Article:
Oluboyede, Y, Smith, AB, Hill, A orcid.org/0000-0003-3192-0427 et al. (1 more author) (2019) The weight-specific adolescent instrument for economic evaluation (WAItE): psychometric evaluation using a Rasch model approach. Quality of Life Research, 28 (4). pp. 969-977. ISSN 0962-9343

https://doi.org/10.1007/s11136-018-2074-2

Title: The weight specific adolescent instrument for economic evaluation (WAItE): Psychometric evaluation using a Rasch model approach

Authors: Yemi Oluboyede[1], Adam B Smith[2] Andrew Hill[3], Claire Hulme[3]

- [1] Institute for Health and Society, Newcastle University, Newcastle Upon Tyne, UK
- [2] York Health Economics Consortium Ltd, University of York, York, UK
- [3] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

Corresponding author:

Yemi Oluboyede: Institute of Health & Society, Newcastle University, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne, NE2 4AX. Tel: +44 (0) 191 208 7349. Fax: +44 (0) 191 208 6043. E-mail: yemi.oluboyede@ncl.ac.uk

Key words: Obesity; Quality of Life; Economic evaluation; Adolescents; Condition-specific measure; Rasch analysis; Adults

Running title: A Rasch model approach to validate the WAItE

The WAItE is available from the corresponding author upon request

INTRODUCTION

Paediatric obesity is of global concern currently. Children and adolescents who are above healthy weight are more likely to become overweight or obese adults and it is well recognised that obesity has a negative impact on health-related quality of life (HRQoL) [1,2]. Obesity in adulthood adds to the burden on healthcare budgets through higher risks of morbidity, disability and premature mortality [1]. Dietary and lifestyle interventions are the main approaches to the treatment of paediatric obesity (Ho et al., 2012), however, policy-makers increasingly require evidence of cost-effectiveness. In the United Kingdom (UK) and elsewhere, the recommended method of cost-effectiveness analysis is the quality-adjusted life-year (QALY) [3], typically derived from a generic health-related preference-based measure (PBM). There are a number of well-established weight-related HRQoL instruments for younger respondents (e.g. KINDL-Obesity module [4]; Impact of Weight on Quality of Life–Kids version (IWQOL-Kids) [5]; Moorehead-Ardelt Quality of Life Questionnaire II (M-A-QoL Q) [6]; Sizing Me Up [7]; Youth Quality of Life – Weight (YQOL-W) [8]). However, there is no weight-specific preference based measure for adolescents with obesity. This is needed as preference values can be derived for use in the QALY calculation [9]. Accordingly, the **W**eight-specific **A**dolescent **I**nstrumen**t** for **E**conomic-evaluation (**WAItE**) was developed for adolescents living with obesity. The WAItE is a short, 7-item measure which was developed based on the views and experiences of UK adolescent girls and boys aged 11 to 18 years. Preliminary psychometric assessments on the final set of seven items comprising the WAItE have been encouraging [10]. However, further psychometric investigation is necessary.

Evidence of measurement properties is critical for the field of patient-reported outcomes as use of unsuitable or poor quality outcome measurement instruments may introduce bias. Rasch analysis can be used in the evaluation of the psychometric properties of new and existing instruments. Few of the existing weight-related tools have employed Rasch analysis in their assessment of measurement properties [11]. Approaches most frequently used in instrument development and the assessment of psychometric properties rely on statistical procedures based on Classical Test Theory (CTT). However, two major conceptual limitations of CTT have been pointed out: the lack of an explicit ordered continuum of items that represent a unidimensional construct, and the lack of additivity of rating scale data [12]. Rasch analysis does not suffer from the aforementioned limitations, but instead facilitates examination of the hierarchical structure, unidimensionality and additivity of HRQOL measures.

As is the case for adolescents, there currently exists no weight-specific preference based measure (PBM) for adult with obesity. The resources and time required to create such a tool are significant. Therefore, if there is evidence to support use of the WAItE for the adult population via assessments of psychometric properties, this will diminish the need for the development of a new instrument and the resource implications attached to this. The aims of the present study therefore were: to assess the performance of the WAItE in a sample of adolescents with obesity engaged in weight management and to assess the applicability and validity of the WAItE in a general adult sample.

