Institute of Education

A measurement evaluation of a six item measure of quality of life (CASP6) across different modes of data collection in the 1958 National Child Development Survey (NCDS) Age 55 years.

By Wiggins, R.D., Brown, M. and Ploubidis, G.B

Centre for Longitudinal Studies
Working paper 2017/2

CENTRE FOR LONGITUDINAL STUDIES

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Contact the author
Dick Wiggins
Professor in Quantitative Social Science
Department of Social Science
UCL Institute of Education
d.wiggins@ucl.ac.uk

This working paper was first published in June 2017
by the Centre for Longitudinal Studies,
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institution of Education (IOE). It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ioe.ac.uk.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: clsfeedback@ioe.ac.uk

Contents

Acknowledgements

This work was funded as part of Economic and Social Research Council support for the Centre for Longitudinal Studies and NCDS. Many thanks to the 1958 NCDS cohort members whose continued participation in the study is greatly appreciated.

Abstract

The purpose of this paper is to evaluate the performance of a reduced six-item self-report (CASP6) of a broader 12-item version of a quality of life measure (CASP-12v2). The focus of our analytical assessment was to examine the impact of the mode of data collection on the measurement properties of CASP6 in the context of an evaluation of the 'sequential mixed-mode' design adopted for the National Child Development Study Age 55 Survey, where cohort members were first invited to complete the survey online, and then by telephone if they had not completed the online survey after 5 weeks. A general - specific measurement model including two first order factors to capture both the unidimensionality of the scale and a specific method factor to identify negatively worded items across modes revealed a good fit (CFI=0.993, TLI=0.984 & RMSEA=0.068, N= 9023). Similar assessments for either online or telephone alone and mixed mode with a telephone option revealed confirmatory results for the use of CASP6 as a standalone measure of quality of life.

Keywords

Quality of life, scale evaluation, item response theory, MPlus.

Section 1: Introduction

This paper provides an evaluation of changes in the length of well-established versions of a self-reported measure of quality of life consisting of 19 or 12 items (CASP19 and CASPv1 & v2, see Hyde et al, 2015 and www.casp19.com) to a shorter six-tem version. This evaluation is conducted in the context of the National Child Development Study (NCDS) when cohort members were aged 55 years. During that sweep NCDS adopted a 'sequential mixed mode' strategy for data collection. This presented a unique opportunity to also evaluate the use of the shorter version of the study under two different modes of data collection namely, online and telephone. Typically, administrations of CASP are conducted in the context of a face to face interview or as a stand-alone self-report via a paper self-completion questionnaire (Hyde et al., 2015)

During the previous NCDS Age 50 survey cohort members (CMs) were asked to complete 14 of the original 19-item so as to provide users with the option of using either the 12-tem version of CASP (CASP-12 v1) based on a preliminary evaluation of the scale reported in 2003 by Hyde et al and used extensively in the Study of Health and Retirement in Europe (Bŏrsch-Supan et al., 2005) or the subsequently preferred 12-tem version (CASPv2) reported by Wiggins et al., 2008. Summative measures of quality of life based on either version actually hold up as useful indicators of quality of life. These questions were included in a paper self-completion questionnaire which was posted to CMs to complete in advance of a face-to-face interview.

The paper continues with a brief history of the development of CASP and the rationale for reducing the scale to six items [section 2] followed by a description of the sample sources upon which our evaluation was based taking into account the mode of data collection or otherwise [section3]. Section 4 outlines the underlying theoretical context for our evaluation (item response theory) followed by our assessment strategy [section 5], our findings are presented in section 6 in terms of the measurement properties of CASP6, the impact of mode of data collection upon these properties and the performance of the measure in terms of the relationship each item has with an underlying latent trait for quality of life (factor loadings), the amount of 'information' and effective measurement range conveyed by CASP6 across sample source and its comparison with CASP12v2 in a simple probit regression model. Finally in our concluding section 7, we examine the implications for users, who are interested in using the short version of CASP in a survey involving either online or telephone based data collection – whether in a single or mixed mode strategy.

