Article:
Oluboyede, Y, Hulme, C and Hill, A (2017) Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. Quality of Life Research, 26 (8). pp. 2025-2039. ISSN: 0962-9343

https://doi.org/10.1007/s11136-017-1561-1

Title: Development and refinement of the WAItE - a new obesity specific quality of life measure for adolescents

Authors: Yemi Oluboyede[1], Claire Hulme[2] & Andrew Hill[2]

[1] Institute of Health and Society, Newcastle University, Newcastle Upon Tyne, UK

[2] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

Corresponding author:

Yemi Oluboyede

Institute of Health and Society

Newcastle University

Baddiley-Clark Building

Richardson Road

Newcastle upon Tyne, NE2 4AX

E-mail: yemi.oluboyede@ncl.ac.uk

Key words: Obesity, Quality of Life, Economic evaluation, Adolescents, Condition-specific measure, Rasch analysis, Instrument development, Instrument refinement

ABSTRACT

Background: Few weight specific outcome measures, developed specifically for obese and overweight adolescents, exist and none are suitable for the elicitation of utility values used in the assessment of cost effectiveness.

Objectives: The development of a descriptive system for a new weight specific measure.

Methods: Qualitative interviews were conducted with 31 treatment seeking (above normal weight status) and a non-treatment seeking (school sample) of adolescents aged 11-18 years, to identify a draft item pool and associated response options. 315 eligible consenting adolescents, aged 11 to 18 years, enrolled in weight management services, and recruited via an online panel, completed two version of a long list 29 item descriptive system (consisting of frequency and severity response scales). Psychometric assessments and Rasch analysis were applied to the draft 29 item instrument to identify a brief tool containing the best performing items and associated response options.

Results: Seven items were selected, for the final item set, all displayed internal consistency, moderate floor effects, and the ability to discriminate between weight categories. The assessment of uni-dimensionality was supported (t-test statistic of 0.024, less than the 0.05 threshold value).

Conclusions: The **W**eight specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (**WAItE**) focuses on aspects of life affected by weight that are important to adolescents. It has potential for adding key information to the assessment of weight management interventions aimed at the younger population.

INTRODUCTION

The prevalence of obesity in childhood and adolescence is rising [1]. Pediatric obesity is associated with reduced overall weight-specific quality of life (QoL), higher risk of morbidity, disability, and premature mortality in adulthood [2-4]. Obese or overweight children and adolescents report lower QoL compared with their lean counterparts, typically manifesting in physical and social functioning dimensions of QoL measures and studies have also reported decrements in emotional functioning [2, 3]. Whilst there is evidence to support dietary and lifestyle interventions in treating childhood obesity [5], policymakers increasingly require evidence to assess value for money [6], and evidence on the cost effectiveness of these interventions is currently lacking. Cost effectiveness analysis is an incremental assessment comparing the costs and outcomes of two or more health interventions (see Drummond et al. [7] for further explanation). The current lack of a suitable outcome measure for economic evaluation in weight management interventions aimed at the younger population is a hindrance [8].

The recommended currency of outcome measurement in cost effectiveness analysis is the quality adjusted life year (QALY) [6, 7]. The QALY is a measure of mortality and morbidity; the latter is typically measured using a generic health related QoL, preference based measure. Preference based measures differ from non–preference-based measures in the way the scoring algorithms have been derived, in that they are estimated from the values people place on different aspects of health rather than a simple summative scoring procedure [9]. In the adult population, generic measures have performed poorly in discriminating between different Body Mass Index (BMI) subgroups relative to weight specific measures [10, 11].

There exists no condition-specific *preference-based* measure for assessment of QoL in obese adolescents, although there are a small number of weight specific instruments e.g. KINDL-Obesity module [12], Impact of Weight on Quality of Life–Kids version (IWQOL-Kids) [13], Moorehead-Ardelt

Quality of Life Questionnaire II (M-A-QoL Q) [14], Sizing Me Up [15], Youth Quality of Life – Weight (YQOL-W) [16] and Oxford Paediatric Obesity Instrument (OPOI) [17]. None of these measures are designed for the calculation of preference weights for use in cost effectiveness analysis. Moreover, at the time the current study was being undertaken, the only validated instrument that was developed incorporating the views of children and adolescents aged 11 – 18 was the YQOL-W [16] and no existing validated instrument had incorporated the views of adolescents living outside the US. Cultural factors play an important role in perceptions of weight and weight-related health consequences, hence the generalizability of the YQOL-W to the UK population must be in question. Furthermore, although the content of the OPOI was informed by adolescents living in the UK, the tool had not been validated at the time the current study was undertaken.

Mapping can be used to estimate preference values where none are available [6]. Preference weights can be mapped onto validated non-preference based instruments, such as in the YQOL-W [16]. This technique has been undertaken in the adult population using the IWQOL-Lite questionnaire [11], where preference weights from the Short-Form Six-Dimension (SF-6D) were mapped to it [10]. For the 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 [9]. Brazier *et al*. emphasise that, '*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*', a view reflected in the National Institute for Health and Care Excellence (NICE) guide [6, 10]. When the information is not available (i.e. where a validated condition specific preference based tool does not exist), mapping between instruments might be the only alternative. An added complication for the adolescent population is that there is no agreed gold standard generic preference based tool to map to [6].

In their current forms neither the YQOL-W nor OPOI can be used to undertake a preference valuation study, nor were they created for this purpose. This is because they do not fulfil a number of specific

requirements that need to be considered in creating a preference-based measure. These include having a limited number of multidimensional items (ideally 5 to 9 items, see Brazier et al. [9]). The YQOL-W and OPOI comprise 21 and 30 items respectively. In order to elicit preference values for different characteristics of a particular condition, individual items should also be heterogeneous with associated ordered categorical response options that are mutually exclusive and can be used to describe the weight specific health states that are important to adolescents. The YQOL-W does not meet this requirement as it utilises a numerical 11-point scale (anchored by *not at all* and *very much*) for all items, limiting its suitability for the purposes of preference valuation. Health states comprise a descriptive system made up of items within dimensions, and response categories for each item. Health states are valued by individuals using preference elicitation techniques [7].

The overall aim of this study therefore was to develop a new descriptive system that met the aforementioned specific requirements for generating health states and that could be utilised in a preference elicitation study. As a first step towards achieving this aim, qualitative methods were used with the population of interest to collect information about the impact of weight on health-related quality of life. Adolescent's views were crucial to the development of the content in order to focus on aspects of life affected by weight that were important to them. Following this, the next step involved the identification of dimensions, and the final process comprised item selection. The focus of this paper is the identification of the final descriptive system, with the ultimate aim of producing a condition-specific preference-based measure for use in cost effectiveness analysis (the **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (**WAItE**)).