METHODS

Data: Participants and procedures

Adolescent sample

Adolescents (females n= 155 & males n= 123; mean (SD) age = 13.3 yrs (1.7 yrs) and 13.1 yrs (1.7 yrs) respectively) were enrolled on two weight management programmes in north of England between 2012 and 2015 (the More Life [13] and Watch It [14] weight management programmes). Both programmes were multicomponent lifestyle interventions (i.e. included educational, dietary and physical activity components). Adolescents came from all over the United Kingdom through a range of sources, including self/parental referral, medical referral, or referral from social services, primary care trusts (PCTs) or educational organizations. Weight status and acceptance into the weight management programme was contingent on having an age and gender adjusted body mass index (BMI) indicating overweight or obesity [15]. In the main, health screening was performed by the family general practitioner [16]. All adolescents were eligible for inclusion in the study unless the staff delivering the weight management intervention indicated otherwise (e.g. unable to self-complete the questionnaire due to learning difficulties). Staff explained to families that completion of the WAItE was optional and was administered at two time points to consenting participants: baseline (T1) and at the end of the programme (follow-up T2). As per the consenting procedures employed within their own organisations, firstly implicit consent from all parents was obtained by weight management staff as part of the baseline face-to-face meeting with families. After that adolescents who chose to participate and gave their consent were given the opportunity to complete the WAItE at the two time points. Details regarding the weight and height of each study participant were obtained from the records kept by the weight management service and was accessed after patents and adolescents gave consent for the records to be shared with the research team. In the main, data were inputted by weight management staff including data on descriptive characteristics, weight status and response to the WAItE and an anonymised database was then provided. No identifiable information was sent to researchers.

Adult sample

An adult sample (females n= 236 & males n= 231; mean (SD) age = 41.2 yrs (13.9 yrs) & 44.3 yrs (14.3 yrs) respectively) completed a web-based survey incorporating an electronic version of the WAItE in 2012. Participants were recruited from a consumer panel. All were over 18 years and recruitment was based on quotas in terms of gender and age in order to obtain a balanced sample of respondents. Weight status of the adult sample were as follows: mean BMI = 27.8, from which 33.6% were classified as overweight and 25.1% with obesity. After obtaining consent from participants, questions on descriptive characteristics, self-report weight and height and the WAItE instrument were administered. Participants completing the survey were given a nominal payment of £1.75 by the survey company if they fully completed the survey.

Ethical approval was provided by the University of Leeds, School of Medicine Research Ethics Committee for both the adolescents and adult studies (Ref: HSLTLM/11/049).

Measures The Weight-specific Adolescent Instrument for Economic-evaluation (WAItE) was developed in conjunction with adolescents living in the UK. Adolescents' views were crucial to the development of the content of the WAItE in order to focus on aspects of life affected by weight that were important to them. There were 2 phases to the development of the WAItE and the study by Oluboyede et al. provides details of this [10].

The WAItE comprises seven items: 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 and 7) People treat me differently when I go out. There was a five-level response scale: Never, Almost never, Sometimes, Often and Always.