Section 2: A brief history of the development of CASP and our rationale for introducing a six item version

The original version of CASP contained nineteen items spanning four life domains (Hyde et al., 2003) and was developed in the context of a being a self-report typically administered in the context of a face-to-face interview. The rationale for the original item selection was driven by theoretical concerns about the meaning of life in early old age for a group of individuals born during or around post-world war II had reached a time of life when they were relatively free of family responsibilities and had time and resources to experience and enjoy new opportunities. In measurement terms CASP-19 was articulated as a second-order factor analytic model where four life domains, control, autonomy, self-realisation and pleasure had an underlying association with an intrinsic measure of quality of life (QoL). Subsequent measurement evaluations (e.g., Wiggins et al., 2008) gave support to the use of domain specific scores and/or a single measure of QoL (see Hyde et al., 2015 and www.casp19.com for a fuller account of the evolution and use of various versions of CASP). The practical demands of administering large scale multi-purpose studies for longitudinal research gave rise to requests or attempts to reduce the length of the original item set in order to allow survey managers to either reduce respondent burden and/or allow more 'schedule' space to include other questions in their final schedules for a given budget. As a result two shorter 12-item versions of CASP (CASP12v1 and CASP12v2) were developed which broadly followed traditions of scale reduction as exemplified in the development of the twelve item versions of the General Health Questionnaire (GHQ, Goldberg and Williams, 1988) and the SF-36 (Ware et al., 2002) along with the six item Depression-Happiness scale (Joseph et al., 2004) and given wide account in DeVellis (2012). The twelve item versions of CASP retained some of the elegance of the original scale as well as providing a robust single summative or latent score for a person's expressed QoL. However, the strict separation of control and autonomy did not really hold firm. For the age 50 years sweep of NCDS fourteen CASP items were administered via a paper self-completion questionnaire which was posted to CMs in advance of a face-to-face interview, to allow users to adopt comparisons with the twelve item versions of the scale. In particular, the survey of health and retirement in Europe (SHARE), (Börsch-Supan et al., 2005) had adopted CASP12v1 based on an earlier evaluation of the scale (Hyde et al., 2003). The appendix lists all of the CASP items in a way which allows the reader to identify the various versions of CASP that have been administered since 1999 including CASP6.

When developing the NCDS Age 55 Survey, budget constraints meant that the questionnaire had to be limited to no more than 25 minutes duration. This brought additional pressure to reduce the number of CASP items to be administered. A decision was taken to retain some of the CASP items but at the same time recognise that a smaller scale would rule out the use of specific life domain scores. The six item version contains a balanced set of both negatively and positively worded statements from the original scale. In particular, CASP-6 contains three items which were originally employed to operationalise 'control' and three items which were used to operationalise 'self-realisation'. All of the items correlated well with an overall index for QoL based on the age 50 year responses especially, the statement 'I feel that the

future looks good for me'. Although it would be fair to add that all of the original items from the life domains 'pleasure' and 'self-realisation' are strongly associated with the total score. The three negatively worded items taken from the original 'control' life domain are less well correlated with the total score but do provide a balance of negative and positive aspects of ageing. Table 1 below provides the correlations for each CASP12v2 item as administered in the age 50 years sweep with the summative score for QoL based and in the adjacent column their corresponding correlations for each of the CASP6 items and their respective total QoL score for CMs aged 55 years. The relationship each individual item has with its relevant total QoL score is remarkably consistent over time and the correlations for each of the six items selected for CASP6 with their corresponding total are very similar. In order to trace the history of the versions of CASP see Appendix table A.1 for item content for CASP12 versions 1 & 2 and CASP6.

Table 1: Correlation of each individual CASP item with total scores for quality of life based on CASP12v2 at age 50 years and CASP6 at age 55 years

	Age 50 years Overall CASP-12v2 score	**Age 55 years Overall CASP6 score**
My age prevents me from doing the things I would like to	**.522****	**.605****
I feel what happens to me is out of my control	**.650****	**.710****
I feel left out of things	**.644****	**.701****
I can do the things I want to do	.567**	NA
I feel that I can please myself what I do	.484**	NA
Shortage of money stops CM doing things wants to do	.510**	NA
I look forward to each day	.722**	NA
I feel that my life has meaning	.668**	NA
I enjoy the things that I do	.695**	NA
I feel full of energy these days	**.713****	**.695****
I feel that life is full of opportunities	**.719****	**.724****
I feel that the future looks good for me	**.782****	**.745****

