and thus they cannot include all aspects of QoL affected by weight status. Future validation on the WAItE should consider the recruitment of respondents from different sites and from different weight management programmes in order to reduce the problem of selection bias (that could be generated from the recruitment of participants from one site or when assessing only one type of weight management intervention). One obvious problem with the study design of the assessment of the measurement properties of the WAItE was the absence of an alternative intervention for weight management such as surgery or pharmaceutical interventions. This poses a number of problems with regard to generalising the results of this preliminary study to other types of interventions. However, this was beyond the scope of a small study, and indeed should be addressed and explored in a future validation study. In future work a larger sample size is needed to establish whether the preliminary findings that have been reported can be replicated. Overall, however, the preliminary findings show that the WAItE is able to detect change associated with weight management and weight loss.

An internet panel was used to recruit respondents for two of the studies: refining the weight specific 29 item LLDS and the DCE feasibility valuation study. The degree to which an internet sample reflects the wider population is uncertain. For the DCE methodological study, the sample did not necessarily have to reflect the general population of adults (being a feasibility study), though other studies utilising the DCE methodology have also recruited in the same way (Bansback *et al*., 2012 and Norman *et al*., 2012). In terms the recruitment of adolescents from an internet panel, additional uncertainties in terms of whether the adolescents self completed the survey or whether they were accessing weight management interventions at the time the survey was being completed, are unknown. These issues could have impacted on the selection of the items and response options making up the WAItE, though it is not possible to predict what differences would have been observed. Further psychometric testing of the WAItE in future work will provide evidence for the suitability of the final instrument in the measurement of weight specific QoL. This work will also enable the assessment of the generalisability of the WAItE to other types of weight management interventions.

Finally, across all studies, participants (both adolescents and adults) were predominantly British Caucasians. The findings in the qualitative study may under represent the impact of weight status on the QoL of ethnic groups. The WAItE was also developed and tested mainly with the white Caucasian population, and may not represent the impact of weight status on the QoL of adolescents from other communities. Furthermore, the findings from the DCE methodological valuation study may have been different if the study sample reflected the true percentage of adults from other communities and cultures living in the UK.

9.5 Directions for future research

There are two main areas where further research is required: firstly carrying out robust psychometric testing of the WAItE, and secondly further investigations into the feasibility of the valuation of the WAItE. The following hypotheses could be assessed:

Further psychometric testing of the WAItE

- How does the WAItE perform when assessing its measurement properties in terms of: reliability, validity, and responsiveness according to the COSMIN taxonomy (Mokkink *et al*., 2010b)?
    - How does the WAItE perform compared to other weight specific measures?
- Is the WAItE suitable for younger or older age groups?
- How does the WAItE perform on specific patient populations, such as patients accessing weight management interventions in the hospital or school setting?

Valuation of the health states described by the WAItE

- Is it possible for preference values for the health states described by the WAItE to be elicited using conventional methods such as SG or TTO?
- Is it possible to elicit preference values for the WAItE using computer-assisted personal interviews in the application of the TTO and DCE elicitation methods that are currently in development for the valuation of the EQ-5D-5L (Herdman *et al*., 2011)

9.5.1 Further psychometric testing of the WAItE

The psychometric testing of the WAItE should be informed by the COSMIN taxonomy and criteria for measurement properties (Mokkink *et al*., 2010). Utilisation of the COSMIN checklist in informing the psychometric testing of the WAItE will highlight key issues that can lead to research of a high methodological quality. The performance of the WAItE against existing weight specific QoL instruments such as the YQOL-W should be compared directly, as this allows further testing on comparable psychometric parameters.

The WAItE was developed with 11 – 18 year olds, but there is no reason that it may not also be suitable for other age groups. However, it may be the case that there are aspects of QoL affected by weight status that are missing from the instrument that are relevant to older or younger age groups (pre-adolescents and young adults for example). Conversely, there may be items in the WAItE that are not applicable for other age groups. This is something that could also be tested empirically.

A fundamental area of future research would be to test the performance of the WAItE for alternative weight management interventions in other settings such as the hospital or school settings. This will provide evidence to support or reject the use of the WAItE in different overweight and obese adolescent populations.

In its current form, the WAItE can be used to undertake assessments of cost effectiveness weight management interventions. Although outcomes will not be measured in terms of a QALY, there is the ability to calculate an overall WAItE score and to provide an assessment of whether baseline and post intervention scores have improved, stayed the same or worsened. The incremental cost effectiveness ratio, a calculation based on the ratio of the change in costs (e.g. $C_{\text{intervention group}} - C_{\text{control group}}$) to incremental benefits ($E_{\text{intervention group}} - E_{\text{control group}}$), could also be calculated. However, instead of measuring benefits in terms of QALYs these can be based on the WAItE total score gained or lost between the intervention and the control group (the control group could represent adolescents on a waiting list seeking treatment).

9.5.2 Valuation

The ultimate goal is for the calculation of QALYs to be possible using the WAItE, to carry out CUA, as per the NICE recommendations (NICE 2013). In **Chapter 2**, details of alternative preference elicitation techniques were presented. TTO has been successfully utilised in the elicitation of disease specific preference values (Yang *et al*., 2011 and 2009). Although valid and robust utility values could not be obtained for the WAItE using the DCE_{TTO} or $DCE_{NoDuration}$ elicitation methods, it may be that the application of SG or TTO may be more appropriate, though this would need to be empirically tested. The feasibility of utilising alternative computer based methods for the valuation of the WAItE can also be assessed by employing the computer-aided personal interview techniques applying the TTO and DCE elicitation techniques that are currently being developed by the EuroQol Group for the valuation of the EQ-5D-5L (Herdman *et al*., 2011). The application of novel methods in the design of elicitation tasks as well as new information in terms of how best to model preference data might provide the necessary tools with which to successfully elicit valid and robust preference values for states described by the WAItE.

Another option for future valuation work would be to use informed adult preferences. The adults in the DCE study did not know that they were valuing a descriptive system describing weight specific health states for adolescents. Individuals were asked to imagine themselves in each of the states, i.e. *as an adult*. It may be that if the respondents knew the states were in adolescence, and that they were describing being in above normal weight status, they might

value them differently. Again this is something that could be tested empirically. The priority of future work should be in obtaining preference values for WAItE from the adult population as this is what is currently recommended by NICE. However, once this has been addressed, future work could also look into the elicitation of preference values for WAItE states from adolescents. There is a paucity of exiting studies that have addressed the problem of obtaining preference values from the younger age group, as stated by Ratcliffe *et al*. (2011). Eliciting preference values for WAItE states will provide information about the importance of the different aspects of weight specific QoL to adolescents, and thus inform the development of new weight specific interventions by tapping into aspects of QoL that are most important to them.

9.6 Conclusion

In the face of infinite demand on health care resources, given a limited supply, the identification of the most cost effective interventions to fund is crucial. The global obesity epidemic presents a challenge in its prevention and management. At present there are a growing number of weight management interventions that have been developed, but there is no clear consensus on the most economically effective management and prevention strategies. Economic evaluations provide a formal method of comparing alternative health care interventions with regard to their resource utilisation (costs) and outcomes (effectiveness). This thesis has described the development of a new weight specific QoL instrument, suitable for the adolescent population (aged 11 to 18 years), in order to provide a measure of the effectiveness of the weight management interventions, expressed in terms of weight specific QoL scores. The new instrument has been developed with a specific purpose in mind: to be a preference based measure of QoL that can be used in the CUA of weight management interventions, though a valuation study needs to be undertaken in order to facilitate the calculation of QALYs. The **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (**WAItE**) could potentially be used to evaluate the difference in outcomes between alternative weight management interventions in order to calculate the additional cost per WAItE improvement: the ratio of the costs to the effectiveness of alternative interventions. This will serve as an important tool to help guide decisions about allocating scarce resources across competing weight management programmes.

References

ANDRICH, D, SHERIDAN, B & LUO, G 2010. Rumm 2030: Rasch Unidimensional Measurement Models [computer software]. http://www.rummlab.com.au/: RUMM Laboratory Perth, Western Australia.

ARNESEN, T & NORHEIM, O 2003. Quantifying quality of life for economic analysis: time out for time tradeoff. *Medical Humanities,* 29**,** 81-6.

BAARS, R, ATHERTON, C, KOOPMAN, H, BULLINGER, M, POWER, M & DISABKIDS-GROUP 2005. The European DISABKIDS project: development of seven condition-specific modules to measure health related quality of life in children and adolescents. *Health Quality of Life Outcomes,* 13**,** 3-70.

BAKKER, C & VAN-DER-LINDEN, S 1995. Health related utility measurement: an introduction. *Journal of Rheumatology,* 22**,** 1197-1199.

BANSBACK, N. 2010. *Investigating the use of discrete choice experiments to estimate societal health state utility values.* University of Sheffield.

BANSBACK, N, BRAZIER, J, TSUCHIYA, A & ANIS, A 2012. Using a discrete choice experiment to estimate health state utility values. *Health Econ,* 31**,** 306-18.

BLEICHRODT 2002. A new explanation for the deference between time trade-off utilities and standard gamble utilities. *Health Economics,* 11**,** 447-456.

BLIEMER, M, ROSE, J & HESS, S 2008. Approximation of Bayesian Efficiency in Experimental Choice Designs. *Journal of Choice Modelling,* 1**,** 98-127.

BRAZIER, J, DEVERILL, M, HARPER, R & BOOTH, A 1999. A review of the use of health status measures in economic evaluation. *Health Technology Assessment,* 3.

BRAZIER, J, HARPER, R & THOMAS, K 1998. Deriving a preference-based single index measure from the SF-36. *Journal of Clinical Epidemiology,* 51**,** 1115-29.

BRAZIER, J, KOLOTKIN, R, CROSBY, R & WILLIAMS, G 2004. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value in Health,* 7**,** 490-498.

BRAZIER, J, RATCLIFFE, J, SALOMON, J & TSUCHIYA, A 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*, Oxford University Press.

BRAZIER, J, ROBERTS, J & DEVERILL, M 2002. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics,* 21.

BROD, M, TESLER, L & CHRISTENSEN, T 2009. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res,* 18**,** 1263-1278.

BUTLAND, B, JEBB, S, KOPELMAN, P, MCPHERSON, K, THOMAS, S, MARDELL, J & PARRY, V 2007. Foresight. Tackling Obesities: Future Choices – Project Report. London: Department of Innovation Universities and Skills: http://www.foresight.gov.uk/Obesity/Obesity_final/Index.html.

CARLTON, J 2011. Development of the Child Amblyopia Treatment Questionnaire (CAT-QoL): a disease-specific health related quality of life (HRQOL) measure for amblyopia. *HEDS Discussion paper* University of Sheffield.

CASTANEDA-GONZALEZ, L, CAMBEROS-SOLIS, R, BACARDI-GASCON, M & JIMENEZ-CRUZ, A 2010. Long-term randomized clinical trials of pharmacological treatment of obesity: systematic review. *Colombia Medica,* 41**,** 17**–**25.

CHOICEMETRICS 2012. Ngene. 1.1.1 ed.: http://www.choice-metrics.com/index.html.

COAST, J, FLYNN, T, NATARAJAN, L, SPROSTON, K, LEWIS, J, LOUVIERE, J & PETERS, T 2008. Valuing the ICECAP capability index for older people. *Social Science & Medicine,* 67**,** 874-82.

COLE, T, BELLIZZI, M, FLEGAL, K & DIETZ, W 2000. Establishing a standard definition for child overweight and obesity worldwide: International survey. *BMJ,* 320**,** 1240-3.

COLE, T, FLEGAL, K, NICHOLLS, D & JACKSON, A 2007. Body mass index cut-offs to define thinness in children and adolescents. *BMJ,* 335**,** 194-7.

COLE, T, FREEMAN, J & PREECE, M 1995. Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood,* 73**,** 25-29.

COLE, T, FREEMAN, J & PREECE, M 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine,* 17**,** 407-29.

CRAIG, B, BUSSCHBACH, J & SALOMON, J 2009. Keep it simple: ranking health states yields values similar to cardinal measurement approaches. *Journal of Clinical Epidemiology,* Mar 62**,** 296-305.

CROSS-GOVERNMENT-OBESITY-UNIT 2009. Healthy weight, healthy lives: child weight management programme and training providers framework. Department-of-Health-and-the-Department-for-Children-Schools-and-Families.

DCSF & NATIONAL-CENTRE-FOR-SOCIAL-RESEARCH 2009. Longitudinal study of young people in England: waves one to four, 2004–2007. Computer file - 7th Edn. SN: 5545 ed. Colchester, Essex UK Data Archive.

DE-BEKKER-GROB, E, RYAN, M & GERARD, K 2012. Discrete choice experiments in health economics: a review of the literature. *Health Economics,* 21**,** 145-172.

DE-ONIS, M, ONYANGO, A, BORGHI, E, SIYAM, A, NISHIDA, C & SIEKMANN, J 2007. Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization,* 85**,** 649-732.

DIETZ, W 1998. Health consequences of obesity in youth: Childhood predictors of adult disease. *Pediatrics,* 101**,** 518-525.

DINSDALE, H, RIDLER, C & ELLS, L 2011. A simple guide to classifying body mass index in children. *In:* NATIONAL-OBESITY-OBSERVATORY (ed.). Oxford.

DOLAN, P, CUDEX, C, KIND, P & WILLIAMS, A 1996. The time trade – off method: results from a general population study. *Health Economics,* 5**,** 141 - 154.

DOLAN, P & GUDEX, C 1995. Time preference, duration and health state valuations. *Health Economics,* 4**,** 289-99.

DOYLE, S 2011. Development of a new health related quality of life instrument for use in paediatric obesity. *In:* OBESITY-REVIEWS-CONFERENCE (ed.) *18th European Congress on Obesity (ECO).* Istanbul Turkey.

DRUMMOND, M, STODDART, G & TORRANCE, G 1996. *Methods for the Economic Evaluation of Health Care Programs,* Oxford, United Kingdom, Oxford University Press.

EUROQOL-GROUP 1990. EuroQol -- a new facility for the measurement of health-related quality of life. *Health Policy,* Dec 16**,** 199-208.