116 ***Analysis***

117 A Rasch analysis was undertaken using *Winsteps* version 3.81.1 software [17]. Rasch models [18,19]
118 are a family of probabilistic logistic models which map item difficulty or location, person measure or
119 score along the same latent trait. The Partial Credit Model (PCM) [19] was applied to the data. This is
120 a Rasch model for ordinal items and is appropriate for analysing polytomous data where response
121 categories are reversed (i.e. problematic level orderings where responders find it difficult to distinguish
122 between item response levels. In the context of Rasch analysis response categories are reversed in
123 situations in which the scale locations of incremental item threshold parameters do not monotonically
124 increase) or differ across items. The following steps were employed in the analysis:
125
126 1. Category disordering was assessed through an analysis of the response categories for each item.
127 The assumption within the model is that the level of latent trait increases monotonically with
128 response categories for each item. Category disordering occurs when this monotonic
129 relationship breaks down and response categories may be combined to overcome this problem.
130 Disordering may occur where the number of responses per category is low. Therefore, the
131 number of responses <10 were noted for each item category.
132
133 2. Secondly, item fit to the Rasch model was evaluated. The most commonly used statistics to
134 determine item fit are the infit and outfit mean squares which are Chi-squared statistics divided
135 by the degrees of freedom. The expected value of the mean squares is 1. Mean squares greater
136 than 1 indicate misfit to the model, whereas values less than 1 indicate overfit. A range of 0.7
137 to 1.3 is usually used to assess fit [20].
138
139 3. A principal components analysis was subsequently applied to the residuals to determine
140 whether the domains constituted a unidimensional structure, i.e. whether there were any
141 additional dimensions present. An eigenvalue <2 for the first contrast, i.e. once the variance
142 explained by the Rasch structure has been factored out, and >50% of the variance explained by
143 the Rasch structure are indicative of a unidimensional structure [17,21].

4. Uniform Differential Item Functioning (DIF) was assessed to determine whether the items performed equally across gender (male/female) and age group (2 levels for the adolescent's sample (age 11-14 and 15-18) and 3 for the adults (age 18-34, 35-54, and 54+). The Welch t-test was used to evaluate DIF: item location parameters were estimated separately for a reference group and focal group(s) through logistic regression. The difference between these estimates was then tested for statistical significance [22]. The Bonferroni correction was applied to account for multiple testing ($p \leq 0.01$ after adjustment). A criterion of a difference between item location estimates of ≤ 0.5 logits was also used to evaluate DIF [23]. The impact of any DIF was evaluated by estimating the person measures separately comparing those derived from the entire sample with those derived using items displaying DIF.

Steps 1-4 were repeated for the two datasets from adolescent's responses, as well as the adult dataset. The difference between item locations for the two time points in the adolescent's datasets was used to evaluate the stability of the item location estimates: a difference <0.5 logits was deemed to be evidence of item stability. The change in person measures over time was also evaluated for the adolescent's dataset using a paired t-test. Cronbach's alphas were derived as a measure of internal reliability (>0.7 indicating good internal reliability)

RESULTS

Adolescent sample

Cronbach's alpha was 0.80 for the combined (T1 and T2) adolescent data, suggesting good degree of internal reliability. Category disordering was observed for only one item, namely item 1, "I get tired." At time 1 this was observed for response category 2 "Almost never", and at time 2 this was observed for response category 5 *Always*. However, in both instances the number of responses per category >10. The datasets from the two time points were therefore combined and the analysis re-run. No category disordering was observed for the combined sample. For time 1 the eigenvalues in the first contrast amounted to 1.91. For time 2 this value was 1.73 suggesting no further dimensionality was present in the factor structure. Item fit is shown in Table 1. All items fit fell within the criterion range both at time 1 and time 2 indicating no item misfit. Table 2 shows the results of the DIF analysis. No DIF was displayed by any other items either by gender or by age except for a single item (item 1). Item 1, which displayed a small degree of DIF, was more easily endorsed by younger adolescents (<11 ages) at time 1. Differences in item locations for time 1 and 2 are shown in Table 1. There was minimal change in item locations over time with all differences <0.5 logits. The mean person measure at time 1 was -0.48 (standard deviation (SD) = 1.12) and -0.78 at time 2 (SD = 1.15) indicating a reduction in scores over time. This difference was statistically significant: t (277) = 5.66, p<0.001 (mean difference = -0.30, SD of the difference = 0.87).