Section 3: Sample sources for the examination of the measurement properties of CASP6

The fieldwork strategy for the NCDS Age 55 survey was that CMs were first invited to participate online. Non-responders after 5 weeks were contacted by telephone and asked to take part via telephone instead. Disentangling mode effects from selection effects can be difficult, so in order to robustly assess the impact of mode on measurement and data quality a random sub-sample of CMs (for whom a telephone contact number was held) were randomly assigned to a 'telephone only' protocol in order to establish a mode comparison group. A total of 10,586 non-emigrant cases for whom a telephone number was held at the point the sample was issued were included in the mode experiment. 9110 CMs were allocated to the mixed mode protocol where they were first invited to participate online with telephone follow-up if they had not responded by web after 5 weeks. The remaining 1476 CMs were allocated to a telephone only protocol and were not given the opportunity to participate online. A further 367 emigrants, 572 cases with no telephone number and 268 others were not included in this experiment. The 'others' included 240 CMs who has been part of a mixed mode pilot study, six months prior to the main sweep of data collection. With the exception of the 'emigrant' group all of these additional and pilot study cases were all approached using the mixed mode protocol. Emigrants were invited to participate online but were not followed up by telephone. Altogether the total number of interviews achieved online was 5995 and those obtained by telephone numbered 3142.

The number of cases issued and the number of cases achieved broken down by mode of data collection are given in Table 2 below:

Table 2: A summary of the number of available NCDS cases issued and achieved by mode of data collection during the age 55 years sweep.

	Experiment		Excluded from experiment			Total
	Mixed Mode (Web>Tel)	Telephone only	Emigrants	Cases with no telephone number	Other	
Issued	9110	1476	367	572	268	11,793
Achieved online	5612	Nil (by design)	194	28	161	5995
Achieved by telephone	1935	1149	Nil (field decision)	16	42	3142
Total achieved interviews	7547	1149	194	44	203	9137

Our results for the performance of CASP-6 are based on *six* sample sources: 'W&T/ALL' ignores the influence of mode on item responses and combines data from both experimental and non-experimental sources, 'W/ALL' focuses on the item responses for all those who responded online, again ignoring the distinction between experiment and non-experiment, 'T/ALL' focuses on all those responding by telephone either whether as part of the mixed mode or telephone only protocol, 'W&T/E' refers to the item responses obtained from the experimental group alone irrespective of mode and 'T/E' focuses upon the item response from those CMs in the experiment who were only offered the telephone as a mode of response. The subsample 'T/OW' refers to those CMs who were originally allocated to the web mode under the experimental conditions but subsequently completed the survey by telephone follow-up (essentially, 'non-compliers'). The term 'ALL' is used interchangeably with 'pooled'. Table 3 distinguishes the sample sizes for each analysis that follows. Any discrepancies in the marginal totals between those in table 3 and the previous table (2) are due to the fact that some CMs did not answer any CASP items and were dropped from the analysis. Thus only CMs with 1 or more CASP item responses were included in the analysis.

Table 3: ‘labelled ‘samples used for the measurement evaluation of CASP6.

Sample description & label	**Sample size**
Web & Tel.: combined experimental and non-experimental data (‘Pooled’ or W&T/ALL)	**9023**
Web alone : combined experimental and non-experimental data (W/ALL)	**5932**
Tel Only: combined experimental and non-experimental data (T/ALL)	**3091**
Web & Tel: experimental data alone (W&T/E)	**7455**
Tel. Only : experimental data (TO/E)	**1142**
Tel. Only: originally allocated to web mode (TO (MM)/E)	**1892**

Key: W=Web, T= Telephone, O=Only, ALL =all available cases irrespective of mode or experimental origins (in other words ‘Pooled’), E = Experimental cases, MM = mixed mode, initially offered opportunity to respond on-line where no response after 5 weeks offered telephone mode