METHODS

There were 2 phases to this research. Qualitative interviews were used to generate a draft item pool and associated response options. One-to-one interviews were conducted to gather in depth information from individual perspectives as to how being overweight or obese impacts upon different

aspects of life (and to minimize any potential social desirability of responses). A focus group with non-treatment seekers was employed to allow for further exploration of and validation of information gathered during individual interviews. Finally, additional focus groups with treatment seekers were conducted to refine the draft of 29 items. Following this, a quantitative survey approach was utilised to reduce the item set and select an appropriate response scale.

Phase 1 – Qualitative study generating the draft 29-Item scale

Participants

Ethical approval for the qualitative study was provided by the University of Leeds, School of Medicine Research Ethics Committee. One-to-one and focus group interviews were conducted with treatment seeking (above normal weight status, aged 11 to 18) and non-treatment seeking adolescents (school sample, varied and unmeasured weight status, aged 11 to 14). We sought the views of a non-treatment seeking sample (the school sample) as the majority of overweight and obese adolescents are not engaged in treatment. By including the views of adolescents regardless of their weight we were as open as possible to the range of relevant issues for this age group.

Consenting adolescents were recruited from three UK-based weight management services and one school. Different parental consenting procedures were employed between the weight management and school samples. For the weight management sample, direct consultation with parents was undertaken and children were included if the parent raised no objection. For the school sample, face-to-face communication with parents was not possible and, in addition to this, the topic area and discussions could be viewed as more sensitive in a school setting compared with a treatment setting where there is parental, peer and staff support readily at hand. Therefore opt-in written parental consent was sought from the school sample.

Interview protocol

All the interviews were undertaken by the same researcher (YO). The qualitative interviews firstly assessed adolescent's own experiences of how weight status affected different aspects of their lives. The second half of the interview assessed issues that were not raised by adolescents in the first section of the interview. Probe questions were used to identify how different aspects of life were affected by weight status in the one-to–one interviews (Table 1). The interviewer asked adolescents about their thoughts on further issues identified from questions in existing weight specific instruments (such as the IWQOL-Kids, Sizing Me Up, OPOI and YQOL-W) (as per section C in Table 1). This provided the opportunity to assess the applicability of the content of existing instruments in the UK and non-US contexts.

Interviews took place in the most convenient location for the adolescent and their family including home, weight management service and University premises. Adolescents were sampled purposively by gender and age (two groups; 11 - 14 year olds and 15 – 18 year olds). Purposeful sampling was undertaken to ensure that the views of the full range of the adolescent population were incorporated. Weight status, using growth reference charts, was recorded for all one-to-one interview participants enrolled in weight management services. Once consent was obtained, height and weight data for the weight management sample were provided by the weight management services. The British 1990 (UK90) growth reference chart [18] was used in the classification of weight status. Table 2 summarises characteristics of the participants involved in the qualitative study.

One-to-one and focus groups interviews

Sixteen one-to-one interviews with adolescents in weight management services (each receiving a £15 retail Love-To-Shop voucher for participating in the study), 2 focus groups with treatment seeking adolescents, and 1 focus group with non-treatment seeking adolescents were undertaken (Table 2). The initial item development involved one-to-one interviews with adolescents in weight management services and a focus group with a school sample of non-treatment seekers. One-to-one interviews were used with adolescents in weight management services as it was thought this would allow participants to feel confident and speak freely about the impact of weight on their lives. A focus group was used in the school sample so that adolescents could deliberate between themselves about the impact of weight status on the lives of adolescents. The school sample focus group interview utilised similar questions as in Table 1. One difference was the use of body shape drawings (line drawings of body shapes that ranged from very thin to obese) that described different body sizes for boys and girls. Adolescents were asked to consider the body size synonymous to an obese weight status and were asked to think about someone they may know who fitted that silhouette (a friend at school or at home, or it could be a relative). They were then asked to think about how that particular individual would be affected by their size. Taking this approach, it was felt that a focus group would be best suited to encourage discussion. Two further focus group interviews with adolescents enrolled in weight management services enabled the refinement of items and response options. Here adolescents were asked to comment on language of the draft items and response options. Unclear language was highlighted and alternative words and phrases were provided by adolescents.

Analysis

The qualitative analysis was carried out in two phases; adolescents' own first-hand experiences were assessed and then the data generated from the interviews that were informed by the existing weight specific instruments (IWQOL-Kids, Sizing Me Up and YQOL-W) was analysed (as per section C in Table

1). The list of dimensions of QoL described in Table 1 section C was generated from the aforementioned existing instruments. If any of the dimensions in this list were not mentioned in the first half of the interview then adolescents were asked if any of the remaining dimensions affected them. This approach enabled a neutral stand point on the first phase analysis of the interviews, enabling the adolescents themselves to raise issues about how their weight might affect them. A neutral stand point was enabled by allowing adolescents to raise whatever issues they felt were important in how their weight affected them (as opposed to asking leading questions such as explaining how their weight affects their physical activity).

The interviews were transcribed verbatim and framework analysis was used to identify themes [19]. This method has been successfully implemented in similar studies [20, 21]. A theme and case-based chart that summarised all the data into a matrix of cases (represented by each row) and themes (represented by each column) was generated. Themes were identified from listening to interviewed recordings and reading through transcripts using an iterative process. Initially three broad themes were identified (physical, social and psychological). These broader themes were then refined as quotes within each theme were re-assessed. For example, the social theme was disaggregated into three themes (non-school based physical activities, school-based physical activities, and barriers to physical activities). The matrix summarised and synthesised the data generated from the interviews whilst retaining the terminology and language used by participants. The analysis and coding were validated by a second reviewer. The second reviewer assessed the generation and refinement of themes and coding of transcripts for the first two interviews and any suggested changes were agreed and finalised before the remainder of the analysis was undertaken. The second reviewer was closely involved throughout the qualitative analysis (e.g. meeting frequently to review and agree the coding of all interviews to ensure consistency) not just for the coding for the first two interviews, given their role as a supervisor. Consensus was reached on all occasions between first and second reviewers,

occasionally after discussion. A draft item set used interview quotations to craft a 29-item instrument. The 29 items (details provided below), grouped into dimensions of QoL [22], covered a variety of obesity specific aspects of QoL including: symptoms, physical function, psychological wellbeing (covering appearance-related items and food-related items), cognitive functioning, social wellbeing, and future prospects. Dimension names and the grouping of items within dimensions, for presenting the 29-item instrument in the survey in Phase 2, was informed by existing literature [22]. This also provided the opportunity for double-checking the suggested grouping of items from the factor analysis. Five-point Likert scales describing frequency (1=Never, 2=Almost Never, 3=Sometimes, 4=Often, 5=Always) and severity (1= Not at all, 2= A little, 3= Quite a bit, 4= A lot, 5= Very much) were applied to each item.