FALLON, E, TANOFSKY-KRAFF, M, NORMAN, A, MCDUFFIE, J, TAYLOR, E, COHEN, M, YOUNG-HYMAN, D, KEIL, M, KOLOTKIN, R & YANOVSKI, J 2005. Health-related quality of life in overweight and non-overweight black and white adolescents. *Journal of Pediatrics,* 147**,** 443-50.

FIEBIG, D, KEANE, M, LOUVIERE, J & WASI, N 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science,* 29**,** 393-421.

FITZPATRICK, R, DAVEY, C, BUXTON, M & JONES, D 1998. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment,* 2.

FLYNN, T, LOUVIERE, J, MARLEY, A, COAST, J & PETERS, T 2008. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Population Health Metrics,* 6.

FOOD-AND-DRUG-ADMINISTRATION. 2009. Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance. Available: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf.

FREEDMAN, D, KETTEL, L, SERDULA, M, DIETZ, W, SRINIVASAN, S & BERENSON, G 2005. The relation of childhood BMI to adult adiposity: the Bogalusa Heart Study. *Pediatrics***,** 22-27.

FURLONG, W, FEENY, D, TORRANCE, G & BARR, R 2001. The Health Utilities Index (HUI) System for Assessing Health-Related Quality of Life in Clinical Studies. *Annals of Medicine,* 33**,** 375-384.

GATELY, P, COOKE, C, BARTH, J, BEWICK, B, RADLEY, D & HILL, A 2005. Children's residential weight-loss programs can work: a prospective cohort study of short-term outcomes for overweight and obese children. *Pediatrics,* 116**,** 73-7.

GATELY, P, COOKE, C, BUTTERLY, R, MACKRETH, P & CARROLL, S 2000. The effects of a children's summer camp programme on weight loss, with a 10 month follow-up. *International journal of obesity and related metabolic disorders,* 24**,** 1445-52.

GOLDBERGER, A 1991. *A Course in Econometrics,* Cambridge MA, Harvard University Press.

GORSUCH, R 1983. *Factor analysis (2nd ed.)*, Hillsdale, NJ: Lawrence Erlbaum Associates.

GORTMAKER, S, SWINBURN, B, LEVY, D, CARTER, R, MABRY, P, FINEGOOD, D, HUANG, T, MARSH, T & MOODIE, M 2011. Changing the future of obesity: science, policy, and action Review Article. *The Lancet,* 378**,** 838-847.

GOTHWAL, V & BAGGA, D 2012. Andhra Pradesh Eye Disease Study - Visual Function Questionnaire: Further Improvements in Psychometric Properties using Rasch Analysis. *Ophthalmic Epidemiology,* 19**,** 306-16.

GREWAL, I, LEWIS, J, FLYNN, T, BROWN, J, BOND, J & COAST, J 2006. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science and Medicine,* 62**,** 1891-1901.

GRIFFITHS, L, PARSONS, T & HILL, A 2010. Self-esteem and Quality of Life in Obese Children and Adolescents: a Systematic Review. *International Journal of Paediatric Obesity,* 5**,** 282-304.

GUDEX, C, DOLAN, P, KIND, P & WILLIAMS, A 1996. Health state valuations from the general public using the visual analogue scale. *Quality of Life Research,* 5**,** 521-31.

GUJARATI, D 2003. *Basic Econometrics (4th) Edition,* New York, McGraw-HiII / lrwin.

HAKIM, Z & PATHAK, D 1999. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Econ,* 8**,** 103-116.

HAUBER, A, MOHAMED, A, JOHNSON, F, OYELOWO, O, CURTIS, B & COON, C 2010. Estimating importance weights for the IWQOL-Lite using conjoint analysis. *Quality of Life Research* 19**,** 701-9.

HEALTH-SURVEY-FOR-ENGLAND. 2008. *Table 3 - Body mass index (BMI), by survey year, age and sex* [Online]. Available: http://www.ic.nhs.uk/webfiles/publications/HSE07/ADULT%20TREND%20TABLES%202007.xls 2010].

HERDMAN, M, GUDEX, C, LLOYD, A, JANSSEN, M, KIND, P, PARKIN, D, BONSEL, G & BADIA, X 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res,* 20**,** 1727-1736.

HILL, A 2010. Psychosocial issues in obese children and adults. *In:* Crawford, Jeffery, Ball & Brug (eds.) *Obesity Epidemiology: From Aetiology to Public Health.* 2nd ed.: Oxford University Press.

HOBART, J & CANO, S 2009. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment.*

HOLE, A, KOLSTAD, J & HANSEN, D 2012. Inferred vs Stated Attribute Non-Attendance in Choice Experiments: A Study of Doctors' Prescription Behaviour. *Sheffield Economic Research Paper (SERP) Series.* Sheffield: SERP Number: 2012010.

HUSKISSON, E 1974. Measurement of pain. *Lancet ii***,** 1127-1131.

HUSSAIN, S & BLOOM, S 2011. The pharmacological treatment and management of obesity. *Postgraduate Medicine* 123**,** 34-44.

KHAMBALIA, A, HARDY, L & BAUMAN, A 2012. Accuracy of weight perception, life-style behaviours and psychological distress among overweight and obese adolescents. *Journal of Paediatrics and Child Health,* 48**,** 220-7.

KLARENBACH, S, PADWAL, R, WIEBE, N, HAZEL, M, BIRCH, D, MANNS, B, KARMALI, S, SHARMA, A & TONELLI, M 2010. Bariatric Surgery for Severe Obesity: Systematic Review and Economic Evaluation. *Canadian Agency for Drugs and Technologies in Health (Technology report; no. 129).* Ottawa.

KOLOTKIN, R & CROSBY, R 2002. Psychometric evaluation of the Impact of Weight on Quality of Life-Lite Questionnaire (IWQOL-Lite) in a community sample. *Quality of Life Research,* 11**,** 157-171.

KOLOTKIN, R, CROSBY, R, KOSLOSKI, K & WILLIAMS, G 2001. Development of a brief measure to assess quality of life in obesity. *Obesity Research,* 9**,** 102-111.

KOLOTKIN, R, HEAD, S, HAMILTON, M & TSE, C 1995. Assessing impact of weight on quality of life. *Obesity Research,* 3**,** 49-56.

KOLOTKIN, R, ZELLER, M, MODI, A, SAMSA, G, QUINLAN, N, YANOVSKI, J, BELL, S, MAAHS, D, DE-SERNA, D & ROEHRIG, H 2006. Assessing weight-related quality of life in adolescents. *Obesity,* 14**,** 448-57.

KRÄGELOH, C, KERSTEN, P, REX BILLINGTON, D, HSU, P, SHEPHERD, D, LANDON, J & FENG, X 2012. Validation of the WHOQOL-BREF quality of life questionnaire for general use in New Zealand: confirmatory factor analysis and Rasch analysis. *Quality of Life Research,* September 15.

LANCASTER, K 1966. A New Approach to Consumer Theory. Journal of Political Economy. *Journal of Political Economy,* 74**,** 132-157.

LANCSAR, E & LOUVIERE, J 2006. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics,* 15**,** 797-811.

LANCSAR, E & LOUVIERE, J 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics,* 26**,** 661-77.

LANDGRAF, J & ABETZ, L 1997. Functional status and well-being of children representing three cultural groups: initial self-reports using the CHQ-CF87. *Psychology & Health,* 12**,** 839-54.

LEHNERT, T, SONNTAG, D, KONNOPKA, A, RIEDEL-HELLER, S & KÖNIG, H 2012. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. *Obesity Reviews,* 13**,** 537-53.

LINACRE, J. 1994. Sample Size and Item Calibration (or Person Measure) Stability. *Rasch measurement and transactions* [Online], 7. Available: http://www.rasch.org/rmt/rmt74m.htm.

LINACRE, J 2002. Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement* 3**,** 85-106.

LOBSTEIN, T, BAUR, L & UAUY, R 2004. Lobstein, T., Baur, L., Uauy, R. (2004). Obesity in children and young people: A crisis in public. *Obesity Reviews,* 5 (Suppl 1)**,** 4-104.

MATZA, L, SWENSEN, A, FLOOD, E, SECNIK, K & LEIDY, N 2004. Assessment of Health Related Quality of Life in Children: A Review of Conceptual, Methodological, and Regulatory Issues. *Value in Health,* 7**,** 79-92.

MAYALL, B 2002. *Towards a sociology for childhood: thinking from children's lives,* Buckingham, Open University Press.

MCCABE, C, BRAZIER, J, GILKS, P, TSUCHIYA, A, ROBERTS, J, O'HAGAN, A & STEVENS, K 2006. Using rank data to estimate health state utility models. *Journal of Health Economics,* May 25**,** 418-31.

MCCABE, C, STEVENS, K, ROBERTS, J & BRAZIER, J 2005. Health State Values for the HUI2 descriptive system: results from a UK Survey. *Health Economics,* 14.

MCCOLL, E 2005. Developing questionnaires. *In:* FAYERS, P. & HAYS, R. (eds.) *Assessing Quality of Life in Clinical Trials.* 2 ed.: Oxford University Press.

MCFADDEN, D 1974. Conditional logit analysis of qualitative choice behavior. *In:* ZAREMBKA, P. (ed.) *Frontiers in Econometrics.* New York: Academic Press.

MICROSOFT 2011. Office - Word 2010 ed.: Microsoft

MILLER, G 1994. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review,* 101**,** 343-352.

MODI, A & ZELLER, M 2008. Validation of a Parent-proxy, Obesity-specific Quality-of-life Measure: Sizing Them Up. *Obesity,* 16**,** 2624-33.

MODI, A & ZELLER, M 2011. The IWQOL-Kids(Copyright): establishing minimal clinically important difference scores and test-retest reliability. *International Journal of Pediatric Obesity,* 6**,** e94-6.

MOKKINK, L, TERWEE, C, PATRICK, D, ALONSO, J, STRATFORD, P, KNOL, D, BOUTER, L & DE-VET, H 2010-a. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study *Quality of Life Research,* 19**,** 539-549.

MOKKINK, L, TERWEE, C, PATRICK, D, ALONSO, J, STRATFORD, P, KNOL, D, BOUTER, L & DE-VET, H 2010-b. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes *Journal of Clinical Epidemiology,* 63**,** 737-745.

MOOREHEAD, M, ARDELT-GATTINGER, E, LECHNER, H & ORIA, H 2003. The validation of the Moorehead-Ardelt Quality of Life Questionnaire II. *Obesity Surgery,* 13**,** 684-92.

MORALES, L, EDWARDS, T, FLORES, Y, BARR, L & PATRICK, D 2011. Measurement properties of a multicultural weight-specific quality-of-life instrument for children and adolescents. *Quality of Life Research,* 20**,** 215-24.

MÜLLER, R, CIEZA, A & GEYH, S 2012. Rasch analysis of the Hospital Anxiety and Depression Scale in spinal cord injury. *Rehabilitation Psychology* 57**,** 214-23.

MUST, A, DALLAL, G & DIETZ, W 1991. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht²) and triceps skinfold thickness. *American Journal of Clinical Nutrition* 53**,** 839-46.

NATIONAL-AUDIT-OFFICE-(NAO) 2001. Tackling Obesity in England. *In:* THE-STATIONERY-OFFICE (ed.). London.

NATIONAL-INSTITUTE-FOR-CARE-EXCELLENCE April 2013. Guide to the Methods of Technology Appraisal.

NATIONAL-INSTITUTE-FOR-CLINICAL-EXCELLENCE June 2006. Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children. *Clinical guidelines CG43.*

NATIONAL-INSTITUTE-FOR-HEALTH-AND-CLINICAL-EXCELLENCE-(NICE), CENTRE-FOR-CLINICAL-PRACTICE & CENTRE-FOR-PUBLIC-HEALTH-EXCELLENCE December 2011. Review of Clinical Guideline (CG43) – Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children

NATIONAL-OBESITY-OBSERVATORY-(NOO) 2009. Body Mass Index as a measure of obesity.

NATIONAL-OBESITY-OBSERVATORY-(NOO).
http://www.noo.org.uk/NOO_about_obesity/obesity_and_health [Online]. [2011-b].

NATIONAL-OBESITY-OBSERVATORY-(NOO). A simple guide to classifying body mass index in children. http://www.noo.org.uk/uploads/doc/vid_11601_A_simple_guide_to_classifying_BMI_in_children.pdf. [Accessed 2011-a].

NATIONAL-OBESITY-OBSERVATORY-(NOO). Statistics on Obesity, Physical Activity and Diet - England, 2013 report. http://www.noo.org.uk/news.php?nid=2244. [Accessed 2013].

NEWELL, A & SIMON, H 1972. *Human Problem Solving*, Englewood Cliffs NJ: Prentice Hall.

NORMAN, R, CRONIN, P & VINEY, R 2012. Deriving utility weights for the EQ-5D-5L using a discrete choice experiment. *CHERE Working Papers.*

OGDEN, C, CARROLL, M, KIT, B & FLEGAL, K 2012. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. *Journal of the American Medical Association,* 307**,** 483-490.

O'MEARA, S, RIEMSMA, R, SHIRRAN, L, MATHER, L & TER RIET, G 2002. The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment. *Health Technology Assessment,* 6**,** 1-97.

ORME, B. 1998. Sample Size Issues for Conjoint Analysis Studies. *Sawtooth Software Technical Papers* [Online].

OUDE-LUTTIKHUIS, H, BAUR, L, JANSEN, H, SHREWSBURY, V, O'MALLEY, C, STOLK, R & SUMMERBELL, C 2009. Interventions for treating obesity in children. *Cochrane database of systematic reviews,* 21.

OYEBODE, O & MINDELL, J March 2013. Use of data from the Health Survey for England in obesity policy making and monitoring. *Obesity Reviews*.

PADWAL, R, KLARENBACH, S, WIEBE, N, HAZEL, M, BIRCH, D, KARMALI, S, SHARMA, A, MANNS, B & TONELLI, M 2011. Bariatric surgery: a systematic review of the clinical and economic evidence. *Journal of General Internal Medicine* 26**,** 1183-1194.

PAN, H & COLE, T 2011. LMSgrowth, a Microsoft Excel add-in to access growth references based on the LMS method. Version 2.76. http://www.healthforallchildren.co.uk/.