Adult sample

Cronbach's alpha was 0.83 for the adult data sample, suggesting good degree of internal reliability. A small degree of category disordering was observed for item 1 between the first (-2.36 logits) and second response categories (-2.38 logits**).** This was not associated with low item category responses (>10). The amount of variance explained by the first contrast was <2.0 suggesting a unidimensional structure. No item misfit was observed for any of the 7 items (Table 3). Although 3 items did demonstrate statistically significant DIF by gender (items 1, 4 and 5) the difference between item locations did not exceed the

<0.5 logits threshold. It may therefore be concluded that no DIF was observed by gender (Table 4). Three items demonstrated DIF by age category, namely items 3, 5 and 7 (Table 4). For instance, item 3 was more easily endorsed by individuals aged 55+ compared to those in the 18 to 34 age group categories. The average differences in person estimates for the 35-54 group and the 55+ age group were small: -0.08 logits (SD 0.16) and -0.09 (SD 0.18), respectively although they were statistically significant ($t(220) = 7.49$, $p<0.001$) and $t(97) = 5.20$, $p<0.001$).

Adolescents and Adults

The variance explained by the Rasch structure amounted to 49.5%, 50.7% and 59.8% for the adolescents (time 1 and 2) and adults, respectively.

DISCUSSION

The aims of this study were to further extend the psychometric assessment of the WAItE in adolescents with obesity and to determine the applicability of the WAItE in an adult population. The results demonstrated that the WAItE has a unidimensional structure (both for adolescents and adults). Item misfit has the potential to distort the measurement properties of an instrument, in other words to negatively impact on the accuracy of the measures or scores produced by respondents. The results showed there was no item misfit observed for either samples and no differential item functioning was present by age or gender for the adolescents. For the adolescent sample stable item locations were observed over time. These assessments of the measurement properties of the WAItE indicate favourable findings in terms of the psychometric evaluation and tests of reliability that have been performed. The tool can be used in the accurate assessment of weight specific QoL with adolescents. Further research assessing other measurement properties such as external validity are underway. We observed that there might be a potential issue with item 1 in terms of category disorder (further research can be undertaken to determine which if any categories need to be collapsed). Some DIF was also observed in the adult sample (3 items), although this appeared to have little or no impact on the person measure estimates.

Existing studies show that instruments can be appropriate for use with a group for which the measure was not directly involved in its development [24]. For example, a recent study by Ratcliffe et al., 2012 found that the CHU9D, a generic instrument originally developed with young people aged 7–11 years, demonstrated properties of reliability and validity when used with was adolescents aged 11–17 years. Given that the content of the WAItE was developed with 11-18 year olds, the feasibility of using the tool with and older age groups was therefore also evaluated. The findings from this study on the performance of the WAItE for adults are promising. In future work it would be beneficial to supplement these findings with qualitative interviews with adults to serve as a further check on the appropriateness of the WAItE content. Future qualitative work would benefit from including adults with obesity engaged in weight management.

Only a minority of the well-known generic QoL instruments for adolescents have employed Rasch analysis in their assessment of measurement properties. The KIDSCREEN52 [25] and Paediatric Quality of Life Inventory (PedsQL) [26] have been subjected to item-response-theory analysis. Rasch analysis is yet to be performed on any of the existing weight-specific tools where the content has been informed by adolescents. The WAItE therefore is the only weight specific measurement of QoL that has been developed with adolescents and whose internal structure has been confirmed by Rasch analysis. Its value will become apparent from use in future assessments of weight management services that engage adolescents with obesity.

In terms of study limitations, for the adolescent participant sample recruitment was limited to one geographical location within the UK. This, together with a lack of information on the socio-economic status of adolescent participants, might have implications on generalisability of findings. Similarly, this sample did not include adolescents with severe obesity who require treatment in a hospital setting. However, the applicability of the WAItE in these adolescents is something that can be tested in future research. A key strength of the study was that all adolescents were engaged with and recruited from

241 community-based weight management services. Potential limitations pertaining to the adult
242 participants include concerns about data quality due to the web-based method of administration of the
243 survey. However, it has been noted that potential problems that might arise from a web-based mode of
244 administration are not unique as they may also arise with self-report pen and paper surveys [24]. Key
245 advantages of a web-based method of survey administration are the ability to recruit from a wide
246 geographical distribution and to set recruitment quotas reflective of background characteristics, for
247 example, recruiting to achieve an even split across gender.