Phase 2 – Quantitative study to identify a reduced item set

Participants and Procedure

Participants were recruited (between October 2011 and January 2012) from three adolescent weight management programmes and via an internet panel (the same procedure was employed for both samples, the only difference being the mode of application of the survey). The survey collected information on background characteristics and perceived weight status. Self-reported weight and height data were collected and weight status determined using growth reference charts [18]. Adolescents who were aged 11-18 and could self-complete the survey were eligible for inclusion. In addition, a sampling quota was used for recruiting the on-line panel, comprising a balanced sample by gender and age. Given the target population, the proportion of adolescents perceiving themselves to be overweight and obese were purposefully oversampled to represent approximately half the total sample. The market research company converted the paper questionnaire to an on-line survey. The online survey included an introductory section providing background info about the study and a consent page for participants and parents (no generic instructions were provided at the beginning of

the survey, but each question, where necessary, had a prompt (e.g. please tick one box or tick)). The company uses a number of different methods to recruit individuals onto their panels (e.g. email and targeting websites). As per their normal procedures incentives were paid to participants (£1-3 shopping vouchers). Data were returned to the research team in an anonymised format. All participants gave written informed consent. Implicit consent from parents was assumed from both samples as parents were present when adolescents chose to participate and gave their consent (for the on-line survey sample, parents were sent the survey and the adolescents would only be able to access the survey if the parent made the choice to allow them to view the survey). Ethical approval was provided by the University of Leeds, School of Medicine Research Ethics Committee (Ref - HSLTLM/11/006).

The study sample was 341 adolescents, 25 adolescents from weight management programmes and 316 from the internet panel (see Table 3). Data from 26 adolescents in the underweight category were excluded from the analysis. This left a final sample of 315 adolescent participants. Ultimately, the new tool was for use in economic evaluation of weight management. If an intervention were successful then adolescents should be moving from above healthy weight to healthy weight. Specifically, we would expect healthy weight individuals to record either no impairment or minimal impairment on items compared to their above healthy weight counterparts, who we would expect to record having some level of impairment. The items in the reduced final item set should therefore be able to clearly distinguish between different weight categories. In order to test this adolescents who were healthy weight were included in the survey so that their responses could be compared with the responses of those above healthy weight.

Analysis

A three step process (a method that has been successfully used in the development of preference based measures from pre-existing disease specific QoL instruments [23-29]) was applied to the 29

items utilising a combination of Rasch analysis [30] and classical psychometric testing. It has been recommended that a sample size of around five to ten participants per item is required in Rasch analysis [31, 32]. Given the 29 item scale a minimum sample size of 290 was sufficient. Psychometric assessments were carried out in SPSS Version 18 [33] whilst Rasch analysis was conducted using the RUMM2030 [34] software package.

Step 1 - Identification of dimensions

In Step 1, factor analysis was used to establish instrument dimensions by identifying the underlying factors that explained patterns of correlation within a set of observed variables. Exploratory Factor Analysis (EFA) was undertaken to establish dimension structure, thus facilitating the grouping of the 29 items into acceptable and justifiable categories. Methodological issues concerning the extraction method, determination of the number of factors to include in the analysis, and type of rotation to use, were also considered. In terms of the factor extraction method, both the maximum likelihood estimation and the Principal Axis Factors method were implemented. The selection of the number of factors was further 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 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 [35]. The factor structure of the scale estimated from the EFA was then tested using confirmatory factor analysis (CFA) (in Amos [33]) [29, 35] (results available upon request).

Step 2 - Item reduction

Rasch models were then applied to each of the dimensions identified in order to exclude poorly performing items [36] in Step 2. Rasch model goodness of fit was assessed in three ways: the Chi-squared (X^2) test statistic was used to assess the differences between the observed and expected responses (P-value for the overall model X^2 statistic should be >0.01). The level of discrimination

amongst different groups of respondents was assessed using the person separation index (PSI); a PSI value of 0.7 or above indicates a well-fitting Rasch model. Fit residuals, providing estimates of the amount of divergence between the expected and observed responses, were assessed and a mean value of approximately zero with a standard deviation approximately equal to one being adequate [29].

In addition, Differential Item Functioning (DIF) was examined to establish whether responses systematically differed across patient characteristics (e.g. for an item asking about physical abilities, boys might select less severe item response options than girls). Gender (male/female) and age (younger adolescents (11-15 years)/older adolescents (16-18 years)) were examined for DIF using item characteristic curves and complementary P-values from ANOVA (under 0.05 indicates that DIF is observed indicating significant deviations between the observed data and the predictions of the Rasch model [37]). Finally, identification of potentially problematic level orderings (i.e. identification of items where responders were unable to distinguish between item-response levels) was sought to ensure the response ordering of health states (that would be included in the preference valuation study in future work) was robust. Disordered item levels could indicate the inability of respondents to distinguish between item levels. 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. Items that did not meet the goodness of fit tests or showed evidence of DIF were not selected in the final item set. The process of fitting Rasch models was repeated (independently on each of the QoL dimensions identified in Step 1) until only well-fitting items remained and the overall item-trait goodness of fit of the model (see below) was non-significant.

Step 3 - Item selection

Step 3 used combined criteria based on Rasch analysis, classical psychometric testing, and data generated from the qualitative interviews to select the final item set. Item selection was predominantly based upon the spread of item levels across the latent space, in order to span the full range of condition severity (appropriate to a wider patient population). The threshold probability curves and item goodness of fit statistics were re-examined. In addition to giving preference to the best overall performance of items across Rasch and psychometric tests, the interview transcripts from the qualitative study (Phase 1) that informed the development of the 29 item instrument were consulted. Transcripts were re-assessed to reflect on the importance of items not excluded after Steps 1 and 2. The importance of items was gauged in two ways: in terms of the number of times participants mentioned the issue under consideration, and the phrasing used in terms of how serious the issue was perceived to be. Item wordings were also reviewed to retain the original phrasing used by adolescents.

Validation

To validate the final item selection, Rasch model goodness of fit statistics were assessed. Additionally, the assumption of uni-dimensionality was tested, i.e. that scale scores could be meaningfully aggregated across the final item set. A test of local dependency was also 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) [36].