PATRICK, D, SKALICKY, A, EDWARDS, T, KUNIYUKI, A, MORALES, L, LENG, M & KIRSCHENBAUM, D 2011. Weight loss and changes in generic and weight-specific quality of life in obese adolescents. *Quality of Life Research,* 20**,** 961-8.

PETROU, S 2003. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Economics,* 12**,** 697-702.

PICOT, J, JONES, J, COLQUITT, J, LOVEMAN, E & CLEGG, A 2012. Weight loss surgery for mild to moderate obesity: a systematic review and economic evaluation. *Obesity Surgery,* 22**,** 1496-506.

QSR-INTERNATIONAL 2011. NVIVO9 ed.

RASCH, G 1960. *Probabilistic models for some intelligence and attainment tests,* Chicago, University of Chicago Press. [Reprinted 1980].

RATCLIFFE, J, BRAZIER, J, TSUCHIYA, A, SYMONDS, T & BROWN, M 2009. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Economics,* 18**,** 1261-76.

RATCLIFFE, J, COUZNER, L, FLYNN, T, SAWYER, M, STEVENS, K, BRAZIER, J & BURGESS, L 2011. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare Best-Worst Discrete Choice Experiment, Standard Gamble and Time Trade Off methods. *Applied Health Economics and Health Policy,* 9**,** 15-27.

RAVENS-SIEBERER, U & BULLINGER, M 1998. Assessing health related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Quality of Life Research,* 7**,** 399-407.

RAVENS-SIEBERER, U, REDEGELD, M & BULLINGER, M 2001. Quality of life after in-patient rehabilitation in children with obesity. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity,* 25 Suppl 1**,** S63-5.

RITCHIE, J, SPENCER, L & O'CONNOR, O 2003. Carrying out qualitative analysis *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative research practice: a guide for social science students and researchers.* London: Sage.

ROUX, L, UBACH, C, DONALDSON, C & RYAN, M 2004. Valuing the benefits of weight loss programs: an application of the discrete choice experiment. *Obesity research* 12**,** 1342-51.

ROWLINSON, M. 2011. *Obesity stigma in young children.* D.Clin.Psychol, University of Leeds.

RYAN, M, BATE, A, EASTMOND, C & LUDBROOK, A 2001. Use of discrete choice experiments to elicit preferences. *Quality in Health Care,* Suppl 1.

RYAN, M & GERARD, K 2003. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy,* 2**,** 55-64.

RYAN, M, NETTEN, N, SKATUN, D & SMITH, P 2006. Using discrete choice experiments to estimate a preference-based measure of outcome--An application to social care for older people. *Journal of Health Economics,* 25**,** 927-944.

SALEM, L, JENSEN, C & FLUM, D 2005. Are bariatric surgical outcomes worth their cost? A systematic review. *Journal of the American College of Surgeons,* 200**,** 270-278.

SCHWIMMER, J, BURWINKLE, T & VARNI, J 2003. (2003). Health related quality of life of severely obese children and adolescents. *Journal of the American Medical Association,* 289**,** 1813 - 1819.

SIM, J, BARTLAM, B & BERNARD, M 2011. The CASP-19 as a measure of quality of life in old age: evaluation of its use in a retirement community. *Quality of Life Research,* 20**,** 997-1004.

SPSS-INC. Released 2009. PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc.

STANIFORD, L, BRECKON, J & COPELAND, R 2012. Treatment of childhood obesity: a systematic review. *Journal of Child and Family Studies,* 21**,** 545-564.

STARFIELD, B, BERGNER, M, ENSMINGER, M, RILEY, A, RYAN, S, GREEN, B, MCGAUHEY, P, SKINNER, A & KIM, S 1993. Adolescent health status measurement: development of the Child Health and Illness Profile. *Paediatrics,* 91**,** 430-5.

STEVENS, K 2009. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research,* 18**,** 1105-1113.

STEVENS, K 2010. Working with children to develop dimensions for a preference-based, generic, paediatric health-related quality-of-life measure. *Qualitative Health Research,* 20**,** 340-51.

STEVENS, K 2012. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics,* 30**,** 729-747.

STEVENS, K & PALFREYMAN, S 2012. 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 in Health,* 15**,** 991-8.

THE-HEALTH-AND-SOCIAL-CARE-INFORMATION-CENTRE 2009. National Child Measurement Programme: England, 2008/09 school year. Department-of-Health.

THE-KIDSCREEN-GROUP-EUROPE 2006. *The-KIDSCREEN-Questionnaires: Quality of life questionnaires for children and adolescents*, Handbook Lengerich, Pabst Science Publishers.

TORRANCE, G, THOMAS, W & SACKETT, D 1972. A utility maximization model for evaluation of health care programs. *Health Services Research,* 7**,** 118-33.

TSIROS, M, OLDS, T, BUCKLEY, J, GRIMSHAW, P, BRENNAN, L, WALKLEY, J, HILLS, A, HOWE, P & COATS, A 2009. Health-related quality of life in obese children and adolescents. *International Journal of Obesity,* 33.

VARNI, J, SEID, M & KURTIN, P 2001. PedsQL 4.0: reliability and validity of the Paediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Medical Care,* 39**,** 800-12.

VINEY, R, SAVAGE, E, KING, M & HOSSAIN, I 2007. Using Choice Experiments to Estimate QALYs: An Application to Prostate Cancer. *iHEA 2007 6th World Congress: Explorations in Health Economics Paper* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=994063.

VON-NEUMANN, J & MORGENSTERN, O 1947. *Theory of games and economic behaviour (2nd ed),* London, Princeton University Press.

WANG, Y, MCPHERSON, K, MARSH, T, GORTMAKER, S & BROWN, M 2011. Health and economic burden of the projected obesity trends in the USA and the UK Review Article *The Lancet,* 378**,** 815-825.

WARE, J & SHERBOURNE, C 1992. The MOS 36-item short form health survey (SF-36): conceptual framework and items selection. *Medical Care,* 30**,** 473-83.

WILLE, N, BADIA, X, BONSEL, G, BURSTRÖM, K, CAVRINI, G, DEVLIN, N, EGMAR, A, GREINER, W, GUSI, N, HERDMAN, M, JELSMA, J, KIND, P, SCALONE, L & RAVENS-SIEBERER, U 2010. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. . *Quality of Life Research,* 19**,** 875-86.

WORLD-HEALTH-ORGANISATION-(WHO). *Global Infobase: data on overweight and obesity, mean BMI, healthy diets and physical inactivity. Prevalence of overweight & obesity map. https://apps.who.int/infobase/* [Online]. 2013-b].

WORLD-HEALTH-ORGANISATION-(WHO). *http://www.who.int/nutrition/topics/obesity/en/* [Online]. 2012].

WORLD-HEALTH-ORGANISATION-(WHO) 2006. Multicentre Growth Reference Study Group. WHO Child Growth Standards. *Acta Paediatrica,* 450 Supplement

WORLD-HEALTH-ORGANISATION-(WHO). May 2012. *Obesity and overweight - Fact sheet No.311. http://www.who.int/mediacentre/factsheets/fs311/en/* [Online]. 2013-a].

WORLD-HEALTH-ORGANIZATION 1948. Constitution of the World Health Organization. Geneva: Basic Documents. WHO.

YANG, Y, BRAZIER, JE., TSUCHIYA, A. AND YOUNG, TA. 2011. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. *Medical Decision Making,* 31**,** 281- 291.

YANG, Y, BRAZIER, J, TSUCHIYA, A & COYNE, K 2009. Estimating a preference-based single index from the overactive bladder questionnaire. *Value in Health,* 12**,** 159-166.

YOUNG, T, ROWEN, D, NORQUIST, J & BRAZIER, J 2010. Developing Preference-Based Health Measures: Using Rasch Analysis to Generate Health State Values. *Quality of Life Research,* 19**,** 907-17.

YOUNG, T, YANG, Y, BRAZIER, J & TSUCHIYA, A 2011. The use of rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Medical Decision Making,* 31**,** 195-210.

YOUNG, T, YANG, Y, BRAZIER, J, TSUCHIYA, A & COYNE, K 2009. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Quality of Life Research,* 18**,** 253-65.

ZELLER, M & MODI, A 2009. Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up. *Obesity,* Jun 17**,** 1171-7.

Appendices

Chapter 3 - Literature review

3.1 Literature review A - Example search strategy

Ovid MEDLINE(R) <1950 to November Week 3 2009>

1	adolescen*.tw. (121588)
2	youth*.tw. (25411)
3	child*.tw. (778435)
4	young person*.tw. (2117)
5	young people.tw. (10867)
6	teen*.tw. (16635)
7	(young adj (adult* or people* or person* or girl* or boy*)).tw. (55920)
8	(pediatric or paediatric or pediatrics or paediatrics).tw. (144734)
9	or/1-8 (967281)
10	(obese or obesity).tw. (108059)
11	bmi.tw. (39581)
12	body mass index.tw. (59341)
13	overweight*.tw. (21027)
14	body fat.tw. (15324)
15	or/10-14 (168665)
16	(qol$ or pqol or hrqol or qaly$ or qls).tw. (16937)
17	"health-related quality of life".tw. (12094)
18	"pediatric quality of life inventory".tw. (172)
19	"weight related quality of life".tw. (16)
20	(sf-36 or "short form 36").tw. (9649)
21	("quality of life" adj5 (measure* or scale* or rating* or value*)).tw. (11297)
22	"Adolescent Psychology".tw. (53)
23	"Child Psychology".tw. (212)
24	"Self Concept".tw. (2761)
25	PedsQL.tw. (222)
26	"IWQOL-Lite".tw. (25)
27	"Impact of weight on quality of life-lite".tw. (25)
28	Happiness.tw. (2320)
29	(life adj1 satisf*).tw. (2810)
30	"Health State".tw. (1642)
31	"Activities of Daily Living".tw. (10207)
32	("Reference Values" adj5 (quality or qol$ or pqol or hrqol or qaly$ or qls)).tw. (39)
33	(("well being" or wellbeing) adj5 (obesity or obese or bmi or "body mass index" or overweight* or "body fat")).tw. (110)
34	or/16-33 (54070)
35	9 and 15 and 34 (281)
36	adolescent/ or child/ (1926558)
37	exp Obesity/ (100259)
38	Overweight/ (4559)
39	exp body fat distribution/ or body mass index/ or skinfold thickness/ or waist-hip ratio/ (56885)
40	37 or 38 or 39 (137773)
41	"Quality of Life"/ (82243)
42	quality-adjusted life years/ (4287)
43	41 or 42 (85815)
44	36 and 40 and 43 (354)
45	35 or 44 (560)
46	from 45 keep 1-560 (560)

3.2 Literature review B - COSMIN taxonomy of definitions*

Measurement property (Box from the COSMIN checklist)	Definition
Internal consistency (box A)	The degree of the interrelatedness among the items
Reliability (box B)	The proportion of the total variance in the measurements which is because of ''true'' differences among patients
Measurement error (box C)	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Content validity (box D)	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
Structural validity (box E)	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
Hypotheses testing (box F)	Idem construct validity - The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
Cross-cultural validity (box G)	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
criterion validity (box H)	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a ''gold standard''
Responsiveness (box I)	Idem responsiveness - The ability of an HR-PRO instrument to detect change over time in the construct to be measured

*From Mokkink *et al*., 2010b, *p.* 743 (Table 2); HR-PRO = Health Related Patient Reported Outcomes

3.3 Literature review A - Summary of search findings

Database	Dates covered	Hits
MEDLINE (OVIDSp)	1950 to November Week 3 2009	560
Medline in Process (OvidSp)	January 14, 2010	19
EMBASE Classic+EMBASE (OVIDSp)	1947 to 2009 December 18	462
AMED Allied and Complementary Medicine	1985 to December 2009	32
CINAHL	Inception to January 2010	984
PsycINFO	1806 to December Week 4 2009	217
Cochrane Library – Methods studies (WILEY)	Inception to December 2009	261

3.4 Literature review B - Summary of search findings

Database	Dates covered	Hits
MEDLINE & Medline in Process (OVIDSp)	1946 to January Week 2 2012	1021
EMBASE Classic+EMBASE (OVIDSp)	1947 to 2012 January 20	1464
AMED Allied and Complementary Medicine	1985 to January 2012	42
CINAHL	Inception to January 24, 2012	76
SPORTDiscus	Inception to January 24, 2012	166