248

Faced with finite and decreasing budgets, decision makers are tasked with ensuring efficiency in the allocation of resources. As it stands, the WAItE can be implemented in assessments of cost-effectiveness of weight management interventions aimed at both adolescents and adults to derive an incremental cost per WAItE score calculation. The WAItE score can be calculated to evaluate whether there is in an improvement or deterioration between the intervention groups being compared. Future research involving a preference valuation study [9] to elicit weight-specific utility values for states described by the WAItE will be needed to facilitate cost-utility analysis of weight management interventions for adolescents and adults.

Overall, given the results from the Rasch analysis, the WAItE showed sufficient psychometric properties to encourage further use in adolescents and adults with obesity.

REFERENCES

1. 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*(4), 282-304.
2. Tsiros, M., Olds, T., Buckley, J., Grimshaw, P., Brennan, L., Walkley, J., et al. (2009). Health-related quality of life in obese children and adolescents. *International Journal of Obesity, 33*(387-400).
3. National-Institute-for-Clinical-Excellence (April 2013). Guide to the Methods of Technology Appraisal.
4. Ravens-Sieberer, U., Redegeld, M., & Bullinger, M. (2001). Quality of life after in-patient rehabilitation in children with obesity. [Research Support, Non-U.S. Gov't]. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity, 25 Suppl 1*, S63-65.
5. Kolotkin, R., Zeller, M., Modi, A., Samsa, G., Quinlan, N., Yanovski, J., et al. (2006). Assessing weight-related quality of life in adolescents. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Validation Studies]. *Obesity, 14*(3), 448-457.
6. Moorehead, M., Ardelt-Gattinger, E., Lechner, H., & Oria, H. (2003). The validation of the Moorehead-Ardelt Quality of Life Questionnaire II. *Obesity Surgery, 13*(5), 684-692.
7. 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*(6), 1171-1177.
8. 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. [Validation Studies]. *Quality of Life Research, 20*(2), 215-224.
9. Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation*: Oxford University Press.
10. Oluboyede, Y., Hulme, C., & Hill, A. (2017). Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. [journal article]. *Quality of Life Research, 26*(8), 2025-2039, doi:10.1007/s11136-017-1561-1.
11. Oluboyede, O. (2013). *Quality of life assessment in adolescent obesity: Development of a new instrument for economic evaluation*. University of Leeds, Leeds.
12. Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes, 1*(1), 27, doi:10.1186/1477-7525-1-27.
13. More-Life. http://www.more-life.co.uk/ Accessed September 2017.
14. Watch-It-Programmes. https://www.leedscommunityhealthcare.nhs.uk/our-services-a-z/watch-it-family-weight-management-service/watch-it-programmes/. Accessed September 2017.
15. Cole, T., Freeman, J., & Preece, M. (1995). Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood, 73*, 25-29.
16. 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*(1), 73-77.
17. Linacre, J. (2014). A user's guide to Winsteps.
18. Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. Reprinted 1980.
19. Masters, G. N. (1982). A rasch model for partial credit scoring. [journal article]. *Psychometrika, 47*(2), 149-174, doi:10.1007/bf02296272.
20. Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*, 33-33, doi:10.1186/1471-2288-8-33.
21. Raîche, G. (2005). Critical Eigenvalue Sizes (Variances) in Standardized Residual Principal Components Analysis. *Rasch Measurement Transactions, 19:1*, p. 1012.
22. Lord, F. (1980). *Applications of item response theory to practical testing problems*: Lawrence Erlbaum Associates.