RESULTS

Phase 2 – Quantitative study to identify a reduced item set

Step 1 - Identification of dimensions

Table 4 presents the findings from the psychometric assessments. The frequency and severity scales were close in their performance across all of the items. The 29 items could be grouped into seven factors where item 27 displayed disagreement regarding the factor loading between the two scales. Overall, the frequency scale consistently had a lower percentage of responses answering *never* compared to *not at all* across all the 29 items. T-tests showed that items were better at discriminating between participants who were normal weight and overweight (25 items with the frequency scale and 28 items with the severity scale) than between those overweight and obese (five items when either the frequency or severity scales were used) for both scales. The severity scale displayed more items with factor loadings lower than the 0.4 threshold (four items) than the frequency scale (two items). Consequently, the frequency scale was chosen over and above the severity scale, and formed the basis of the remainder of the analyses.

Step 2 - Item reduction

Seven models were estimated independently (F1 to F7) in line with the item groupings identified. The goodness of fit statistics for each of these models is reported in Table 5. Models F2, F3, F6 and F7 did not have to be re-estimated. 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. 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.

Step 3 - Item selection

Table 6 summarises the combined results of psychometric and Rasch analysis undertaken by item. Eight items were excluded from further analysis based on the results of the previous psychometric and Rasch analysis (items 4, 13, 14, 16, 17, 18, 23 and 26 – in italics in the table). None of the items included in model F1 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. In order to ensure that the impact of weight status on all of the 7 dimensions of QoL identified from the factor analysis, one item needed to be selected from each of the identified factor groupings. 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 remaining 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).

Validation

The results of the psychometric and Rasch validation assessments on the final seven item scale appeared to be adequate (see Table 7). The seven items displayed internal consistency, moderate floor effects, and the ability to discriminate between weight categories. Rasch analysis on the final seven items showed that the scale fitted the Rasch model after Bonferroni adjustment ($p<0.01$). 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. Combining the final set of seven items showed a number of items with dis-ordered thresholds. In the main, these statistics showed that the seven items conformed to the underlying assumptions of the Rasch model. The assessment of uni-dimensionality of the seven items was supported (t-test statistic of 0.024, less than the 0.05 threshold value). Similarly the test of local dependency demonstrated that the seven items were independent, despite the fact that the correlated item-total correlations ranged between 0.62 and 0.75. The final Weight specific Adolescent Instrument for Economic-evaluation (WAItE) comprised seven items and a five level frequency response scale.

DISCUSSION

The aim of this study was to develop a weight specific instrument appropriate for undertaking economic evaluation of weight management. Psychometric assessment of the draft 29 items, generated from qualitative interviews, identified seven dimensions of QoL. Using a combination of Rasch analysis, psychometric assessment, and re-visiting the qualitative material, one item was selected from each dimension. Accordingly, the result was a short, 7-item measure based on the views and experiences of adolescent girls and boys aged 11 to 18 years. The involvement of adolescents in the development of the WAItE is in line with the Food and Drug Administration guidelines [38] on patient reported outcome measures. The WAItE was tested against the Rasch model assumptions showing that the WAItE adequately met the various test requirements, that each of the seven items was independent, and that the scores could be aggregated across the seven items meaningfully.

The WAItE therefore has the 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 it is a brief instrument in comparison to

existing weight-specific instruments that have an excess of 20 items [13, 16]. Hence, it is likely to reduce burden in completion and minimise missing data [39]. 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. 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 [16], 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. In terms of the WAItE, adolescents are able to express that exercising may have an impact upon 'feeling tired' or perhaps their ability to 'keep up with others', for example. The WAItE allows respondents to identify the consequences resulting from performing different activities both in the physical function dimension and in the symptoms dimension. One dimension that was identified in the WAItE, 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, despite evidence of a negative relationship between obese weight status and educational attainment and achievement [2]. Furthermore, checks with the end users of the scale in terms of understanding what the items mean and the clarity of the response options, was conducted iteratively over the development of the WAItE.

Study limitations include the narrow age range of the school sample (11-14 years) in the qualitative study. It would have been beneficial to have a second focus group with 15-18 year olds. However, it is worth noting there was a great degree of agreement in terms of the effect of weight status on QoL between the school and treatment seeker samples. Nevertheless, it would be necessary to collect data from the older age group as well in order to assess the level of agreement in this age group. Furthermore, considering that approximately half of the respondents, for the Phase 2 – Quantitative

study, reported their perceived weight status as slightly overweight to slightly underweight, this may have resulted in an under-representation of the views of adolescents with obesity in the QoL descriptive system. Additionally, the majority of the sample for this study was White British and the lack of information on their socio-economic status might have implications on generalisability of findings.

No measure of weight status was taken for the school sample in the Phase 1 - Qualitative study and for the Phase 2 - Quantitative study, self-reported measures of weight and height were collected. The use of objective measures of weight and height can be further investigated in future research. However, a requirement for assessed body weight is one of several barriers to the recruitment of adolescents and young people in research [40]. For example, adolescents are self-conscious of their bodies and concerned about their appearance, this may then lead to a hindrance in recruitment. This has consequent implications for sample representativity.

The parallel analysis used to inform the factor structure in addition to the CFA was not conducted using polychoric correlations, a limitation given that the analysis would assume the categorical data was interval-scaled. This may have impacted on the loading for each factor, though it is difficult to predict in which direction this may lie (i.e. under or over-estimation of the factor loadings). Empirical testing comparing of the performance of maximum likelihood and robust categorical least squares methodology, for estimating confirmatory factor analysis models, found that the maximum likelihood method was favourable where items had five or more associated response categories [41].

In terms of the final seven items that were chosen, the following were observed. There was a large number of 'never' responses observed for some of the items. This may be because approximately

45% of the sample for the quantitative study was classified as normal weight after adjusting for age and gender. Analysis of the final seven selected items showed that a number of items in the final set displayed disordered thresholds. There is a debate surrounding what to do when disordered thresholds are observed and there is no clear consensus on this issue. Sometimes it is argued that categories need to be collapsed to address this issue [42, 43]. It has recently been shown empirically that reversed thresholds do not violate the key assumptions of the item response theory and that collapsing categories due to reversed thresholds should be carefully considered [44, 45]. It has been argued that disordered thresholds have been shown to be a consequence of (at least) one category not being the most likely category along the trait continuum and thus, whether threshold parameters are ordered or disordered depends solely on the number of respondents endorsing each response category [45]. Particularly of note, a large number of 'never' responses were observed in the current study and could have affected the threshold ordering. As such the levels for the final seven items were not collapsed. Sensitivity of the measure to changes in BMI will need to be assessed as this is an important determinant of the validity of the new measure. Future research will involve the further validation of the WAItE in terms of further assessments of the psychometric properties, such as validity and reliability. A valuation study, to assign weight-specific preference values to states described by the WAItE to generate QALYs will also be necessary.