3.5 Literature review A - Identified studies

Reference (n=34)	Instruments identified (N=11)
1. ABRESCH, R. T., MCDONALD, D. A., WIDMAN, L. M., MCGINNIS, K. & HICKEY, K. J. 2007. Impact of spinal cord dysfunction and obesity on the health-related quality of life of children and adolescents. *Journal of Spinal Cord Medicine,* 30 Suppl 1, S112-8.	PedsQL
2. ARIF, A. A. & ROHRER, J. E. 2006. The relationship between obesity, hyperglycemia symptoms, and health-related quality of life among Hispanic and non-Hispanic white children and adolescents. *BMC Family Practice,* 7, 3.	KINDL
3. BRANCATISANO, A., WAHLROOS, S. & BRANCATISANO, R. 2008. Improvement in comorbid illness after placement of the Swedish Adjustable Gastric Band. *Surgery for Obesity and Related Diseases,* 4, S39-S46.	SF-36
4. BRAZIER, J. E., KOLOTKIN, R. L., CROSBY, R. D. & WILLIAMS, G. R. 2004. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value in Health,* 7, 490-8.	SF-36 IWQOL-Lite
5. COLLINS, J., MATTAR, S., QURESHI, F., WARMAN, J., RAMANATHAN, R., SCHAUER, P. & EID, G. 2007. Initial outcomes of laparoscopic Roux-en-Y gastric bypass in morbidly obese adolescents. *Surgery for Obesity & Related Diseases,* 3, 147-52.	Moorehead-Aldelt QoL Questionnaire
6. DE BEER, M., HOFSTEENGE, G. H., KOOT, H. M., HIRASING, R. A., DELEMARRE-VAN DE WAAL, H. A. & GEMKE, R. J. B. J. 2007. Health-related-quality-of-life in obese adolescents is decreased and inversely related to BMI. *Acta Paediatrica,* 96, 710-4.	CHQ PedsQL
7. DE ZWAAN, M., LANCASTER, K. L., MITCHELL, J. E., HOWELL, L. M., MONSON, N., ROERIG, J. L. & CROSBY, R. D. 2002. Health-related quality of life in morbidly obese patients: Effect of gastric bypass surgery. *Obesity Surgery,* 12, 773-780.	SF-36
8. DOLL, H. A., PETERSEN, S. E. & STEWART-BROWN, S. L. 2000. Obesity and physical and emotional well-being: associations between body mass index, chronic illness, and the physical and mental components of the SF-36 questionnaire. *Obesity Research,* 8, 160-70.	SF-36
9. FLODMARK, C. E. 2005. The happy obese child. *International Journal of Obesity,* 29 Suppl 2, S31-3.	KINDL PesdQL
10. FONTAINE, K. R., CHESKIN, L. J. & BAROFSKY, I. 1996. Health-related quality of life in obese persons seeking treatment. *Journal of Family Practice,* 43, 265-70.	SF-36
11. *FULLERTON, G., TYLER, C., JOHNSTON, C. A., VINCENT, J. P., HARRIS, G. E. & FOREYT, J. P. 2007. Quality of life in Mexican-American children following a weight management program. *Obesity,* 15, 2553-6. (DUPLICATED)	PedsQL
12. GARCIA-MORALES, L. M., BERBER, A., MACIAS-LARA, C. C., LUCIO-ORTIZ, C., DEL-RIO-NAVARRO, B. E. & DORANTES-ALVAREZ, L. M. 2006. Use of sibutramine in obese mexican adolescents: a 6-month, randomized, double-blind, placebo-controlled, parallel-group trial. *Clinical Therapeutics,* 28, 770-82.	SF-36
13. KOLOTKIN, R. L., ZELLER, M., MODI, A. C., SAMSA, G. P., QUINLAN, N. P., YANOVSKI, J. A., BELL, S. K., MAAHS, D. M., DE SERNA, D. G. & ROEHRIG, H. R. 2006. Assessing weight-related quality of life in adolescents. *Obesity,* 14, 448-57.	IWQOL
14. LARSSON, U., KARLSSON, J. & SULLIVAN, M. 2002. Impact of overweight and obesity on health-related quality of life--a Swedish population study. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity,* 26, 417-24.	SF-36
15. LOFRANO-PRADO, M. C., ANTUNES, H. K. M., DO PRADO, W. L., DE PIANO, A., CARANTI, D. A., TOCK, L., CARNIER, J., TUFIK, S., DE MELLO, M. T. & DAMASO, A. R. 2009. Quality of life in Brazilian obese adolescents: effects of a long-term multidisciplinary lifestyle therapy. *Health & Quality of Life Outcomes,* 7, 61.	SF-36
16. LOUX, T. J., HARICHARAN, R. N., CLEMENTS, R. H., KOLOTKIN, R. L., BLEDSOE, S. E., HAYNES, B., LEATH, T. & HARMON, C. M. 2008. Health-related quality of life before and after bariatric surgery in adolescents. *Journal of Pediatric Surgery,* 43, 1275-9.	SF-36
17. MCAULEY, K. A., TAYLOR, R. W., FARMER, V. L., HANSEN, P., WILLIAMS, S. M., BOOKER, C. S. & MANN, J. I. 2010. Economic evaluation of a community-based obesity prevention program in children: the APPLE project. *Obesity,* 18, 131-6.	HUI
18. MODI, A. C., LOUX, T. J., BELL, S. K., HARMON, C. M., INGE, T. H. & ZELLER, M. H. 2008. Weight-specific health-related quality of life in adolescents with extreme obesity. *Obesity,* 16, 2266-71.	PedsQL IWQOL
19. MODI, A. C. & ZELLER, M. H. 2008. Validation of a parent-proxy, obesity-specific quality-of-life measure: sizing them up. *Obesity,* 16, 2624-33.	PedsQL Sizing them up IWQOL-Kids
20. PINHAS-HAMIEL, O., SINGER, S., PILPEL, N., FRADKIN, A., MODAN, D. & REICHMAN, B. 2006.	PedsQL

	Health-related quality of life among children and adolescents: associations with obesity. *International Journal of Obesity,* 30, 267-72.	
21.	RAVENS-SIEBERER, U., REDEGELD, M. & BULLINGER, M. 2001. Quality of life after in-patient rehabilitation in children with obesity. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity,* 25 Suppl 1, S63-5.	KINDL
22.	SILBERHUMER, G. R., MILLER, K., KRIWANEK, S., WIDHALM, K., PUMP, A. & PRAGER, G. 2006. Laparoscopic adjustable gastric banding in adolescents: the Austrian experience. *Obesity Surgery,* 16, 1062-7.	Moorehead-Aldelt QoL Q
23.	TYLER, C., JOHNSTON, C. A., FULLERTON, G. & FOREYT, J. P. 2007. Reduced quality of life in very overweight Mexican American adolescents. *Journal of Adolescent Health,* 40, 366-8.	PedsQL
24.	VARNI, J. W., LIMBERS, C. A. & BURWINKLE, T. M. 2007. Impaired health-related quality of life in children and adolescents with chronic conditions: a comparative analysis of 10 disease clusters and 33 disease categories/severities utilizing the PedsQL 4.0 Generic Core Scales. *Health & Quality of Life Outcomes,* 5, 43.	PedsQL
25.	VISSERS, D., DEVOOGDT, N., GEBRUERS, N., MERTENS, I., TRUIJEN, S. & VAN GAAL, L. 2008. Overweight in adolescents: differences per type of education. Does one size fit all? *Journal of Nutrition Education & Behavior,* 40, 65-71.	SF-36
26.	WILLE, N., ERHART, M., PETERSEN, C. & RAVENS-SIEBERER, U. 2008. The impact of overweight and obesity on health-related quality of life in childhood--results from an intervention study. *BMC Public Health,* 8, 421.	KINDL-Obesity
27.	WILLIAMS, J., WAKE, M., HESKETH, K., MAHER, E. & WATERS, E. 2005. Health-related quality of life of overweight and obese children. *JAMA,* 293, 70-6.	PedsQL
28.	YACKOBOVITCH-GAVAN, M., NAGELBERG, N., DEMOL, S., PHILLIP, M. & SHALITIN, S. 2008. Influence of weight-loss diets with different macronutrient compositions on health-related quality of life in obese youth. *Appetite,* 51, 697-703.	PedsQL
29.	YACKOBOVITCH-GAVAN, M., NAGELBERG, N., PHILLIP, M., ASHKENAZI-HOFFNUNG, L., HERSHKOVITZ, E. & SHALITIN, S. 2009. The influence of diet and/or exercise and parental compliance on health-related quality of life in obese children. *Nutrition Research,* 29, 397-404.	PedsQL
30.	ZABELINA, D. L., ERICKSON, A. L., KOLOTKIN, R. L. & CROSBY, R. D. 2009. The effect of age on weight-related quality of life in overweight and obese individuals. *Obesity,* 17, 1410-3.	IWQOL
31.	ZELLER, M. H. & MODI, A. C. 2006. Predictors of health-related quality of life in obese youth. *Obesity,* 14, 122-30.	PedsQL
32.	ZELLER, M. H. & MODI, A. C. 2009. Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up. *Obesity,* 17, 1171-7.	PedsQL Sizing Me Up Sizing Them Up
33.	ZELLER, M. H., MODI, A. C., NOLL, J. G., LONG, J. D. & INGE, T. H. 2009. Psychosocial functioning improves following adolescent bariatric surgery. *Obesity,* 17, 985-90.	PedsQL
34.	ZELLER, M. H., ROEHRIG, H. R., MODI, A. C., DANIELS, S. R. & INGE, T. H. 2006. Health-related quality of life and depressive symptoms in adolescents with extreme obesity presenting for bariatric surgery. *Pediatrics,* 117, 1155-61.	PedsQL

*Duplicated reference

Chapter 5 - Creation of the weight specific LLDS

5.1 Information sheet and consent - adolescents

YOUNG PERSON'S INFORMATION SHEET

AND CONSENT FORM

HOW DOES WEIGHT AFFECT YOUNG PEOPLE?

We would like you to take part in a research study to look at how weight can affect young people and the things they like to do. Before you decide if you want to take part it's important to know why we are doing the study. And also what you will have to do. So please think about the information below carefully. You can talk about it with your family or friends if you want to.

PART 1 – What is the study about and what will happen?

Why are we doing this research?

We would like to know more about how weight affects young people. Mainly, how weight may affect health and other things in life. Such as: family life, school and social life.

Why have I been invited to take part?

You have been chosen because you are attending the WATCH IT programme. We are hoping to speak to a number of young people

from the programme over the next few months so they can talk to us about how weight affects their life.

Do I have to take part?

No, it is up to you if you take part or not. Even if you say yes at the start, you can stop taking part at any time during the study. You do not have to give a reason.

What will happen to me if I take part?

If you would like to take part we will ask you to sign a form saying you agree to take part. This is the form on the last page. Then the researcher will arrange a time to speak with you. At this time the researcher will ask you questions about how your weight affects your health and other things in your life. This should last around 30-40 minutes.

Is anything bad likely to happen to me if I take part?

No, we don't think so. You might sometimes find it upsetting to talk about how your weight has affected you. You can stop talking at any time you like.

You don't have to answer all the questions and can decide to skip any you don't want to answer.

Why should I be in it?

You will help us know how weight can affect young people. This will help in offering better services to young people who need a hand with managing their weight.

What will happen to the information I give?

In some cases we may have to tell others what you have said but we would never do this without talking to you about this first. No one outside the study team will be shown any personal information about you.

We will put the information from all the young people together in a report. In this report we may use some quotes of the things you say but nobody will be able to tell whom the information is from or about.

What are the likely benefits of taking part?

The main benefit from taking part in the study is that you will help us learn more about how weight affects young people. You might not directly benefit from taking part, but you might find it helpful to talk about how weight has affected you.

Thank you for reading this.

If you want more information please read part2.

PART 2: Additional information about the study

Who is managing and paying for the study?

The study is part of a PhD being paid for by the National Institute for Health Research (NIHR).

Who has checked the study?

Before any study goes ahead it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. This project has been checked by the Leeds University Research Ethics Committee.

Contact details

If you have any questions or would like more information you can speak to the researchers involved in the study. Contact details are:

Yemi Oluboyede
NIHR Doctoral Research fellow
University of Leeds
Charles Thackrah Building
101 Clarendon Road
Leeds LS2 9LJ
Tel: +44 (0)113 343 0823
E mail: y.oluboyede@leeds.ac.uk

Professor Christopher McCabe

or

Professor Andrew Hill
Tel: +44 (0) 113 343 6966
E mail: hssjad@leeds.ac.uk

Thank you for reading this – please ask any questions if you need to.

CONSENT

FORM

Please circle '**yes**' or '**no**' for each of these questions:

Have you read (or had information read to you) about this project?	**Yes**	**No**
Has somebody explained this project to you?	**Yes**	**No**
Do you understand what this project is about?	**Yes**	**No**
Have you asked all the questions you want?	**Yes**	**No**
Have you had your questions answered in a way you understand?	**Yes**	**No**
Do you understand it's OK to stop taking part at any time?	**Yes**	**No**
Are you happy to be contacted again about this project?	**Yes**	**No**
Are you happy to take part?	**Yes**	**No**

Only sign your name if you want to take part!

If you DO want to take part, please write and sign your name below:

Print Your Name _____

Your Signature _____

Today's Date _____

The person who explained this study to you needs to sign too:

Print Name & signature: _____

Today's Date: _____

Thank you for your help

5.2 Opt in/opt out – parents and carers

OPT-OUT

FORM

Please fill this form in if you DO NOT want your child to take part in the study

How does weight affect young people?

I do not want my child to take part in the above named study.

Your name/signature: _____

Print Name of Child: _____

Today's Date: _____

(Please either give this completed form back to a member of the WATCH IT team or send it to us using the stamped addressed envelope provided)

CONSENT

FORM

How does weight affect young people?

		Please initial the boxes
1.	I confirm that I have read and understand the Information Sheet for this study and have had the opportunity to ask questions.	
2.	I understand that my child's participation is voluntary and that I and/or my son/daughter are free to withdraw at any time without giving any reason and without our legal rights being affected.	
3.	I understand that no identifiable information from the data collected from me and my son / daughter will be published in any papers sent to research journals.	
4.	I understand that if we withdraw from this study, the data collected from me and my son / daughter will not be used in analysing the results of the study.	
5.	I understand that a copy of this consent form will be sent to the Research Office (the University of Leeds)	
6.	I agree to my child taking part in this study.	

PARTICIPANT:

Name of participant (IN CAPITALS)	Date	Signature

RESEARCHER:

Name of Researcher	Date	Signature

(Completed form to be given to XXX from the school)

5.3 Topic guide

Introduction

Aim: To introduce the research and set the context for the discussion to follow.