In addition to comparing the measurement properties of these six samples defined in terms of combining modes or isolating mode across experimental and non-experimental origins we invoked specific tests to assess the assumption of measurement invariance or equivalence which is implicit in any group comparison. Formally, measurement equivalence exists ‘when the relations between the observed test scores (indicators) and the latent attribute measured by a test are identical across subpopulations’ (Meredith, 1993; Ploubidis & Grundy, 2009). Specifically, group comparisons were carried out for online (Web) versus Telephone based on all cases participating via these modes irrespective of experimental status and Web and Telephone versus Telephone for the experimental data alone. These comparisons are respectively labelled in subsequent tables of results as ‘Web vs. Tel. Invariance’ and ‘Mixed Mode Invariance’. We will now turn to a brief account of the underlying statistical approach to our evaluation.

Section 4: An outline of the underlying latent variable model within an item response theory (IRT) framework

Formally, CASP-6 contains six ordinal item responses and our measurement objective is to ascertain whether or not a single summated index or latent score based on these item responses is robust across different modes of data collection. To achieve this assessment we draw upon both item response theory (IRT) and confirmatory factor analysis (CFA). We modelled the probability of response to the binary Malaise inventory items with a 2 parameter probit unidimensional latent variable measurement model (Muthen, 1984; Rabe-Hesketh & Skrondal, 2008) with an “underlying variable” probit specification which has been shown to be equivalent

to the item response function specification which is typically invoked in Item Response Theory applications (Bartholomew, Knott, & Moustaki, 2011).

For simplicity we will present the binary case where we are simply modelling the adoption of response category 1 over 0. This approach can be extended, of course, for a full latent response formulation for ordinal response categories for a four point scale where we are modelling the probability of adopting response category 1 over 0, 2 over 1 and 3 over 2. These probabilities relate to underlying thresholds or positions on an underlying latent continuum for each particular item.

To continue with the binary case our categorical or binary observed indicators (y_{ij}) are related to continuous latent variables (η_j) via a normal ogive response model, such that:

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > \tau_i \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

where $y_{ij}^* = \beta_i + \lambda_i \eta_j + \varepsilon_{ij}$

for i=1,…, I_j (I_j being the number of observed indicators for latent variable j) and j=1,.., J (J being the number of latent variables,; here J=6) where the value of y^*_{ij} exceeds a particular threshold value τ_i for the item response which is directly related to a position η_j on an underlying latent scale for any specific item i. The relationship between y^*_{ij} and η_j is driven by two item specific parameters β_i and λ_i broadly equivalent to the item's level of 'difficulty' and 'discrimination' in an IRT context. Perhaps better described in the context of measuring 'quality of life' (an unobserved trait) as 'how much quality of life' an individual holds if s/he endorses a particular item and the extent to which the item endorsement distinguishes between levels of an individual's quality of life. The expression also contains a unique residual term, ε_{ij} .

We assume that

$\eta_j \sim N(0, \Psi)$, $\varepsilon_{ij} \sim N(0,1)$, $COV\left(\eta_j, \varepsilon_{ij}\right) = 0$

where Ψ is a diagonal matrix and COV stands for covariance.

Model (1) can be equivalently expressed as:

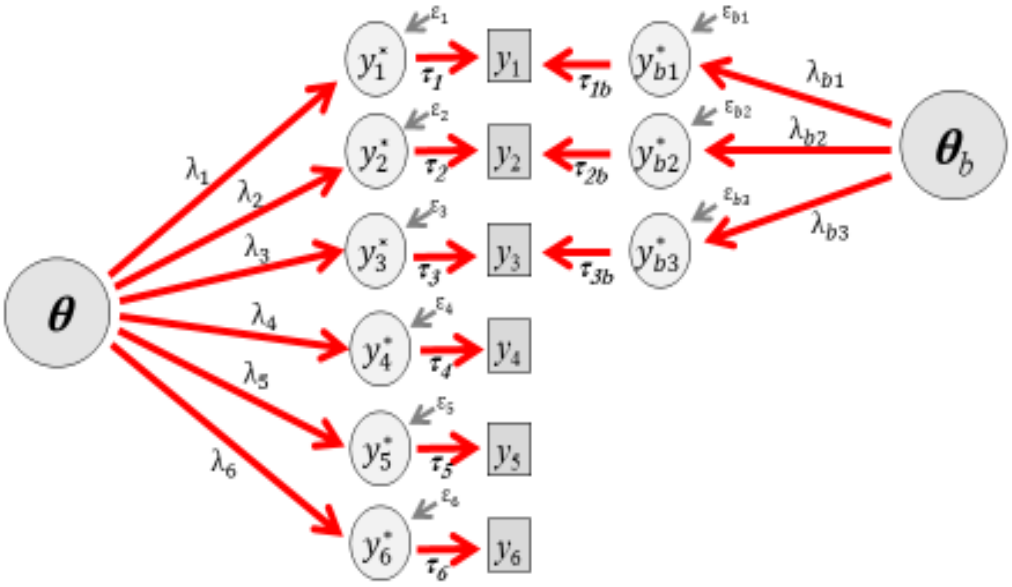
$$\Pr\left(y_{ij} = 1 \mid \eta_j\right) = \Pr\left(y_{ij}^* > \tau_i \mid \eta_j\right) = \Phi\left(\beta_i + \lambda_i \eta_j\right)$$