In the face of high demand on health care resources, the identification of the most cost effective interventions is crucial. There is currently little consensus on the most economically effective management and prevention strategies in adolescent obesity. The WAItE could 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 could serve as an important tool to help guide decisions about allocating scarce resources across competing weight management programmes.

COMPLIANCE WITH ETHICAL STANDARDS

This study was funded by a National Institute for Health Research (NIHR) doctoral fellowship awarded to Yemi Oluboyede (DFR/2009/02/101).

CONFLICT OF INTEREST STATEMENT

Yemi Oluboyede declares that she has no conflict of interest. Claire Hulme declares that she has no conflict of interest. Andrew Hill declares that he has no conflict of interest

ETHICAL APPROVAL

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

1. Ogden, C., M. Carroll, B. Kit, and K. Flegal, *Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010.* Journal of the American Medical Association, 2012. **307**(5): p. 483-490.
2. Griffiths, L., T. Parsons, and A. Hill, *Self-esteem and Quality of Life in Obese Children and Adolescents: a Systematic Review.* International Journal of Paediatric Obesity, 2010. **5**(4): p. 282-304.
3. Tsiros, M., T. Olds, J. Buckley, P. Grimshaw, L. Brennan, J. Walkley, A. Hills, P. Howe, and A. Coats, *Health-related quality of life in obese children and adolescents.* International Journal of Obesity, 2009. **33**(387-400).
4. Wijga, A., S. Scholtens, W. Bemelmans, J. de Jongste, M. Kerkhof, M. Schipper, E. Sanders, J. Gerritsen, B. Brunekreef, and H. Smit, *Comorbidities of obesity in school children: a cross-sectional study in the PIAMA birth cohort.* BMC Public Health 2010. **10**(1).
5. Ho, M., S. Garnett, L. Baur, T. Burrows, L. Stewart, M. Neve, and C. Collins, *Effectiveness of lifestyle interventions in child obesity: systematic review with meta-analysis.* Pediatrics, 2012. **130**(6): p. 1647-71.
6. National-Institute-for-Clinical-Excellence, *Guide to the Methods of Technology Appraisal*. April 2013.
7. Drummond, M., G. Stoddart, and G. Torrance, *Methods for the Economic Evaluation of Health Care Programs*. 1996, Oxford, United Kingdom: Oxford University Press.
8. National-Institute-for-Clinical-Excellence. *Managing overweight and obesity among children and young people: lifestyle weight management services. Review 1: Effectiveness and cost effectiveness of lifestyle weight management services for children and young people.* . April 2013; Available from: http://www.nice.org.uk/guidance/ph47/evidence/review-of-effectiveness-and-cost-effectiveness-430360093.
9. Brazier, J., J. Ratcliffe, J. Salomon, and A. Tsuchiya, *Measuring and Valuing Health Benefits for Economic Evaluation*. 2007: Oxford University Press.
10. Brazier, J., R. Kolotkin, R. Crosby, and G. Williams, *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, 2004. **7**(4): p. 490-498.
11. Kolotkin, R., S. Head, M. Hamilton, and C. Tse, *Assessing impact of weight on quality of life.* Obesety Research, 1995. **3**: p. 49-56.
12. Ravens-Sieberer, U., M. Redegeld, and M. Bullinger, *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, 2001. **25 Suppl 1**: p. S63-5.
13. Kolotkin, R., M. Zeller, A. Modi, G. Samsa, N. Quinlan, J. Yanovski, S. Bell, D. Maahs, D. de-Serna, and H. Roehrig, *Assessing weight-related quality of life in adolescents.* Obesity, 2006. **14**(3): p. 448-57.
14. Moorehead, M., E. Ardelt-Gattinger, H. Lechner, and H. Oria, *The validation of the Moorehead-Ardelt Quality of Life Questionnaire II.* Obesity Surgery, 2003. **13**(5): p. 684-92.
15. Zeller, M. and A. Modi, *Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up.* Obesity, 2009. **Jun 17**(6): p. 1171-7.
16. Morales, L., T. Edwards, Y. Flores, L. Barr, and D. Patrick, *Measurement properties of a multicultural weight-specific quality-of-life instrument for children and adolescents.* Quality of Life Research, 2011. **20**(2): p. 215-24.
17. Doyle, S., *Development of a new health related quality of life instrument for use in paediatric obesity*, in *18th European Congress on Obesity (ECO)*, Obesity-Reviews-Conference, Editor. 2011: Istanbul Turkey. p. 77-78.

18. Cole, T., J. Freeman, and M. Preece, *Body mass index reference curves for the UK, 1990.* Archives of Disease in Childhood, 1995. **73**: p. 25-29.
19. Ritchie, J., L. Spencer, and O. O'Connor, *Carrying out qualitative analysis* in *Qualitative research practice: a guide for social science students and researchers*, J. Ritchie and J. Lewis, Editors. 2003, Sage: London. p. 219-262.
20. Carlton, J., *Identifying Potential Themes for the Child Amblyopia Treatment Questionnaire.* Optom Vis Sci 2013. **90**(8): p. 867-73.
21. Stevens, K. and S. Palfreyman, *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, 2012. **15**(8): p. 991-8.
22. Fitzpatrick, R., C. Davey, M. Buxton, and D. Jones, *Evaluating patient-based outcome measures for use in clinical trials.* Health Technology Assessment, 1998. **2**(14).
23. Rowen, D., J. Brazier, T. Young, S. Gaugris, B. Craig, M. King, and G. Velikova, *Deriving a preference-based measure for cancer using the EORTC QLQ-C30.* Value in Health, 2011. **14**(5): p. 721-31.
24. Rowen, D., B. Mulhern, S. Banerjee, B. van Hout, T. Young, M. Knapp, S. Smith, D. Lamping, and J. Brazier, *Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy.* Value in Health, 2012. **15**(2): p. 346-56.
25. Yang, Y., Brazier, JE., Tsuchiya, A. and Young, TA., *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, 2011. **31**(2): p. 281- 291.
26. Yang, Y., J. Brazier, A. Tsuchiya, and K. Coyne, *Estimating a preference-based single index from the overactive bladder questionnaire.* Value in Health, 2009. **12**(1): p. 159-166.
27. Young, T., D. Rowen, J. Norquist, and J. Brazier, *Developing Preference-Based Health Measures: Using Rasch Analysis to Generate Health State Values.* Quality of Life Research, 2010. **19**(6): p. 907-17.
28. Young, T., Y. Yang, J. Brazier, and A. Tsuchiya, *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, 2011. **31**(1): p. 195-210.
29. Young, T., Y. Yang, J. Brazier, A. Tsuchiya, and K. Coyne, *The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis.* Qual Life Res, 2009. **18**(2): p. 253-65.
30. Rasch, G., *Probabilistic models for some intelligence and attainment tests*. 1960, Chicago: University of Chicago Press. Reprinted 1980.
31. Linacre, J. *Sample Size and Item Calibration (or Person Measure) Stability*. Rasch measurement and transactions, 1994. **7**, 328.
32. Linacre, J., *Understanding Rasch measurement: Optimizing rating scale category effectiveness.* Journal of Applied Measurement 2002. **3**(1): p. 85-106.
33. SPSS-Inc., *PASW Statistics for Windows, Version 18.0*. Released 2009, SPSS Inc: Chicago.
34. Andrich, D., B. Sheridan, and G. Luo, *Rumm 2030: Rasch Unidimensional Measurement Models [computer software]*. 2010, RUMM Laboratory Perth, Western Australia: http://www.rummlab.com.au/.
35. Tabachnick, B.G. and L.S. Fidell, *Using multivariate statistics (5th ed.)*. 2007: Upper Saddle River, NJ: Pearson Allyn & Bacon.
36. Hobart, J. and S. Cano, *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods*, in *Health Technology Assessment*. 2009.
37. RUMM-Laboratory-Pty-Ltd., *Interpreting RUMM2030 Analysis Manual* 2009.
38. Food-and-Drug-Administration *Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance*. 2009.