23. Lai, J.-S., Teresi, J., & Gershon, R. (2005). Procedures for the Analysis of Differential Item Functioning (DIF) for Small Sample Sizes. *Evaluation & the Health Professions, 28*(3), 283-294, doi:10.1177/0163278705278276.
24. Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M. (2012). An assessment of the construct validity of the CHU9D in the Australian adolescent general population. [journal article]. *Quality of Life Research, 21*(4), 717-725, doi:10.1007/s11136-011-9971-y.
25. Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., et al. (2008). The KIDSCREEN-52 Quality of Life Measure for Children and Adolescents: Psychometric Results from a Cross-Cultural Survey in 13 European Countries. *Value in Health, 11*(4), 645-658, doi:http://dx.doi.org/10.1111/j.1524-4733.2007.00291.x.
26. Amin, L., Rosenbaum, P., Barr, R., Sung, L., Klaassen, R. J., Dix, D. B., et al. (2012). Rasch analysis of the PedsQL: an increased understanding of the properties of a rating scale. *Journal of Clinical Epidemiology, 65*(10), 1117-1123, doi:10.1016/j.jclinepi.2012.04.014.

FUNDING

The work presented here was part of a National Institute for Health Research (NIHR) funded fellowship project awarded to the first author (DFR/2009/02/101). This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

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.
Ethical approval was provided by the University of Leeds, School of Medicine Research Ethics Committee for both the adolescents and adult studies (Ref: HSLTLM/11/049).

Informed consent: Both of the weight management services followed their own procedures for obtaining consent. All parents and carers of adolescents provided written or oral consent for adolescents to complete the WAItE. If parents did not object then written or oral assent (under 16 years) /consent (16 years plus) for all participating adolescents was obtained. Anonymised datasets were provided directly from weight management organisations who adhered to strict security protocols.
Adult participants who were recruited from a consumer panel provided consent to the market research company to be approached and complete web surveys.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

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. We would like to acknowledge Aki Tsuchiya (PhD supervisor) for her guidance and support throughout the fellowship project. Finally, we would like to thank all the participants who took part in the research and the parents and staff who supported this research.

TABLES

Table 1. Item fit for adolescent sample - Combined for T1 & T2

Item[a]	Location time 1	IN.MSQ [b]	OUT.MSQ [b]	Location time 2	IN.MSQ	OUT.MSQ	Difference T1-T2
WAItE_1	-1.08	1.06	1.07	-0.85	1.02	1.05	0.23
WAItE_2	0.20	1.06	1.05	0.35	1.03	0.98	0.15
WAItE_3	0.60	1.17	1.19	0.32	0.99	1.01	-0.28
WAItE_4	0.04	1.18	1.13	-0.05	1.20	1.25	-0.09
WAItE_5	-0.13	0.87	0.84	-0.01	0.92	1.05	0.12
WAItE_6	-0.11	0.71	0.67	-0.09	0.73	0.71	0.02
WAItE_7	0.47	0.98	0.87	0.35	1.07	1.09	-0.12

[a]WAItE_1 = Tired; WAItE_2 = Walking; WAItE_3 = Sports; WAItE_4 = Concentrate; WAItE_5 = Embarrassed; WAItE_6 = Unhappy; WAItE_7 = Treated differently
[b] Misfit Indices: IN.MSQ = Infit Mean Square; OUT.MSQ = Outfit Mean Square