$$\Phi^{-1}\Pr\left(y_{ij} = 1 \mid \eta_j\right) = \beta_i + \lambda_i \eta_j$$

Where $\Phi(\cdot)$ is the cumulative standard normal distribution and Φ^{-1} is the probit link.

Our model formulation can be presented diagrammatically as a measurement model with a single underlying factor. Figure 1 expresses the model by adopting familiar conventions for CFA where rectangles represent indicators (items) and circles represent latent variables (factors) and arrowheads indicate item or indicator dependencies. The diagram presents a generic form of the model described using algebra above including Θ to label our substantive factor (here quality of life) and Θ_b to denote our 'method' or 'wording' factor.

Figure 1. CASP6 Age 55 years: preferred measurement model for 2-factors expressed in general algebraic format



In order to test the assumption of measurement invariance referred to in an earlier section we 'fix' (i.e. make equivalent) the value of the parameters β and λ in our item response model and examine the estimated thresholds (τ) for each item category across our group comparisons. Next we will describe our assessment strategy.

Section 5: Assessment Strategy

The first step in our analysis was to assess the goodness of fit of the single factor model for polytomous (ordinal) indicators using the full sample across all modes of data collection. All analyses included only those CMs with 2 or more CASP items under a full categorical (ordinal) specification. Any item missingness was handled using the 'Weighted Least Squares- Mean and Variance corrected' (WLSMV) estimator using Mplus version 7.2 (Muthén and Muthén, 2004) for items which were 'pairwise present'[1]. We used three goodness of fit criteria; the comparative fit index (CFI) where CFI >= 0.95 to denote 'good fit', the Tucker-Lewis Index (TLI) where TLI>= 0.95 to denote 'good fit' and the root mean square error (RMSEA) where RMSEA <= 0.05 to denote 'good fit' (Bollen,1993). Next we examined the performance of the simple model across all sample groups isolating mode wherever appropriate (described above). Strategically at this stage of our evaluation if the simple measurement model were judged to be inappropriate we would endeavour to amend the model formulation. Once a satisfactory measurement model was obtained for each sample source we proceeded to explore how i) the

[1] Analyses were also repeated for samples with complete cases and for 1 or more items missing using FIML (Enders, 2010).

thresholds/"difficulty" parameters for each category shift (0-1, 1-2 and 2-3) varied by sample; (ii) how the factor loadings vary by sample and iii) how the information function for CASP6 behaved for different samples. An information function is constructed across the range of the latent trait for measuring quality of life where the value of the function at any point on the latent scale corresponds to the reciprocal of the standard error of the trait at any particular point along its range. The three positive items 'I feel full of energy these days', 'I feel that life is full of opportunities' and 'I feel that the future looks good for me' were reverse coded to ensure that the more frequent an endorsement received a high rating and any resulting summated score would indicate that a high score matched a 'good quality of life'. Finally, we show how latent scores based on CASP6 and CASP12v2 compare when we predict current employment status in a simple probit model. Additionally, table A.2 in the appendix contains CASP6 inter-item correlations and correlations for key social indicators (health, living alone, work status, social class and tenure) with the total CASP6 score. The sample selections for model evaluation were described above in table 3 and their corresponding labels are used in the results section which follows below.