39. Ravens-Sieberer, U., M. Erhart, N. Wille, R. Wetzel, J. Nickel, and M. Bullinger, *Generic Health-related Quality of life assessment in children and adolescents: Methodological considerations.* Pharmacoeconomics, 2006. **24**(12): p. 1199-1220.
40. Steinbeck, K., L. Baur, C. Cowell, and A. Pietrobelli, *Clinical research in adolescents: challenges and opportunities using obesity as a model.* Int J Obes, 2008. **33**(1): p. 2-7.
41. Rhemtulla, M., P.É. Brosseau-Liard, and V. Savalei, *When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions.* Psychological Methods, 2012. **17**(3): p. 354-373.
42. Andrich, D., *An Expanded Derivation of the Threshold Structure of the Polytomous Rasch Model That Dispels Any ''Threshold Disorder Controversy''.* Educational and Psychological Measurement, 2013. **73**(1): p. 78-124.
43. Luo, G., *The relationship between the Rating Scale and Partial Credit Models and the implication of disordered thresholds of the Rasch models for polytomous responses.* Journal of Applied Measurement, 2005. **6**(4): p. 443-55.
44. Adams, R., M. Wu, and M. Wilson, *The Rasch rating model and the disordered threshold controversy.* Educational and Psychological Measurement, 2012. **72**(4).
45. Wetzel, E. and C. Carstensen, *Reversed Thresholds in Partial Credit Models: A Reason for Collapsing Categories?* Assessment, 2014. **21**(6): p. 765-774.

ACKNOWLEDGEMENTS

We would like to acknowledge the advice and support of the following individuals: Cathy Brennan, Jenny Hewison, Donna Lamping, Christopher McCabe, David Meads, Jennifer Roberts, Katherine Stevens, Alan Tennant, Aki Tsuchiya and members of Academic Unit of Health Economics, University of Leeds. We would like to thank all the young people who took part in the research and the parents and staff who supported this research. The usual disclaimer applies.

The work presented here is part of a National Institute for Health Research (NIHR) funded fellowship project awarded to the first author.

The WAItE is available from the corresponding author upon request.

TABLES

Table 1: Summary of questions included in the topic guide for Phase 1 qualitative study

<table>
<tr><th colspan="2">A. Background and personal circumstances</th></tr>
<tr><td>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, activities, local services)</td><td>Comparing yourself to others your age, where do you fall size wise:
average weight, somewhat overweight or very overweight
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?</td></tr>
<tr><th colspan="2">B. Dimensions of QoL affected by weight – from <u>own life</u></th></tr>
<tr><td>Can you describe the types of things you do on a typical day:
School / work days
Non-school / work days</td><td>Thinking back to the activities you just spoke about, tell me about how your weight might affect these activities</td></tr>
<tr><th colspan="2">C. Dimensions of QoL that obesity affects – from the <u>literature</u></th></tr>
<tr><td>“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”……..</td><td>Physical activity
School
Psychological health
Body esteem - The way you see yourself
Relationships
Social functioning
Eating
Future</td></tr>
</table>

Table 2: Characteristics of the adolescents participating in the interviews

Participants	Baseline Body Mass Index – Standard Deviation Score (BMI-SDS)*
1-to-1 interviews (n=16)	
Girls aged 11-14 years old (n=6)	2.80
	2.43
	3.20
	2.27
	3.56
	2.96
Girls aged 15-18 years old (n=4)	3.16
	3.38
	2.62
	NA (on waiting list)
Boys 11-14 years old (n=4)	2.87
	2.86
	2.75
	1.53
Boys 15-18 years old (n=2)	3.08
	3.67
Focus group interviews (n=3)	
Boys aged 11-18 (enrolled in a weight management service) (n=5; 3 aged 15-18 & 2 aged 11-14)	NA
Girls aged 11-18 (enrolled in a weight management service) (n=5; 3 aged 15-18 & 2 aged 11-14)	NA
Mixed, aged 11-14 years (school sample) (n=2 girls & 3 boys)	NA

***see Cole et al. [18] for further explanation**

Table 3: Participant characteristics N=315*

Characteristics		N	%
Gender	Male	157	49.84
	Female	158	50.16
Age	10[a]	1	0.32
	11	32	10.16
	12	37	11.75
	13	28	8.89
	14	23	7.30
	15	32	10.16
	16	36	11.43
	17	56	17.78
	18	70	22.22
Perceived weight status	Very overweight	26	8.25
	Moderately overweight	84	26.67
	Slightly overweight	55	17.46
	About the right weight	117	37.14
	Slightly underweight	28	8.89
	Moderately underweight	4	1.27
	Very underweight	1	0.32
BMI	Normal weight [b]	152	48.25
	Overweight [b]	51	16.19
	Obese [b]	112	35.56
In full-time education	Yes	299	94.92
	No	16	5.08
Ethnicity	White	256	81.27
	Mixed/dual heritage	9	2.86
	Asian or Asian British	25	7.94
	Black or Black British	15	4.76
	Chinese	5	1.59
	Other	3	0.95
	Preferred not to say	2	0.63
Geographical location	England	275	87.30
	Scotland	13	4.13
	Wales	16	5.08
	Northern Ireland	8	2.54
Self assessed health status	Excellent	76	24.13
	Good	118	37.46
	Fair	85	26.98
	Poor	32	10.16
	Very poor	4	1.27

* Data from 26 adolescents in the underweight category are excluded

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

[b] Underweight <= 2nd centile, Normal weight >= 2nd centile or < than 85th centile, Overweight >= 85th centile or < 95th centile and Obese >= 95th centile [18]