Table 2. Differential Item Functioning (DIF) adolescents

Gender (t1)											
Item	**Item location (Girls)**	**SE[a]**	**Item location (Boys)**	**SE**	**CONTRAST[a]**	**Joint SE**	**t[a]**	**df[a]**	**p[a]**	**MH X2[a]**	**p (MH) [a]**
WAItE_T1_1	-1.12	0.12	-1.03	0.13	-0.09	0.18	-0.49	257	0.62	0.34	0.56
WAItE_T1_2	0.17	0.09	0.24	0.11	-0.07	0.14	-0.52	256	0.61	0.44	0.51
WAItE_T1_3	0.54	0.09	0.68	0.11	-0.14	0.14	-1.00	256	0.32	0.81	0.37
WAItE_T1_4	0.19	0.09	-0.14	0.10	0.32	0.13	2.43	260	0.02	6.54	0.01
WAItE_T1_5	-0.19	0.08	-0.04	0.09	-0.14	0.12	-1.17	256	0.24	1.81	0.18
WAItE_T1_6	-0.11	0.08	-0.13	0.09	0.02	0.13	0.16	258	0.87	0.28	0.60
WAItE_T1_7	0.47	0.09	0.44	0.10	0.02	0.13	0.19	255	0.85	0.03	0.87
Age group[b] (t1)											
Item	**Item location (11-14)**	**SE**	**Item location (15-18)**	**SE**	**CONTRAST**	**Joint SE**	**t**	**df**	**p**	**MH X2**	**p (MH)**
WAItE_T1_1	-0.95	0.10	-1.48	0.18	0.53	0.21	2.59	105	0.01	9.69	0.00
WAItE_T1_2	0.18	0.08	0.28	0.15	-0.10	0.17	-0.63	102	0.53	0.05	0.83
WAItE_T1_3	0.63	0.08	0.50	0.14	0.13	0.16	0.78	104	0.44	0.66	0.42
WAItE_T1_4	0.04	0.08	0.00	0.14	0.04	0.16	0.29	103	0.77	0.02	0.90
WAItE_T1_5	-0.17	0.07	0.01	0.13	-0.18	-0.14	1.23	103	0.22	1.74	0.19
WAItE_T1_6	-0.11	0.07	-0.09	0.13	-0.02	0.15	-0.14	103	0.89	0.42	0.52
WAItE_T1_7	0.43	0.07	0.57	0.13	-0.13	0.15	-0.87	103	0.39	1.70	0.19
Gender (t2)											
Item	**Item location (Girls)**	**SE**	**Item location (Boys)**	**SE**	**CONTRAST**	**Joint SE**	**t**	**df**	**p**	**MH X2**	**p (MH)**
WAItE_T2_1	-0.94	0.11	-0.75	0.13	-0.19	0.17	-1.11	258	0.27	0.46	0.50
WAItE_T2_2	0.39	0.10	0.30	0.11	0.09	0.15	0.59	261	0.56	0.14	0.71
WAItE_T2_3	0.32	0.10	0.32	0.11	0.00	0.15	0.00	260	1.00	0.00	0.98
WAItE_T2_4	0.10	0.09	-0.24	0.10	0.33	0.14	2.41	262	0.02	4.46	0.03
WAItE_T2_5	-0.19	0.08	0.24	0.10	-0.43	0.13	-3.26	249	0.0013	11.90	0.00
WAItE_T2_6	-0.12	0.09	-0.07	0.10	-0.05	0.13	-0.36	258	0.72	0.08	0.77

WAItE_T2_7	0.46	0.10	0.21	0.10	0.26	0.14	1.84	264	0.07	2.86	0.09
Age group (t2)											
Item	**Item location (11-14)**	**SE**	**Item location (15-18)**	**SE**	**CONTRAST**	**Joint SE**	**t**	**df**	**p**	**MH X2**	**p (MH)**
WAItE_T2_1	-0.77	0.10	-1.14	0.17	0.37	0.20	1.88	103	0.06	3.87	0.05
WAItE_T2_2	0.31	0.09	0.46	0.16	-0.14	0.18	-0.79	101	0.43	0.41	0.52
WAItE_T2_3	0.29	0.09	0.39	0.15	-0.10	0.18	-0.55	102	0.58	0.29	0.59
WAItE_T2_4	-0.01	0.08	-0.18	0.14	0.17	0.16	1.05	104	0.30	0.50	0.48
WAItE_T2_5	-0.04	0.07	0.09	0.13	-0.13	0.15	-0.87	102	0.39	0.55	0.46
WAItE_T2_6	-0.13	0.08	0.03	0.14	-0.16	-0.16	1.03	102	0.30	1.69	0.19
WAItE_T2_7	0.38	0.08	0.27	0.14	0.11	0.16	0.67	107	0.50	0.36	0.55

[a] S.E = Standard Error; CONTRAST = difference in logits between the two measures; t = Wald t-statistic; df = Degrees of freedom; p = p-value; MH X2 = Mantel-Haenszel Chi-squared; p (MH) = Mantel-Haenszel p-value