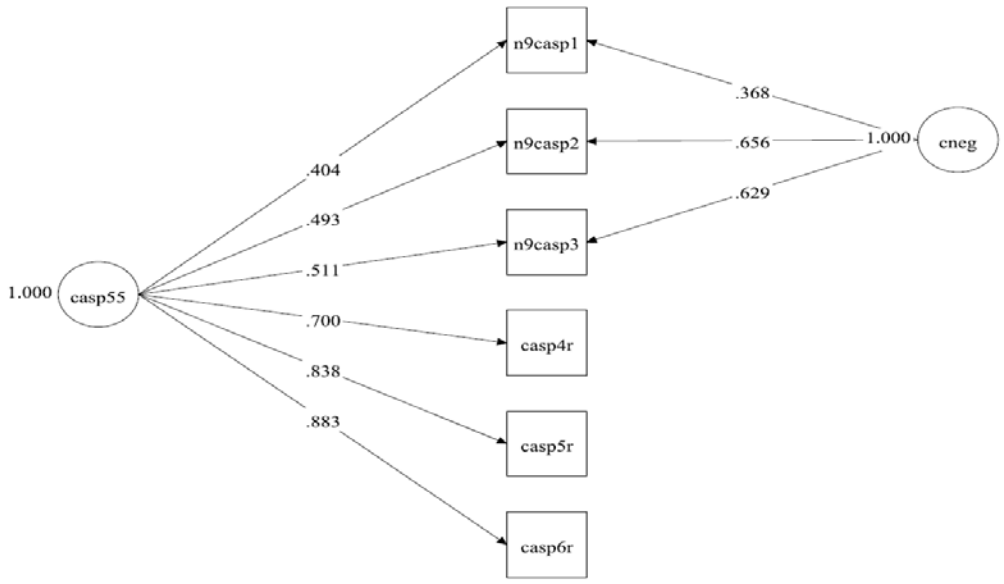
Section 6: Results

6.1 Evaluating our measurement model

A single factor analytic model for the six categorical CASP items was first tested using W&T/ALL cases and found to be unsatisfactory (CFI=0.911, TLI=0.852, RMSEA= 0.204). Upon reflection our next approach was to include a second first order 'Wording' factor in addition to our underlying latent variable for Quality of Life (QoL) to distinguish between the negative and positively worded statements following Hazlett-Stevens, Ullman & Craske (2004), Jo (2000) and Abbott et al., (2006). This adjusted measurement model when applied to the W&T/ALL sample provided a much improved fit (CFI=0.993, TLI=0.984, RMSEA= 0.068)[2]. The item loadings on 'QoL' and 'Wording' are relatively strong and positive. A simplified version of this model is provided in figure 2 below:

[2] The value of RMSEA is much improved over the single factor model and lies below 0.8 which is regarded as acceptable see table 2, Wiggins et al., (2008).

Figure 2 Measurement Model for CASP6 consisting of two first order factors, 'quality of life and 'wording' for NCDS members aged 55 years (visualization taken directly from MPlus graphics).



Notes

casp55 – latent variable (or factor) for quality of life

cneg – latent variable for 'wording' or 'method'

n9casp1, n9casp2, n9casp3 are equivalent to casp1,casp2 and casp4 in table A.1 (all negatively worded hence leading 'n' in label).

Factor loadings are estimated for each path.

Having identified a suitable measurement model based upon the W&T/ALL sample we examined its properties across four alternative sources and included specific comparisons for measurement invariance as described above. For comparative ease we have replicated the goodness of fit criteria for the single factor model and the two first order factor model in the first two lines of table 4 below. The results for the latter model have been emboldened for reference. The measurement model holds up for each separate mode W/ALL and T/ALL as well as for the combined 'W&T/E' and the telephone mode alone 'T/E' based on the sequential mixed mode experimental data. Measurement invariance tests for cross-mode (W vs T) are also satisfied either when the comparison is constructed for the mode comparison irrespective of experimental status or not (CFI=0.979, TLI=0.980, RMSEA=0.077). We are now in a position to examine the performance of our measurement model in three important ways: the pattern of six CASP item loadings across sample sources on QoL, the thresholds for movement between the item response categories (0-1, 1-2 and 2-3) and the 'information content' of our latent variable (QoL) across its numerical range.