Table 4: Summary of psychometric analysis results by response scale N=315 (Step I - Factor analysis to establish dimension structure)

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 factor 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 factor estimation (PAF) Factor loading[c]
17	41.27	<0.01	0.19	0.73	F1	**0.39**	44.44	<0.01	0.21	0.74	F1	**0.33**
21	41.90	<0.01	0.25	0.73	F1	0.59	46.35	<0.01	0.70	0.77	F1	0.64
22	58.10	<0.01	0.39	0.79	F1	0.62	61.59	<0.01	0.32	0.79	F1	0.67
23	54.29	<0.01	0.90	0.77	F1	**0.39**	59.05	<0.01	0.59	0.77	F1	0.46
24	60.00	0.02	0.45	0.74	F1	0.59	62.86	0.02	0.48	0.71	F1	0.72
25	64.13	<0.01	0.27	0.76	F1	0.64	67.62	<0.01	0.22	0.72	F1	0.72
26	40.95	<0.01	0.20	0.77	F1	0.77	42.86	<0.01	0.64	0.74	F1	0.77
28	64.76	<0.01	0.05	0.75	F1	0.56	66.35	<0.01	0.12	0.75	F1	0.53
29	50.79	0.28	0.02	0.77	F1	0.69	54.92	<0.01	0.11	0.75	F1	0.67
8	47.94	<0.01	0.18	0.75	F2	0.66	59.05	<0.01	0.29	0.78	F2	-0.87
9	47.94	<0.01	0.12	0.75	F2	0.73	54.92	<0.01	0.19	0.80	F2	-0.85
10	56.51	0.05	0.15	0.71	F2	0.58	63.49	<0.01	0.35	0.73	F2	-0.60
1	47.94	0.21	0.43	0.63	F3	0.62	60.63	0.12	0.15	0.66	F3	0.42
2	37.46	0.07	0.50	0.72	F3	0.81	45.71	<0.01	0.56	0.72	F3	0.75
3	29.84	0.07	0.28	0.70	F3	0.90	**37.14**	0.02	0.62	0.72	F3	0.79
4	37.46	0.13	0.03	0.69	F3	0.42	47.30	<0.01	0.06	0.73	F3	**0.32**
13	69.21	0.02	0.15	0.66	F4	0.60	71.75	<0.01	0.46	0.67	F4	0.64
14	78.41	0.12	0.40	0.63	F4	0.44	80.95	0.07	0.25	0.66	F4	0.54
15	72.70	<0.01	0.74	0.68	F4	0.68	77.14	<0.01	0.06	0.72	F4	0.82
16	66.35	<0.01	0.87	0.73	F4	0.67	69.84	<0.01	0.63	0.73	F4	0.58
11	60.32	0.03	0.75	0.67	F5	-0.79	68.89	0.03	0.50	0.65	F5	-1.00
12	52.70	0.03	0.85	0.67	F5	-0.93	59.37	<0.01	0.90	0.65	F5	-0.78
18	50.48	0.07	0.44	0.63	F5	-0.42	51.75	<0.01	0.78	0.65	F5	**-0.27**
19	61.27	<0.01	0.31	0.84	F6	-0.62	65.71	<0.01	0.66	0.86	F6	-0.53
20	60.32	<0.01	0.32	0.82	F6	-0.73	65.08	<0.01	0.51	0.86	F6	-0.56
5	69.84	<0.01	0.04	0.73	F7	0.57	77.14	<0.01	0.09	0.73	F7	-0.46
6	66.98	<0.01	0.20	0.71	F7	0.52	75.87	<0.01	0.06	0.72	F7	-0.44
7	76.51	0.06	0.03	0.65	F7	0.70	81.27	0.09	0.03	0.63	F7	-0.54
27	64.76	<0.01	0.36	0.82	F6	-0.62	68.25	<0.01	0.34	0.86	F1	**0.33**

[a] Floor effect = *Never* or *Not at all* less than 40% of responses (e.g. when less than 40% of the sample fall into these categories). [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.

Table 5: Rasch model goodness if fit by dimension N=315* (Step II - Using Rasch analysis to exclude items)

FREQUENCY SCALE							
Model (N)	**X^2 Goodness of fit**	**Degrees of freedom (X^2)**	**P-value (X^2)**	**Person Separation Index (PSI) with extremes (No extremes)**	**Cronbach's Alpha with extremes (No extremes)**	**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	<0.01	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	<0.01	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 the Rasch analysis excludes individuals with extreme scores (individuals responding 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: Summary of Rasch analysis results by item N= 315 (Step II cont. - Using Rasch analysis to exclude items)*

Items	Factor loading	Disordered item levels	DIF[a] by Sex	DIF[a] by Age	Exclude item
Factor 1 (F1)					
Item 17. I worry about my health in the future [b]	*0.39*	*No*	*No*	*No*	*Yes*
Item 21. I feel uncomfortable or embarrassed getting changed in front of others	0.59	Yes	No	No	
Item 22. I feel uncomfortable or embarrassed shopping for clothes [c]	**0.62**	**Yes**	**No**	**No**	
Item 23. I feel uncomfortable or embarrassed meeting new people [b]	*0.39*	*Yes*	*No*	*No*	*Yes*
Item 24. I feel uncomfortable or embarrassed eating in front of others	0.59	Yes	No	No	
Item 25. I feel unhappy because I can't eat what I want	0.64	Yes	No	No	
Item 26. I feel unhappy about the way I look [b]	*0.77*	*No*	*Yes*	*Yes*	*Yes*
Item 28. I feel disappointed because clothes aren't made in the size I need	0.56	Yes	No	No	
Item 29. I struggle to keep in control of what I eat	0.69	Yes	No	No	
Factor 2 (F2)					
Item 8. I struggle to keep up with others when doing physical activity	0.66	*No*	No	No	
Item 9. I struggle to keep up with others when I play sports	0.73	No	No	No	
Item 10. I avoid doing things like running, cycling, swimming or playing sports [c]	**0.58**	**No**	No	No	
Factor 3 (F3)					
Item 1. I have body pain / ache	0.62	No	No	No	
Item 2. I get low energy	0.81	No	No	No	
Item 3. I get tired [c]	**0.90**	**No**	**No**	**No**	
Item 4. I get out of breath [b]	*0.42*	*No*	*No*	*Yes*	*Yes*
Factor 4 (F4)					
Item 13. I get treated differently at school, such as being teased or picked-on or left out [b]	*0.60*	*Yes*	*No*	*Yes*	*Yes*
Item 14. I get treated differently at home, such as being teased or picked-on or left out [b]	*0.44*	*Yes*	*No*	*No*	*Yes*
Item 15. People treat me differently when I go out [c]	**0.68**	**No**	**No**	**No**	
Item 16. I avoid playing / hanging out or socialising with others [b]	*0.67*	*No*	*Yes*	*Yes*	*Yes*