- Introduce self
- Introduce the study & Purpose of the interview
    - Length of the interview
    - Voluntary nature of participation & right to withdraw
    - Confidentiality and how findings will be reported
    - Being contacted at a later date once the draft questionnaire is completed
    - Recording of interview
- Gift-in-kind to be received at the end of the interview
- Any questions they have

TURN ON DIGITAL RECORDER

A. Background and personal circumstances

Aim: To ease respondent into the interview and to give some context to their life that might influence their views on the effect of weight on their lives

- Tell me about <u>yourself</u>, <u>family</u> and <u>area living in</u>:
  - Household circumstances (who they live with, their age)
  - Main daytime activity (self & parents)
  - Area they live in (rural, built up, parks, shops activity services)
- Comparing yourself to others your age, where do you fall size wise:
  - *average weight, somewhat overweight or very overweight*
  - *Sample Probe: Can you tell me a little bit about why you think you are XXXX?*
  - *How about others in your family: mother, father, brothers, sisters?*

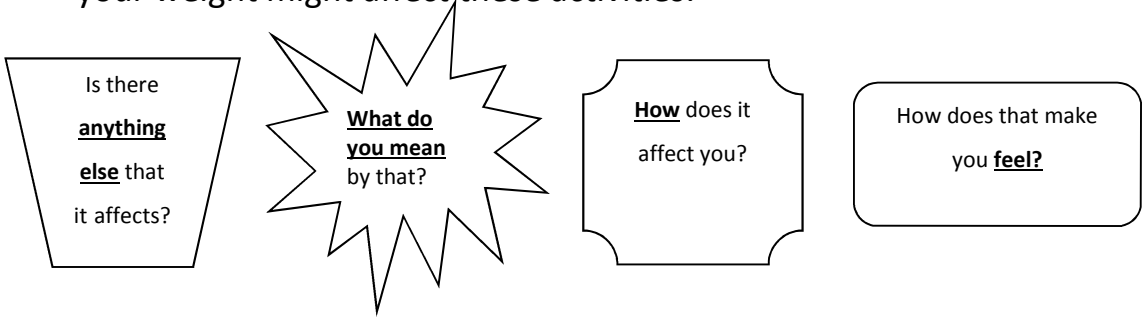
B. Dimensions of QoL affected by weight – from own life

Aim: To identify dimensions of QoL affected by weight status using a typical day in the lives of adolescents

- Can you describe the types of things you do on a typical day:

School or work / *During the week*	Non school or workdays / *weekend & holidays (think back to last weekend if that helps)*
o *Morning*	o *Morning*
o *Daytime*	o *Daytime*
o *Evening*	o *Evening*

- Thinking back to the activities you just spoke about, tell me about your weight might affect these activities:

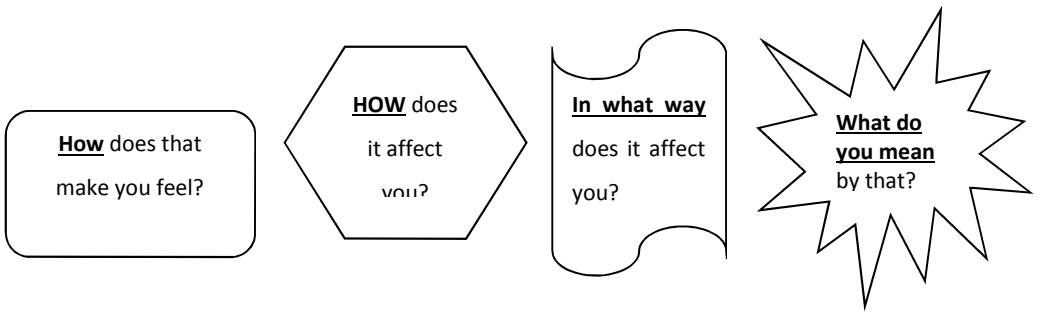


C. Dimensions of QoL that obesity affects – from the Literature

Aim: To explore the respondent's experiences of how weight affects aspects of their lives informed by the literature

"Other young people have said that their weight affects them in other ways too. How do you think weight might affect other aspects, such as"……..

- **Physical activity** (*moving around/taking part in sports/ sports/walking/climbing stairs/bending over)*
- **School** *(attending/joining in activities & lessons)*
- **Psychological health** *(feelings/happy/sad/frustrated/angry/anxious/tense)*
- **Body esteem** - The way you see yourself *(attractiveness/mirror/clothes/dressed /undressed)*
- **Relationships** *(friends/family)*
- **Social functioning** *(social life/going out to places/way treated by people)*
- **Eating** *(control/feelings when eating)*
- **Future** – (e.g. *'limit things I would like to do in the future'*)



'Thank you for sharing these experiences with me. You have given me very valuable information:'

- ***Are there other aspects of your life that weight might affect that we have not already talked about?***

- ***Do you have any questions you would like to ask me?***

TURN OFF DIGITAL RECORDER after ranking is completed

Thank you very much in participating in this study.

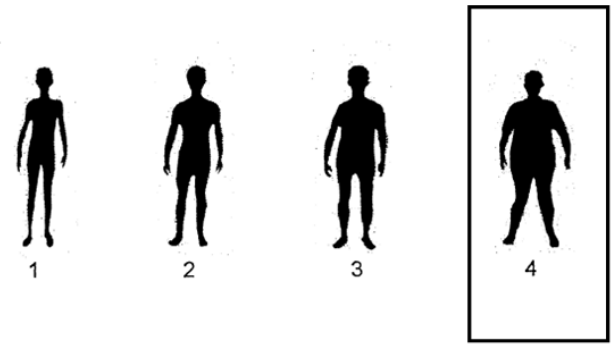
Think:

End the interview in a nice way to get participants thinking positively about themselves again.

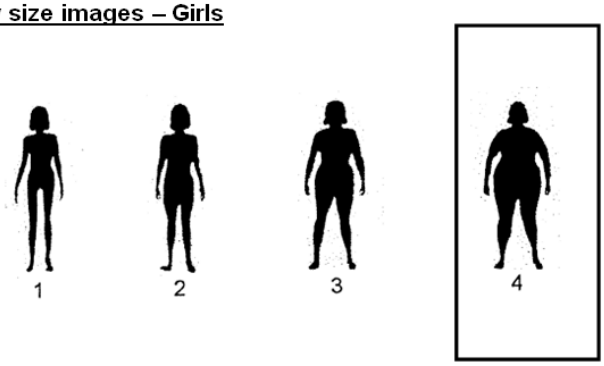
e.g. Some general chat about what they're doing today, something about their hobbies from the interview.

5.4 Body size silhouettes

Body size images – Boys



Body size images – Girls



5.5 Saturation matrix

Section 1 - Themes identified from the experiences of adolescents

Interview no.			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Physical	**Daily routine**	(moving around/walking/climbing stairs/bending over/dressing/washing)													1	1	1		1
	Physical activities	(limitations in: running/taking part in sports)		1	1	1	1	1	1	1	1		1	1	1	1	1	1	1
Psychological well-being	**Feelings**	(feeling inside: unhappy/sad/frustrated/angry/anxious)			1	1	1	1	1		1	1		1	1	1	1	1	1
	Body image / esteem	(attractiveness/mirror/clothes/changing in front of others)					1	1	1		1			1		1	1	1	1
Social	**Social functioning**	(social life/going out to places/way treated by people outside of school)	1				1	1	1		1				1				1
	Relationships	(With: friends/family)						1		1								1	1
	School	(attending/joining in activities & lessons/way treated by people in school)		1	1				1					1					1
Eating		(control/feelings when eating)				1		1	1	1	1	1	1	1	1	1	1	1	1
Future		(e.g. 'limit things I would like to do in the future')																	1

Section 2 - Themes driven by the literature

Order no.			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Physical	Daily routine	(moving around/walking/climbing stairs/bending over/dressing/washing)																	
	Physical activities	(limitations in: running/taking part in sports)				1		1	1		1				1			1	
Psychological well-being	Feelings	(feeling inside: unhappy/sad/frustrated/angry/anxious)	1	1	1	1		1	1	1	1			1	1		1		
	Body image / esteem	(attractiveness/mirror/clothes/changing in front of others)	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	
Social	Social functioning	(social life/going out to places/way treated by people outside of school)	1	1	1	1		1	1	1	1		1	1	1				
	Relationships	(With: friends/family)			1			1	1	1	1			1	1				1
	School	(attending/joining in activities & lessons/way treated by people in school)				1	1			1	1	1	1	1			1		
Eating		(control/feelings when eating)		1	1		1	1			1	1		1	1	1	1	1	
Future		(e.g. 'limit things I would like to do in the future')	1	1	1	1	1							1	1	1		1	

5.6 Advisory Group

The members of the advisory panel included the thesis supervisors and individuals who provided guidance and specialist support at specific phases of the thesis:

Dr Cathy Brennan: Provided expertise and support in terms of the qualitative field work and analysis. Her interests are in qualitative methodologies and the development of methods to engage individuals for whom verbal and/or written communication may be problematic.

Professor Jenny Hewison: With a vast amount of expertise and experience in applied health research and the application of a wide range of methodologies, Jenny provided guidance on the overall methodological demands of the empirical studies in thesis.

Prof Donna Lamping: As a senior psychometrician, Donna was able to provide specialist support in shaping both the content of the descriptive system and addressing the key psychometric issues.

David Meads, Dr Adam Smith & Prof Alan Tennant: Provided guidance on the application of modern psychometric methods such as Rasch and Item-response models in the creation of a new QoL instrument.

Dr Katharine Stevens: Katherine had recent firsthand experience of developing, a new instrument for economic evaluation for the younger population (in her case for 7-11 year olds), the Child Health Utility – Nine dimensions (CHU9D). She was able to provide close guidance on all the key phases the thesis, especially in the design and analysis of the qualitative work.

Thesis supervisors

Current:

- Prof Claire Hulme – University of Leeds supervisor
- Prof Andrew Hill – University of Leeds supervisor
- Prof Aki Tsuchiya – University of Sheffield supervisor

Previous:

- Prof Christopher McCabe – University of Leeds supervisor now at the University of Alberta

5.7 Preliminary 46 items mapped on to QoL dimensions

No.	Item	Sub-Dimension	Dimension
1	I avoid being active e.g. when walking long distances or going up stairs	Daily routine	**Physical function**
2	I have low energy in the morning		
	I don't have energy when I get up in the morning		
3	Reaching down is hard for me		
4	It's hard for me to wash or dress myself		
5	I avoid taking part in activities e.g. when playing games or sports or running	Physical activity	
6	I feel that other people are better than me when I take part in activities e.g. when playing games or sports or running		
7	It's hard for me to keep up with the others when I take part in activities e.g. when playing games or sports or running		
8	I get out of breath easily if I am active e.g. when walking about or going up stairs	Out of breath	**Symptoms**
9	I get out of breath easily if I take part in activities e.g. when playing games or sports or running		
10	I get tired easily if I am active e.g. when walking about or going up stairs	Tired; Weak	
11	I get tired easily if I take part in activities e.g. when playing games or sports or running		
12	I have pain if I am active e.g. when walking about or going up stairs	Pain; hurt; ache	
13	I have pain if I take part in activities e.g. when playing games or sports or running		
14	I get picked on because of my weight	Treated differently	**Social wellbeing**
15	I feel like people avoid me because of my weight		
16	My school work is affected when I get picked on because of my weight	School/ College	**Cognitive functioning**
17	I feel that my weight will get in the way of my future	Future	**Role activities**
	I feel that my weight will stop me from doing what I want in the future		
18	I can't fit into clothes most people my age fit into	Self-esteem; Self-confidence;	**Psychological wellbeing / Personal**
19	I avoid getting changed in front of others		

20	I don't like the way I look	embarrassed; uncomfortable	**constructs**
21	I find it hard to eat food that is healthy	Control over food	
22	I find it hard to not eat too much food		
23	How I feel inside affects my eating		
24	I get **frustrated** being slower than others I get **annoyed** being slower than others e.g. when playing games or sports or running	Frustrated; Annoyed; Angry	**Psychological wellbeing / Personal constructs**
	I feel **upset** because others are faster than me (YOUNGER ADOLECENTS) e.g. when playing games or sports or running	Unhappy; Upset; Sad; Depressed	
25	I **feel disappointed** with myself being slower than others e.g. when playing games or sports or running	Disappointed	
	I feel **upset** because others are faster than me (YOUNGER ADOLECENTS) e.g. when playing games or sports or running	Unhappy; Upset; Sad; Depressed	
26	I feel **frustrated** because walking about or going up stairs is hard for me I feel **annoyed** because walking about or going up stairs is hard for me	Frustrated; Annoyed; Angry	
	I feel **upset** because walking about or going upstairs is hard for me (YOUNGER ADOLECENTS)	Unhappy; Upset; Sad; Depressed	
27	I feel **frustrated** because I get tired easily I feel **annoyed** because I get tired easily	Frustrated; Annoyed; Angry	
	I feel **upset** because I get tired easily (YOUNGER ADOLECENTS)	Unhappy; Upset; Sad; Depressed	
28	I feel **upset** when the clothes I want aren't made in my size	Unhappy; Upset; Sad; Depressed	
	I feel **disappointed** when the clothes I want aren't made in my size	Disappointed	
29	I feel **self-conscious** in shops when I go looking for new clothes to buy	Confidence; Self-conscious; Embarrassed; Uncomfortable	
30	I feel **uncomfortable** when I get changed in front of others		
31	I am **not happy** that I have to get bigger sized clothes than most people my age	Unhappy; Upset; Sad; Depressed	
32	I am **not happy** about telling people my size in clothes		
33	I feel **left out** when I go clothes shopping with my friends	Left out; Unwanted	
34	I am **unhappy** that I can't swap clothes with my friends	Unhappy; Upset; Sad; Depressed	
35	I feel **unhappy** when I can't eat what I want		

36	I feel **guilty** when I eat unhealthy food	Control over food	**Psychological wellbeing / Personal constructs**
37	I feel **guilty** when I eat too much		
38	I feel **self-conscious** when I eat in front of others	Confidence; Self-conscious; Embarrassed; Uncomfortable	
39	I feel **embarrassed** if I eat more than other people		
40	When I am at school I feel **unwanted**	Left out; Unwanted	
	I **feel different** because of my size when I am at school	Confidence; Self-conscious; Embarrassed; Uncomfortable	
41	I feel **upset** when I get picked on at school because of my size	Unhappy; Upset; Sad; Depressed	
42	My size makes me feel **less confident** around people I don't know well	Confidence; Self-conscious; Embarrassed; Uncomfortable	
43	I am not able to **enjoy** myself when I go out		
44	I feel **self-conscious** when I go out		
45	I **stick out** from other people my age		
46	I **worry** about my health in the future	Future	**Role activities**

Chapter 6 – Refinement of the weight specific LLDS

6.1 Survey – paper version

How Does Weight Affect Young People?

Primary researcher: Yemi Oluboyede

Leeds Institute of Health Sciences
Faculty of Medicine and Health
Charles Thackrah Building
University of Leeds
101 Clarendon Road
Leeds LS2 9LJ

y.oluboyede@leeds.ac.uk
0113 343 0823

Welcome

You are invited to take part in a survey looking at how weight affects the lives of young people.

We are interested in finding out how your weight may or may not affect different things in your life. We are also interested in how your weight makes you feel.

This short survey is open to any young person aged between 11 and 18. It will take approximately 15 minutes to complete. The survey does contain some potentially sensitive questions. Please be sure that you are in a space in which you feel comfortable before continuing to complete the survey.

If you have any questions or comments, please contact me using the details below.

Thank you for your time.

Name: Yemi Oluboyede

Role: Primary Researcher

My telephone number: 0113 343 0823

My email address: y.oluboyede@leeds.ac.uk

Consent

Please make sure that you have read and understood the information provided below.