[b]1) =11-14 year olds; 2) = 15-18 year olds

Table 3. Item fit for adult sample

Item	Item location	IN.MSQ	OUT.MSQ
WAItE_1	-1.78	1.13	1.15
WAItE_2	0.44	0.85	0.84
WAItE_3	-0.85	1.38	1.55
WAItE_4	0.20	1.13	1.10
WAItE_5	0.44	0.96	0.93
WAItE_6	0.43	0.64	0.63
WAItE_7	1.12	0.88	0.86

Table 4. DIF Adults

Items	Item location (Males)	SE	Item location (Females)	SE	CONTRAST	SE	t	df	p	MH X2	p (MH)
WAItE_1	-1.55	0.11	-2.00	0.10	0.44	0.15	3.00	481	0.003	10.50	0.00
WAItE_2	0.44	0.09	0.44	0.08	0.00	0.12	0.00	481	1.000	0.12	0.73
WAItE_3	-0.88	0.07	-0.82	0.07	-0.05	0.10	-0.52	482	0.603	0.07	0.79
WAItE_4	0.00	0.09	0.38	0.09	-0.38	-0.12	3.05	482	0.002	7.52	0.01
WAItE_5	0.64	0.09	0.28	0.08	0.36	0.11	3.14	475	0.002	11.45	0.00
WAItE_6	0.37	0.08	0.49	0.08	-0.11	-0.11	1.01	482	0.313	3.14	0.08
WAItE_7	1.04	0.09	1.19	0.09	-0.15	-0.13	1.20	481	0.230	2.35	0.13

CLASS	Group	Item difficulties 1	SE	CLASS	Item difficulties 2	SE 1	CONTRAST	SE 2	t	df	p	X2 (MH)	p
WAITE_1	18-34	-1.81	0.12	35-54	-1.73	0.11	-0.08	0.17	-0.46	366	0.644	0.31	0.58
WAITE_1	18-34	-1.81	0.12	55+	-1.84	0.16	0.03	0.21	0.15	200	0.881	0.12	0.73
WAITE_2	18-34	0.78	0.10	35-54	0.44	0.09	0.34	0.14	2.47	357	0.014	5.38	0.02
WAITE_2	18-34	0.78	0.10	55+	-0.15	0.13	0.93	0.16	5.66	215	0.000	27.37	0.00
WAITE_3	*18-34*	*-0.38*	*0.09*	*35-54*	*-0.95*	*0.08*	*0.57*	*0.12*	*4.84*	*361*	*0.000*	*19.53*	*0.00*
WAITE_3	*18-34*	*-0.38*	*0.09*	*55+*	*-1.44*	*0.12*	*1.06*	*0.15*	*7.17*	*199*	*0.000*	*40.96*	*0.00*
WAITE_4	18-34	-0.23	0.10	35-54	0.28	0.09	-0.51	-0.14	3.67	367	0.000	14.45	0.00
WAITE_4	*18-34*	*-0.23*	*0.10*	*55+*	*0.82*	*0.15*	*-1.05*	*-0.18*	*5.87*	*190*	*0.000*	*29.36*	*0.00*
WAITE_5	18-34	0.17	0.09	35-54	0.44	0.09	-0.27	-0.12	2.16	370	0.031	5.15	0.02
WAITE_5	*18-34*	*0.17*	*0.09*	*55+*	*0.98*	*0.14*	*-0.81*	*-0.17*	*4.8*	*177*	*0.000*	*18.50*	*0.00*
WAITE_6	18-34	0.45	0.10	35-54	0.43	0.08	0.02	0.13	0.17	363	0.869	0.34	0.56
WAITE_6	18-34	0.45	0.10	55+	0.43	0.13	0.02	0.16	0.13	199	0.895	0.07	0.80
WAITE_7	*18-34*	*0.88*	*0.10*	*35-54*	*1.12*	*0.09*	*-0.24*	*-0.14*	*1.73*	*371*	*0.084*	*4.37*	*0.04*
WAITE_7	*18-34*	*0.88*	*0.10*	*55+*	*1.58*	*0.16*	*-0.71*	*-0.19*	*3.77*	*175*	*0.000*	*12.05*	*0.00*