Items	Factor loading	Disordered item levels	DIF[a] by Sex	DIF[a] by Age	Exclude item
Factor 5 (F5)					
Item 11. I struggle to do as well as others at school	-0.79	No	No	No	
Item 12. I struggle to concentrate on school / college work [c]	**-0.93**	**No**	**No**	**No**	
Item 18. I worry about the type of job/career I will be able to have [b]	*-0.42*	*Yes*	*No*	*No*	*Yes*
Factor 6 (F6)					
Item 19. I feel angry or annoyed because I am unable to do the same things as others	-0.62	No	No	No	
Item 20. I feel frustrated because I am unable to do the same things as others	-0.73	No	No	No	
Item 27. I feel unhappy because I am unable to do the same things as others [c]	**-0.62**	**No**	**No**	**No**	
Factor 7 (F7)					
Item 5. I struggle to keep up when I am walking around with others [c]	**0.57**	**No**	**No**	**No**	
Item 6. I struggle when I am going up stairs	0.52	No	No	No	
Item 7. I struggle to reach or bend down	0.70	No	No	No	

*Item 23 was excluded from the Rasch analysis due to very low factor loading. Item-threshold probability curves (a plot of the probability of being in each item level across the latent QoL scale) were examined to identify problematic level orderings and highlight disordered item levels

[a]DIF – Differential Item Functioning were examined using item-characteristic curves and item-by-characteristic ANOVA statistics

[b] Eight items in italics could be excluded from further analysis based on the results of the psychometric assessments and Rasch analysis

[c]. The final seven items that were selected are in bold

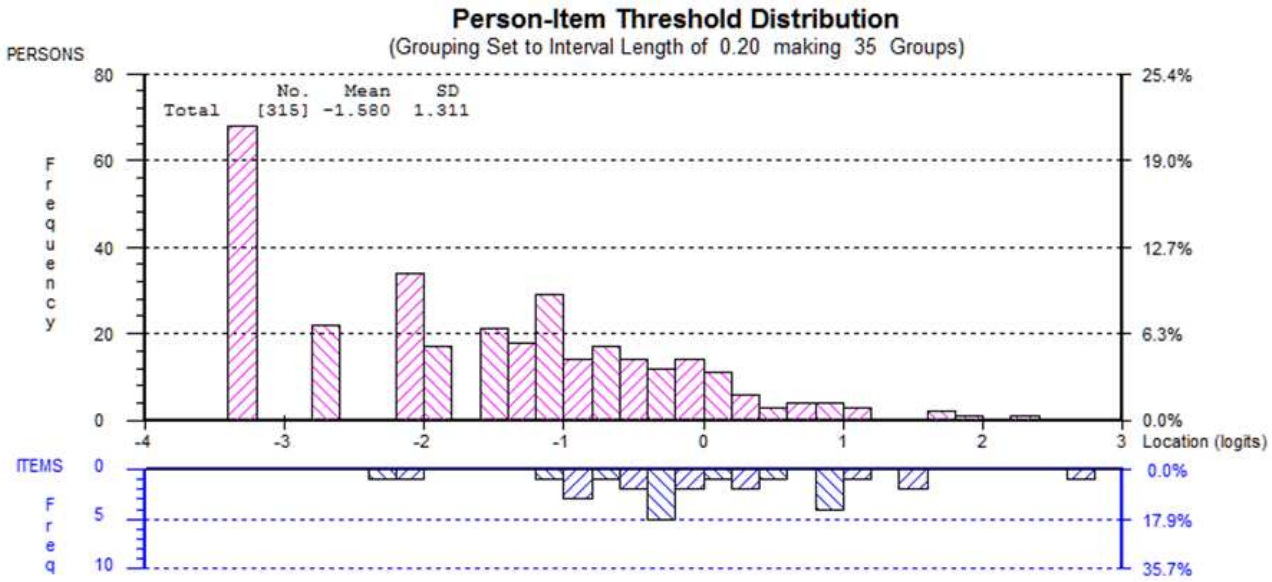
Table 7: Psychometric & Rasch analyses on the final 7 item instrument

Item No.	PSYCHOMETRIC ANALYSIS (N=315)						RASCH ANALYSIS (N=247*)		
	Mean (SD)	Total No. 'Never' responses	Total No. 'Always' responses	Corrected item-total correlation	Cronbach's Alpha if item deleted	Factor loading	FitResid (SE)	ChiSq (P-value)	Disordered thresholds
3. I get tired	2.54 (1.23)	26	16	0.62	0.87	0.89	0.875 (0.08)	3.78 (0.44)	No
5. I struggle to keep up when I am walking around with others	1.54 (0.95)	152	3	0.67	0.87	0.57	-1.76 (0.08)	9.3 (0.05)	Yes
10. I avoid doing things like running, cycling, swimming or playing sports	1.96 (1.27)	110	15	0.68	0.87	0.58	-0.06 (0.07)	4.53 (0.34)	Yes
12. I struggle to concentrate on school / college work	1.93 (1.16)	98	9	0.64	0.87	-0.93	1.14 (0.07)	7.23 (0.12)	Yes
15. People treat me differently when I go out	1.48 (0.90)	161	4	0.64	0.87	0.68	-0.56 (0.09)	6.44 (0.17)	Yes
22. I feel uncomfortable or embarrassed shopping for clothes	1.91 (1.26)	115	16	0.72	0.86	0.62	-0.56 (0.07)	4.56 (0.34)	Yes
27. I feel unhappy because I am unable to do the same things as others	1.72 (1.15)	136	13	0.75	0.86	-0.62	-1.98 (0.07)	13.56 (0.01)	Yes
Total score		**All (N=315)**		**Norm_WT (N=152)**		**Over_WT (N=51)**		**Obese (N=112)**	
Mean (SD)		13.08 (6.12)		10.68 (4.13)		14.37 (7.48)		15.75 (6.47)	
Cronbach's Alpha		0.88							
P-value (from t-test for discrimination between Norm_WT & Over_WT)			<0.01						
P-value (from t-test for discrimination between Over_WT & Obese)			0.26						
P-value (from t-test for discrimination between Normal_WT & Obese)			<0.01						

*Sample size excluding individuals with extreme values N=68 ie. Individuals choosing 'Always' across all seven items. There are a number of items with dis-ordered thresholds but this is likely due to the high number of 'Never' responses.

SUPPLEMENTARY MARTIAL

Figure 1: 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.