1. Your participation is voluntary

2. You can stop taking part in the survey at any time without saying why

3. Your response will be kept confidential (no one will know what answers you have given)

4. The information that you give will be anonymous (no one will be able to identify you from the information you give)

5. If you want to ask any more questions about the survey you can contact the researcher by e-mail on y.oluboyede@leeds.ac.uk or at the address provided at the end of the survey

If you DO want to take part, please write your name below:

Your Name _____

Today's Date _____

Section 1: Questions about you

Q1. - *Date of Birth*: _____

Q2 – *Gender*

Boy	
Girl	

Q3. – *What is your current Weight & Height?*

Weight in Kilograms or Stone *(e.g. 40.2)*	
Height in Meters *(e.g. 1.57)*	

Q4. Are you in full-time education?

Yes	
No	

Q5. What do you feel yourself to be?

Please choose one option	
Very overweight	
Moderately overweight	
Slightly overweight	
About the right weight	
Slightly underweight	
Moderately underweight	
Very underweight	

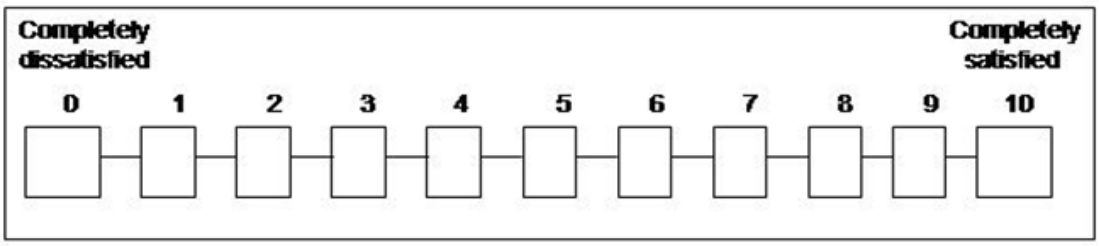
Q6. – *Where do you currently live*?

	England			Outside of England			Other *(please specify)*
	North	**Midlands**	**South**	**Scotland**	**Wales**	**Ireland**	
Please choose one option							

Q7. Which of the following do you class yourself as?

Please choose one option	
White: British / Irish / Any other white background	
Mixed: White & Black Caribbean / White & Black African/ White & Asian / Any other mixed background	
Asian or Asian British: Indian / Pakistani / Bangladeshi / Any other West or South Asian background	
Black or Black British: Caribbean / African / Any other Black background	
Chinese: Chinese / Any other East Asian background	
Other *– (please specify)*	

Q8. Thinking about your own life and personal circumstances, how satisfied are you with your life as a whole?



Q9. Overall, how would you rate your health during the past week?

Please choose one option	
Excellent	
Very good	
Good	
Fair	
Poor	

Q10. Do you have any problems with any of the following?

Please choose all that apply	
Arms / Legs / hands etc.	
Sight	
Skin conditions / allergy	
Chest / breathing	
Heart / blood pressure	
Stomach / digestion	
Diabetes	
Anxiety (worry) / depression	
Other *(please specify)*	

Section 2A: Questions about how your weight affects you

1. The questions in Section 2A ask for your views about some of the things in your life that might be affected by your weight

2. Please read though each of the questions below thinking about your weight and answer each question by **choosing ONE response by placing a tick like this ☑ next to the answer that best describes you at the moment**

3. Certain questions may look alike but each one is different. It's important that you answer ALL the questions in the survey.

4. Some questions ask about problems you may not have, it is important for us to know. Please answer each question even if you think that it does not apply to you.

5. **There are no right or wrong answers. It's what you think that matters.** If you are unsure how to answer a question please give the best answer you can

6. Remember that your individual answers will not be shared with anyone!

Section 2A: Questions about how your weight might affect you?

HOW DOES YOUR WEIGHT AFFECT YOU?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I have body pain / ache					
I get low energy					
I get tired					
I get out of breath					

HOW DOES YOUR WEIGHT AFFECT DAILY ROUTINES?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I struggle to keep up when I am walking around with others					
I struggle when I am going up stairs					
I struggle to reach or bend down					

HOW DOES YOUR WEIGHT AFFECT PHYSICAL ACTIVITIES?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I struggle to keep up with others when doing physical activity					
I struggle to keep up with others when I play sports					
I avoid doing things like running, cycling, swimming or playing sports					

HOW DOES YOUR WEIGHT AFFECT SCHOOL / COLLEGE?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I struggle to do as well as others at school					
I struggle to concentrate on school / college work					

HOW DOES YOUR WEIGHT AFFECT SOCIALISING?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I get treated differently at school, such as being teased or picked-on or left out					
I get treated differently at home, such as being teased or picked-on or left out					
People treat me differently when I go out					
I avoid playing / hanging out or socialising with others					

HOW DOES YOUR WEIGHT AFFECT THE FUTURE?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I worry about my health in the future					
I worry about the type of job/career I will be able to have					

HOW DOES YOUR WEIGHT AFFECT THE WAY YOU FEEL?	Choose one answer for each question				
	Not at all	A little	Quite a bit	A lot	Very much
I feel angry or annoyed because I am unable to do the same things as others					
I feel frustrated because I am unable to do the same things as others					
I feel uncomfortable or embarrassed getting changed in front of others					
I feel uncomfortable or embarrassed shopping for clothes					
I feel uncomfortable or embarrassed meeting new people					
I feel uncomfortable or embarrassed eating in front of others					
I feel unhappy because I can't eat what I want					
I feel unhappy about the way I look					
I feel unhappy because I am unable to do the same things as others					
I feel disappointed because clothes aren't made in the size I need					
I struggle to keep in control of what I eat					

Section 2B: Questions about how your weight affects you

1. The questions in Section 2B ask for your views about some of the things in your life that might be affected by your weight. These are similar to the questions on the previous page; however, a different choice of answers is given.

As with the previous page:

2. Please read though each of the questions below thinking about your weight and answer each question by **choosing ONE response by placing a tick like this ☑ next to the answer that best describes you at the moment**

3. Certain questions may look alike but each one is different. It's important that you answer ALL the questions in the survey.

4. Some questions ask about problems you may not have, it is important for us to know. Please answer each question even if you think that it does not apply to you.

5. **There are no right or wrong answers. It's what you think that matters**. If you are unsure how to answer a question please give the best answer you can

6. Remember that your individual answers will not be shared with anyone!

Section 2B: Questions about how your weight might affect you?

HOW DOES YOUR WEIGHT AFFECT YOU?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I have body pain / ache					
I get low energy					
I get tired					
I get out of breath					

HOW DOES YOUR WEIGHT AFFECT DAILY ROUTINES?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I struggle to keep up when I am walking around with others					
I struggle when I am going up stairs					
I struggle to reach or bend down					

HOW DOES YOUR WEIGHT AFFECT PHYSICAL ACTIVITIES?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I struggle to keep up with others when doing physical activity					
I struggle to keep up with others when I play sports					
I avoid doing things like running, cycling, swimming or playing sports					

HOW DOES YOUR WEIGHT AFFECT SCHOOL / COLLEGE?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I struggle to do as well as others at school					
I struggle to concentrate on school / college work					

HOW DOES YOUR WEIGHT AFFECT SOCIALISING?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I get treated differently at school, such as being teased or picked-on or left out					
I get treated differently at home, such as being teased or picked-on or left out					
People treat me differently when I go out					
I avoid playing / hanging out or socialising with others					

HOW DOES YOUR WEIGHT AFFECT THE FUTURE?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I worry about my health in the future					
I worry about the type of job/career I will be able to have					

HOW DOES YOUR WEIGHT AFFECT THE WAY YOU FEEL?	Choose one answer for each question				
	Never	Almost never	Sometimes	Often	Always
I feel angry or annoyed because I am unable to do the same things as others					
I feel frustrated because I am unable to do the same things as others					
I feel uncomfortable or embarrassed getting changed in front of others					
I feel uncomfortable or embarrassed shopping for clothes					
I feel uncomfortable or embarrassed meeting new people					
I feel uncomfortable or embarrassed eating in front of others					
I feel unhappy because I can't eat what I want					
I feel unhappy about the way I look					
I feel unhappy because I am unable to do the same things as others					
I feel disappointed because clothes aren't made in the size I need					
I struggle to keep in control of what I eat					

Prize draw

Thank you for completing the survey!

You now have the opportunity to be entered into a prize draw to win one of 8 prizes (First Prize: 1 x £30 or Runner up prizes: 7 x £10 of retail Love-To-Shop vouchers).

You will need to provide an e-mail address so we can notify you if you win. Your email will only be used to contact you about the result of the prize draw. Your email address will not be linked to any of the answers you have given and will not be used for any other purpose.

Please enter your e-mail address to take part: _____

Please re-enter your e-mail address: _____

Thank you for taking the time to complete this survey. Please use the details below if you need to contact me.

Name: Yemi Oluboyede
Role: Primary Researcher
My telephone number: 0113 343 0823
My email address: y.oluboyede@leeds.ac.uk

My work address: Leeds Institute of Health Sciences
Faculty of Medicine and Health
Charles Thackrah Building
University of Leeds
101 Clarendon Road
Leeds LS2 9LJ

6.2 Summary of Fit Statistics – Psychometric analysis

Steps	Fit statistic	Interpretation	Assessment level: a) by dimension or b) by Item	Fit test /Criteria
Step I	Item correlation matrix	Items that are not independent will have a high correlation	Item	Identification of correlations exceed 0.32
	Items not discriminating between groups	Individual t-tests comparing mean QoL scores for adjacent weight categories: a) Underweight vs. Normal b) Normal vs. Overweight c) Overweight vs. Obese	Item	**P-values** **Significance level > 0.10** = Not discriminating between weight categories **0.10 ≥ Significance level > 0.05** = Discriminating between weight categories at the 10% level of significance **Significance level ≤** 0.05 = Discriminating between weight categories at the 5% level of significance
	The factor correlation matrix	If correlations exceed 0.32, then there is 10% (or more) overlap in variance among factors, enough variance to warrant oblique rotation unless there are compelling reasons for orthogonal rotation.	Dimension	This should be assessed for correlations around 0.32 and above
	OLS regression with total score as dependent variable	Items not significant in predicting total score	Item	**P-values** **Significance level > 0.10** = Not significant determinant of weight specific QoL **0.10 ≥ Significance level > 0.05** = A significant determinant of weight specific QoL at the 10% level of significance **Significance level ≤** 0.05 = A significant determinant of weight specific QoL at the 5% level of significance
Step III	Feasibility (rate of missing data)		Item	Select items with the minimum amount of missing data
	Internal consistency (correlation between item & dimension score) assessed using each subscale - item total correlations	Check for items with low item-total correlations	Dimension	Alpha >0.80 and items that item-total correlations of between 0.2 and 0.8. if less than 0.2 they are not related if higher 0.8 too related and are redundant
	Distribution of responses		Item	Items with high floor effect (>40%)

6.3 Summary of Fit Statistics – Rasch Analysis*

Steps	Fit statistic	Interpretation	Assessment level: a) by dimension or b) by Item	Fit test /Criteria
Step II	The item–trait interaction	Measures whether data fit the Rasch model for discrete groups of responders	1. Assessment at dimension level 2. Assessment at item level	X^2 test statistic - the convention is that the P-value for the overall model X^2 statistic should be > 0.01 for a well-fitting model
	The person separation index (PSI)	Measures the level of discrimination amongst different groups of responders, in other words the extent to which items distinguish between different levels of functioning	Assessment at dimension level	A PSI of 0.7 or more indicates a well fitting Rasch model (ideally it should be between 0.7 and 0.8)
	Fit Residuals	The difference (or residuals) between the observed data from the instrument and what would be expected from the Rasch Model for each respondent or item response. They are summed over all items (item fit residuals) or summed over all persons (person fit residuals).	Assessment at dimension level	A perfect item or person fit residual should have a mean value of approximately zero with a standard deviation approximately equal to unity
Step III	Threshold probability curves	Used to examine the spread of item levels at a point on the latent scale, typically at the central logit zero	Assessment at item level	Item levels that are closer together, in comparison with other levels, are candidates for item-level collapsing.
	Item goodness of fit statistics	Where the better the goodness of fit (high X^2 value and non-significant P-value) the better the item represents the underlying uni-dimensional latent scale for each dimension	Assessment at item level	X^2 test statistic - the convention is that the P-value for the overall model X^2 statistic should be > 0.01 for a well-fitting model
Additional assessments	Invariance of the items	A check to explore if the ratios of items remain the same	Individual Item, but taken in consideration of all items in the instrument	A non significant c2 Significance level/ P value?
	Item difficulty	The relative difficulty of the items along the Rasch ruler	Individual Item, but taken in consideration of all items in the instrument	A range (or spread) of items, as indicated by the item threshold distribution graph

*Fit statistics were considered in order to test how well the observed data fit the expectations of the Rasch measurement model. This table presents the fit statistics used in RUMM2030 and the criteria on which the assessment of each is made. The Assessment Level column represents what level of the instrument is being evaluated with each fit statistic.

6.4 Correlation matrix for the seven factors in an EFA with Direct Oblimin Rotation

Frequency response scale*

Factor	1	2	3	4	5	6	7
1	1.000						
2	.452	1.000					
3	.552	.511	1.000				
4	.511	.305	.419	1.000			
5	-.573	-.410	-.565	-.492	1.000		
6	.301	.395	.420	.329	-.328	1.000	
7	-.571	-.372	-.440	-.405	.464	-.420	1.000

Severity response scale*

Factor	1	2	3	4	5	6	7
1	1.000						
2	-.074	1.000					
3	.551	-.208	1.000				
4	.540	-.149	.433	1.000			
5	-.547	.149	-.563	-.503	1.000		
6	-.465	.220	-.285	-.324	.369	1.000	
7	-.592	.299	-.493	-.581	.557	.449	1.000

*Note that none of the correlations exceed the upper threshold of 0.80 however, as most of the correlations exceed the lower threshold of 0.32 (as per Tabachnick and Fiddell 2007), this suggests that the factors are not independent and thus a non-orthoginal rotation should be run

6.5 Fit statistics for the CFA

Model	The Tucker Lewis index (TLI)[a]	comparative fit index (CFI)[a]	The root mean square error of approximation (RMSEA)[b]
3 factor solution	0.772	0.79	0.125
4 factor solution	0.843	0.859	0.109
5 factor solution	0.874	0.89	0.099
6 factor solution	0.873	0.887	0.093
7 factor solution	0.908	0.92	0.08

[a] values approaching unity are desirable, with values greater than approximately 0.95 considered to indicate a good fit (as per Sim *et al*., 2011)

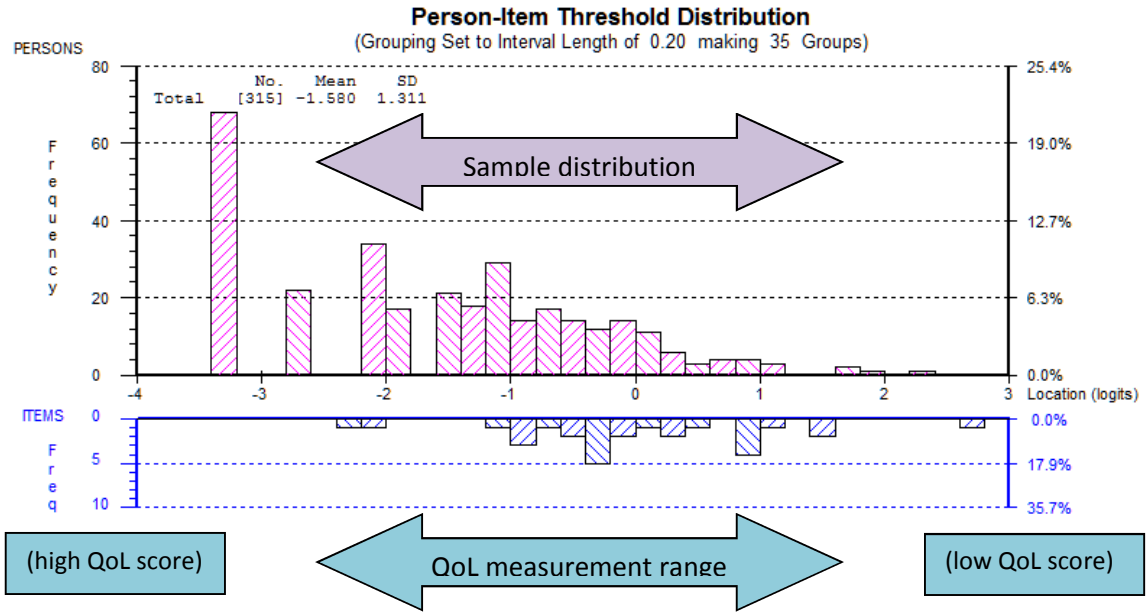
[b] values approaching zero are desirable, preferably below approximately 0.05 (as per Sim *et al*., 2011)

6.6 Residual correlation matrix - final seven items*

Items	3	5	10	12	15	22
3						
5	-0.123					
10	-0.065	-0.032				
12	-0.167	-0.204	-0.166			
15	-0.233	-0.128	-0.277	-0.134		
22	-0.322	-0.127	-0.267	-0.196	-0.016	
27	-0.285	-0.044	-0.192	-0.242	0.009	0.077
Mean	**-0.149**					
Test statistic	**-0.349**[b]					

*The test for local dependency assessed whether the residual correlations were higher than the test statistics **-0.349** (= mean + 0.2), none of which were.

6.7 Person item distribution - final seven items*



*The person-item threshold distribution assesses whether the scale-to-sample targeting is adequate for making judgements about the performance of the seven item scale and the measurement of people. The pink blocks on the upper part of the graph represent groups of respondents and their QoL scores. The blue blocks on the lower part of the scale represent the item locations and their distribution. There is a good overlap between the persons and items, indicating that person locations are covered by items and also that the item locations are covered by persons.

Chapter 7 - Psychometric properties of the WAItE: A preliminary assessment

7.1 Information sheet and consent - adolescents

YOUNG PERSON'S INFORMATION SHEET

AND CONSENT FORM

MEASURING QUALITY OF LIFE IN YOUNG PEOPLE

We would like you to take part in a research study looking at the lives of young people. Before you decide if you want to take part it's important to know why we are doing the study. So please think about the information below carefully.

What does the research involve?
You are invited to complete a short survey looking at different aspects of the lives of young people. It will take less than 5 minutes to complete.

Why have I been invited to take part?
You have been chosen because you are attending the Carnegie summer camp. We are hoping that a number of young people from the camp will complete the survey on 2 occasions, on your first and last days of attending the camp.

Do I have to take part?
No, it is up to you if you take part or not. Even if you say yes at the start, you can stop taking part at any time during the study. You do not have to give a reason.

Contact details
If you have any questions or would like more information you can speak to the researchers involved in the study. Contact details are:

Yemi Oluboyede
NIHR Doctoral Research Fellow
University of Leeds
Charles Thackrah Building
101 Clarendon Road
Leeds LS2 9LJ
Tel: +44 (0)113 343 0823
E mail: y.oluboyede@leeds.ac.uk

Dr. Claire Hulme
or
Professor Andrew Hill
Tel: +44 (0) 113 343 6966
E mail: hssjad@leeds.ac.uk

Thank you for reading this – please ask any questions if you need to.

CONSENT

FORM

Before continuing with the survey, please make sure that you understand and agree with each of these statements below by placing a tick next to each one:

I understand the following.....	*Please tick*
1. My participation is voluntary	
2. I can stop taking part in the study at any time without saying why	
3. My response will be kept confidential (no one will know what answers I have given)	
4. The information that I give will be anonymous (no one will be able to identify me from the information I give)	
5. If I want to ask any more questions about the study I can contact the researchers using the details on the previous page	
6. I am happy to take part	

Only sign your name if you want to take part!

If you DO want to take part, please write your name below:

Print Your Name _____

Today's Date _____

If you need to contact me, here are my details:

My name: Yemi

Tel: 0113 343 0823

E-mail: y.oluboyede@leeds.ac.uk

7.2 The WAItE

Quality of life questionnaire for young people

INSTRUCTIONS:

The questions below ask about different things in your life. Read through each question carefully and **TICK ONE RESPONSE THAT BEST DESCRIBES YOU AT THE MOMENT**.

EXAMPLE:

If you 'sometimes' have body pain then you should tick this box:

	Never	Almost never	Sometimes	Often	Always
I have body pain			✓		

Now please read through each of the 7 questions below and choose <u>ONE</u> answer that best describes you at the moment.

	Never	Almost never	Sometimes	Often	Always
1. I get tired					
2. I struggle to keep up when I am walking around with others					
3. I avoid doing sports					
4. I struggle to concentrate on my studies / work					
5. I feel embarrassed shopping for clothes					
6. I feel unhappy because I am unable to do the same things as others					
7. People treat me differently when I go out					

7.3 Distribution of scores across response options

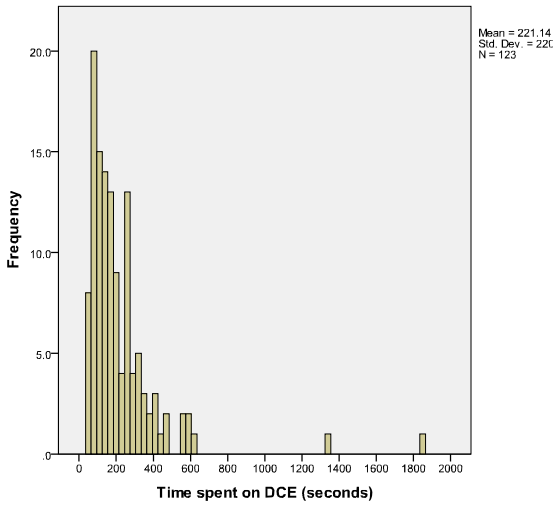
	Level	WAItE items													
		1		2		3		4		5		6		7	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%
T1 (n=30)	1	0	0	9	30.0	15	50.0	7	23.3	6	20.0	5	16.7	11	36.7
	2	3	10.0	7	23.3	7	23.3	10	33.3	6	20.0	5	16.7	8	26.7
	3	17	56.7	7	23.3	3	10.0	6	20.0	9	30.0	9	30.0	6	20.0
	4	6	20.0	5	16.7	3	10.0	5	16.7	4	13.3	4	13.3	3	10.0
	5	4	13.3	2	6.7	2	6.7	2	6.7	5	16.7	7	23.3	2	6.7
T2 (n=27)	1	1	3.3	12	40.0	11	36.7	8	26.7	10	33.3	10	33.3	10	33.3
	2	10	33.3	9	30.0	8	26.7	7	23.3	4	13.3	5	16.7	6	20.0
	3	12	40.0	5	16.7	7	23.3	11	36.7	7	23.3	6	20.0	8	26.7
	4	4	13.3	1	3.3	1	3.3	0	0	3	10.0	6	20.0	0	0
	5	0	0	0	0	0	0	1	3.3	3	10.0	0	0	3	10.0

Chapter 8 - Valuation of the WAItE: A methodological exploration

8.1 Characteristics of respondents in the DCE surveys

Characteristic		Pilot A (N=112)		Pilot B (N=206)		Pilot C (N=171)	
		N	%	N	%	N	%
Marital status	Married/partner	69	61.60	115	55.80	102	59.60
	Other	43	38.40	91	44.20	69	40.40
Employment status	Employed/self employed	64	57.14	135	65.50	114	66.70
	Student	8	7.14	8	3.90	8	4.70
	Not working	40	35.71	63	30.60	49	28.70
Highest level of education	No qualifications	5	4.50	9	4.40	8	4.70
	University undergraduate / Higher degree	37	33.00	86	41.75	59	34.50
	Other	70	62.50	111	53.88	104	60.80
Self assessed health	Excellent	13	11.61	28	13.59	25	14.62
	Good	59	52.68	103	50.00	84	49.12
	Fair	32	28.57	52	25.24	45	26.32
	Poor	8	7.14	19	9.22	15	8.77
	Very Poor	0	0.00	4	1.94	2	1.17
Satisfaction with appearance	0-Completely dissatisfied	3	2.68	2	0.97	2	1.17
	1	.0	0.00	2	0.97	3	1.75
	2	1	0.89	7	3.40	8	4.68
	3	6	5.36	11	5.34	12	7.02
	4	7	6.25	14	6.80	13	7.60
	5	17	15.18	25	12.14	19	11.11
	6	20	17.86	40	19.42	37	21.64
	7	24	21.43	52	25.24	35	20.47
	8	18	16.07	38	18.45	25	14.62
	9	5	4.46	8	3.88	13	7.60
	10-Completely satisfied	11	9.82	7	3.40	4	2.34

8.2 Speeder check based on Pilot A dataset



Time spent on DCE exercise (seconds)		
N	Valid	123
	Missing	0
Percentiles	10	75.00
	25	105.00
	50	168.00
	75	265.00
	90	406.20

8.3 Frequency table – results of VAS anchoring

VAS score		Frequency	Percent	Valid percent	Cumulative percent
Valid	-.50	6	3.5	3.6	3.6
	-.49	1	.6	.6	4.1
	-.41	1	.6	.6	4.7
	-.37	1	.6	.6	5.3
	-.20	4	2.3	2.4	7.7
	-.13	1	.6	.6	8.3
	-.11	2	1.2	1.2	9.5
	-.10	2	1.2	1.2	10.7
	-.01	1	.6	.6	11.2
	.00	5	2.9	3.0	14.2
	.01	2	1.2	1.2	15.4
	.03	1	.6	.6	16.0
	.04	2	1.2	1.2	17.2
	.05	1	.6	.6	17.8
	.06	1	.6	.6	18.3
	.07	2	1.2	1.2	19.5
	.08	1	.6	.6	20.1
	.09	3	1.8	1.8	21.9
	.10	22	12.9	13.0	34.9
	.11	1	.6	.6	35.5
	.12	2	1.2	1.2	36.7
	.14	2	1.2	1.2	37.9
	.17	2	1.2	1.2	39.1
	.18	1	.6	.6	39.6
	.19	2	1.2	1.2	40.8
	.20	17	9.9	10.1	50.9
	.21	2	1.2	1.2	52.1
	.22	1	.6	.6	52.7
	.23	2	1.2	1.2	53.8
	.24	2	1.2	1.2	55.0
	.25	2	1.2	1.2	56.2
	.26	1	.6	.6	56.8
	.27	3	1.8	1.8	58.6
	.28	3	1.8	1.8	60.4
	.29	4	2.3	2.4	62.7
	.30	16	9.4	9.5	72.2
	.31	2	1.2	1.2	73.4
	.32	1	.6	.6	74.0
	.33	1	.6	.6	74.6
	.34	4	2.3	2.4	76.9
	.35	1	.6	.6	77.5
	.36	1	.6	.6	78.1
	.38	1	.6	.6	78.7
	.39	2	1.2	1.2	79.9
	.40	5	2.9	3.0	82.8
	.46	1	.6	.6	83.4
	.50	7	4.1	4.1	87.6
	.58	1	.6	.6	88.2
	.59	1	.6	.6	88.8
	.60	1	.6	.6	89.3
	.62	1	.6	.6	89.9
	.66	1	.6	.6	90.5
	.67	1	.6	.6	91.1
	.68	2	1.2	1.2	92.3
	.69	2	1.2	1.2	93.5
	.70	3	1.8	1.8	95.3
	.71	1	.6	.6	95.9
	.72	1	.6	.6	96.4
	.81	1	.6	.6	97.0
	.82	1	.6	.6	97.6
	.95	2	1.2	1.2	98.8
	.99	1	.6	.6	99.4
	1.00	1	.6	.6	100.0
	Total	**169**	**98.8**	**100.0**	
Missing	**System**	**2**	**1.2**		
Total		**171**	**100.0**		

8.4 Frequency table – results of pairwise anchoring

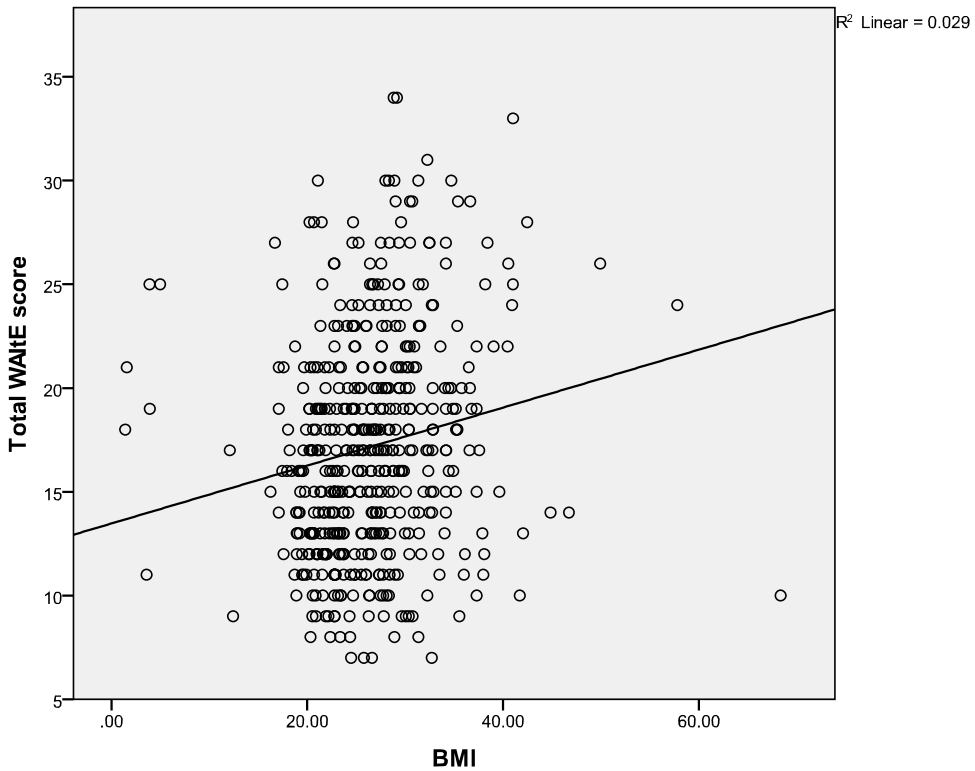
Alternatives presented in Task 1		**Count (n=171)**		**Alternatives presented in Task 2**		**Count (n=55)***	
Years of life in PITS **Life A**	Years of life in full health **Life B**	Individuals choosing **Life A - PITS**	Individuals choosing **Life B**	Years of life in PITS **Life A**	Years of life in full health **Life B**	Individuals choosing **Life A - PITS**	Individuals choosing **Life B**
20	4	8	12	20	15	4	3
20	6	8	7	20	17	3	5
20	17	1	7	20	18	1	4
20	18	1	3	20	19	2	5
50	10	15	17	50	37	2	4
50	15	16	29	50	42	3	6
50	37	4	15	50	45	4	3
50	42	2	18	50	47	1	5
50	45	0	8				
Total		**55**	**116**	**Total**		**20**	**35**

*The two pairwise tasks were randomly selected for each individual to complete. 116 individuals chose to live in Life B (full health) when the first task was presented. This was also the shorter duration from the two pre-selected tasks so these individuals were not presented with a second task. Only the 55 individuals who chose to live in Life A (PITS) for a longer period of time were presented with the second choice task.

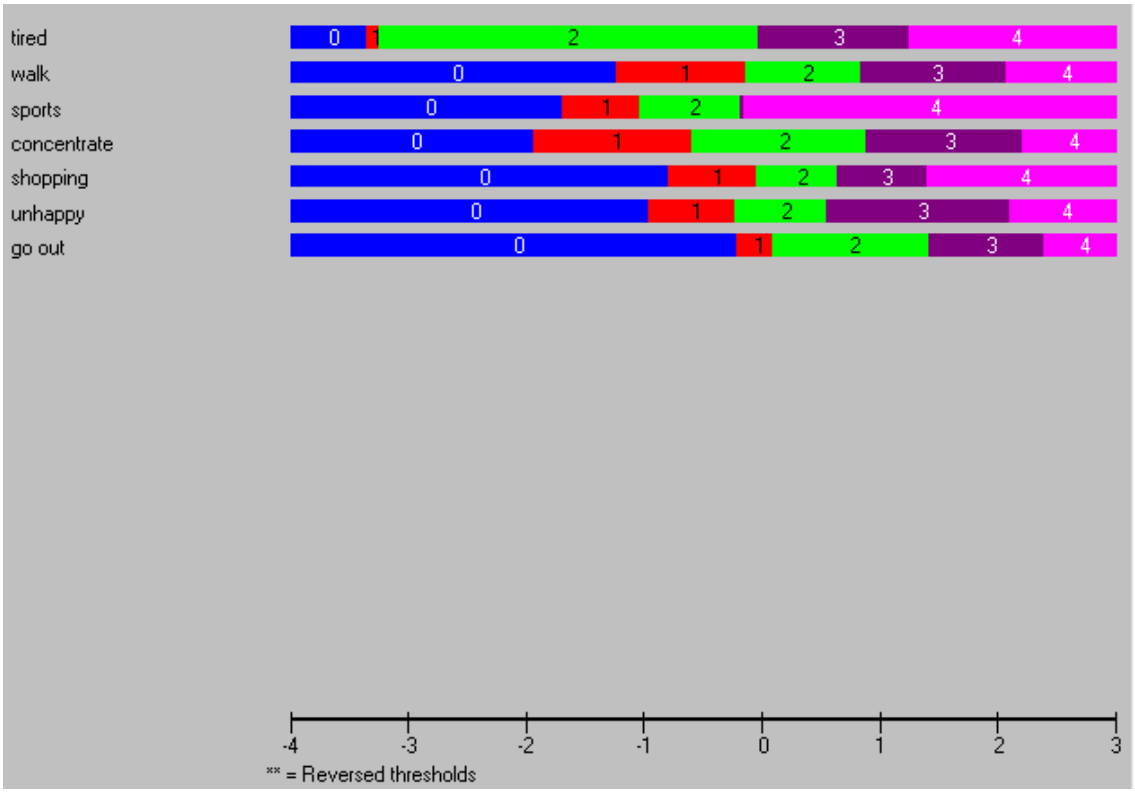
8.5 Psychometric properties of the WAItE

WAItE item N=489	Cronbach's Alpha if Item Deleted	
I get tired	.819	
I struggle to keep up when I am walking around with others	.789	
I avoid doing sports	.832	
I struggle to concentrate on my studies / work	.812	
I feel embarrassed shopping for clothes	.796	
I feel unhappy because I am unable to do the same things as others	.771	
People treat me differently when I go out	.796	
Weight category	**Mean total score**	**SD**
Underweight (N=21)	18.19	4.802
NorWT (N=193)	15.64	4.674
OverWT (N=158)	17.70	5.404
Obese (N=117)	18.79	5.826
t-test (2-tailed)	**t statistic**	**p-value**
Discrimination between Overweight & Obese	1.612	0.108
Discrimination between Normal weight & Overweight	3.826	p<0.00
Discrimination between Normal weight & Obese	4.972	p<0.00

8.6 Scatter plot of BMI and total WAItE total score



8.7 Threshold map of the WAItE*



*The dataset is composed of the combined full sample from all three pilot studies (N=489). Tired = I get tired, walk = I struggle to keep up when I am walking around with others, sports = I avoid doing sports, concentrate = I struggle to concentrate on my studies / work, shopping = I feel embarrassed shopping for clothes, unhappy = I feel unhappy because I am unable to do the same things as others and, go out = People treat me differently when I go out. The threshold map shows that none of the seven items displays threshold disordering.

8.8 Parameter estimates - Main effects models*

Variable	Pilot A 35-54 yeas only		Pilot B Sample 1 35-54 yeas only		Pilot B Sample 3 35-54 yeas only	
	Parameter estimate	P-value	Parameter estimate	P-value	Parameter estimate	p-value
Tired_L2	0.487	0.24	0.032	0.93	0.042	0.90
Tired_L3	1.160	0.09	0.091	0.87	-1.050	0.05
Tired_L4	2.070	0.03	-0.231	0.77	-0.911	0.19
Tired_L5	2.950	0.04	-0.093	0.93	-2.150	0.05
Walking_L2	0.883	0.02	0.503	0.12	-0.541	0.03
Walking_L3	1.790	0.01	0.498	0.32	-1.000	0.06
Walking_L4	2.830	0.01	0.618	0.34	-1.580	0.05
Walking_L5	4.030	0.01	0.801	0.46	-1.940	0.07
Sports_L2	1.060	0.01	0.238	0.41	0.220	0.45
Sports_L3	1.790	0.01	0.420	0.35	-0.263	0.60
Sports_L4	2.620	0.01	0.428	0.51	-0.753	0.28
Sports_L5	3.650	0.01	0.882	0.43	-1.390	0.20
Concentrate_L2	1.070	0.01	0.205	0.51	-0.991	0.00
Concentrate_L3	1.680	0.02	-0.008	0.99	-0.851	0.10
Concentrate_L4	2.840	0.00	-0.028	0.97	-2.100	0.00
Concentrate_L5	3.630	0.02	-0.108	0.92	-2.660	0.01
Embarrased_L2	1.110	0.03	0.520	0.08	-0.485	0.18
Embarrased_L3	1.550	0.02	0.460	0.36	-1.210	0.02
Embarrased_L4	2.320	0.02	0.533	0.47	-1.740	0.02
Embarrased_L5	3.490	0.02	0.526	0.60	-2.210	0.04
Unhappy_L2	1.190	0.01	0.268	0.40	-0.077	0.82
Unhappy_L3	1.470	0.04	-0.351	0.50	-1.300	0.01
Unhappy_L4	2.260	0.02	-0.111	0.87	-1.760	0.02
Unhappy_L5	3.390	0.03	-0.273	0.81	-2.920	0.01
TreatDif_L2	1.230	0.00	-0.063	0.88	-0.847	0.01
TreatDif_L3	1.660	0.01	0.266	0.63	-1.170	0.03
TreatDif_L4	2.520	0.01	0.257	0.71	-1.600	0.02
TreatDif_L5	3.510	0.02	0.210	0.85	-2.190	0.05
Duration	-0.547	0.14	0.048	0.42	0.472	0.00
Constant	-0.024	0.83	-0.019	0.89	0.253	0.05

Main effects models	Pilot A	Pilot B Sample 1	Pilot B Sample 3
	35-54yrs (D10)[a]	35-54yrs (D50)[a]	35-54yrs (D20)[a]
No. observations	350	510	480
No. individuals	35	51	48
Likelihood ratio test	95.107	95.462	102.412
Adjusted rho-square	0.072	0.05	0.064
No.-ve QoL coefficients (%)	0 (0)	9 (32.14)	26 (93)

*Results for individuals aged 35-54 years only. The seven individuals who self reported not understanding the DCE tasks were excluded from the analysis. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by two and D50 = standard duration levels scaled up by five

8.9 Parameter estimates – Interaction effects models*

Variable	Pilot A 35-54 yeas only		Pilot B Sample 1 35-54 yeas only		Pilot B Sample 3 35-54 yeas only	
	Parameter estimate	P-value	Parameter estimate	P-value	Parameter estimate	P-value
Tired_L2xD	0.012	0.76	0.004	0.65	0.026	0.18
Tired_L3xD	0.015	0.74	0.008	0.41	-0.022	0.29
Tired_L4xD	0.073	0.29	0.001	0.93	0.012	0.74
Tired_L5xD	0.070	0.54	0.010	0.59	-0.040	0.45
Walking_L2xD	0.057	0.14	0.013	0.04	-0.019	0.30
Walking_L3xD	0.094	0.14	0.015	0.06	-0.021	0.48
Walking_L4xD	0.167	0.08	0.022	0.09	-0.035	0.30
Walking_L5xD	0.195	0.10	0.029	0.08	-0.035	0.50
Sports_L2 xD	0.070	0.11	0.009	0.23	0.034	0.12
Sports_L3 xD	0.095	0.15	0.014	0.10	0.029	0.26
Sports_L4 xD	0.128	0.14	0.017	0.14	0.019	0.57
Sports_L5 xD	0.146	0.26	0.031	0.10	0.004	0.94
Concentrate_L2xD	0.085	0.03	0.005	0.43	-0.052	0.01
Concentrate_L3xD	0.077	0.26	0.002	0.85	-0.024	0.43
Concentrate_L4xD	0.177	0.02	0.004	0.73	-0.086	0.01
Concentrate_L5xD	0.131	0.28	0.002	0.88	-0.081	0.10
Embarrased_L2xD	0.065	0.16	0.017	0.01	-0.002	0.93
Embarrased_L3xD	0.064	0.29	0.017	0.08	-0.038	0.15
Embarrased_L4xD	0.076	0.37	0.019	0.13	-0.051	0.11
Embarrased_L5xD	0.113	0.36	0.022	0.19	-0.040	0.44
Unhappy_L2xD	0.087	0.03	0.009	0.13	0.028	0.12
Unhappy_L3xD	0.061	0.26	-0.004	0.60	-0.036	0.18
Unhappy_L4xD	0.092	0.29	0.004	0.70	-0.052	0.18
Unhappy_L5xD	0.102	0.39	0.002	0.92	-0.083	0.11
TreatDif_L2 xD	0.085	0.05	0.002	0.80	-0.024	0.10
TreatDif_L3 xD	0.074	0.24	0.014	0.19	-0.023	0.33
TreatDif_L4 xD	0.127	0.13	0.013	0.23	-0.049	0.10
TreatDif_L5 xD	0.135	0.28	0.018	0.31	-0.046	0.38
Duration	-0.375	0.51	-0.031	0.71	0.423	0.10
Constant	0.018	0.87	-0.022	0.88	0.178	0.11

Interaction effects models	Pilot A	Pilot B Sample 1	Pilot B Sample 3
	35-54yrs (D10)[a]	35-54yrs (D50)[a]	35-54yrs (D20)[a]
No. observations	350	510	480
No. individuals	35	51	48
Likelihood ratio test	93.761	97.446	99.534
Adjusted rho-square	0.07	0.053	0.059
No.-ve QoL coefficients (%)	0 (0)	1 (3.57)	21 (75)

*Results for individuals aged 35-54 years only. The seven individuals who self reported not understanding the DCE tasks were excluded from the analysis. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by two and D50 = standard duration levels scaled up by five