Table 4. Goodness of Fit criteria for competing measurement models tested across combined and single modes of data collection.

	CFI	*TLI*	*RMSEA*
Single factor for QoL	0.911	0.852	0.204
Two factor model based on 'pooled sample' (W&T/ALL)	**0.993**	**0.984**	**0.068**
Web for 'pooled sample' (W/ALL)	0.991	0.977	0.087
Tel for 'pooled sample' (T/ALL)	0.994	0.986	0.053
Tel within original web allocation (T/OW)	0.994	0.984	0.057
Web Vs Tel Invariance	0.979	0.98	0.077
Mixed mode experiment (W&T/E)	0.993	0.982	0.074
Mixed mode experiment Tel. Alone (T/E)	0.996	0.989	0.044
Mixed Mode Invariance	0.995	0.995	0.038

Web = online mode, Tel. = Telephone mode, OW = originally allocated to Web

6.2 Item loadings for CASP6 across sample sources

Table 5 provides the factor loadings for each of the six CASP items under our selected measurement model across the complete range of sample sources. The three positive statements 'I feel full of energy these days', 'I feel that life is full of opportunities' and 'I feel that the future looks good for me' have the strongest item specific loadings (all above 0.6) and hold strongly for the online mode. The other statements which capture a negative or less assertive outlook on life, 'My age prevents me from doing the things I would like to do', 'I feel what happens to me is out of my control', and 'I feel left out of things' hold consistently across sample sources but are less strong in general than the positive attributes and similarly less well defined for the telephone mode than the online mode.

Table 5. Factor loadings for the selected measurement model

	Pooled	*Web*	*Tel all*	*Tel (OW)*	*W Vs T Inv*	*Mixed WT*	*Mixed Tel*	*Mixed Inv*
My age prevents me from doing the things I would like to do	0.404	0.440	0.351	0.338	0.351	0.409	0.394	0.413
I feel what happens to me is out of my control	0.493	0.502	0.484	0.473	0.454	0.493	0.487	0.495
I feel left out of things	0.511	0.516	0.508	0.478	0.492	0.502	0.542	0.508
I feel full if energy these days	0.700	0.745	0.631	0.657	0.618	0.714	0.563	0.711
I feel that life is full of opportunities	0.838	0.859	0.812	0.815	0.784	0.841	0.808	0.841
I feel that the future looks good for me	0.883	0.905	0.853	0.849	0.855	0.885	0.858	0.885

6.3 An examination of threshold movement for item response categories across each sample source

Next turning to the estimated threshold values in Table 6. The table is constructed so that there are three numerical values (thresholds) for each item by sample source tested under our selected measurement model. The first row by each statement indicates the position a hypothetical person holds on the underlying latent trait to indicate the point at which s/he moves from 0 to 1 on the observed item rating as recorded, equally the second row indicates the move from 1 to 2 and the third row indicates the move from 2 to 3 on the observed item rating. The values of the thresholds on the latent scale are arbitrary in the sense that their usefulness provides a means of interpreting the impact of mode (combined or otherwise) on the item response. Take for example the statement 'I feel full of energy these days' and compare the threshold values for the comparison between the web alone (W/ALL) and the telephone alone (T/ALL). Firstly, the range for the telephone mode is slightly narrower than the corresponding range for the online mode (-1.574 to 0.654 cp. -1.201 to 0.505) suggesting that the telephone mode is slightly less sensitive to high and low values of QoL. A score of -1.201 for the telephone triggers a move from 0 to 1 (never to rarely) compared to a relatively lower score of -1.574 for the web. Similarly the threshold values for moves from 1 to 2 (rarely to sometimes) and 2 to 3 (sometimes to often) have to be relatively higher for the online mode compared to the telephone mode. A similar interpretation can be applied to all six items under this comparison. Put another way, the same item when responded to via the web is relatively more discriminating along our underlying scale of QoL than when it is administered by means of a telephone interview.

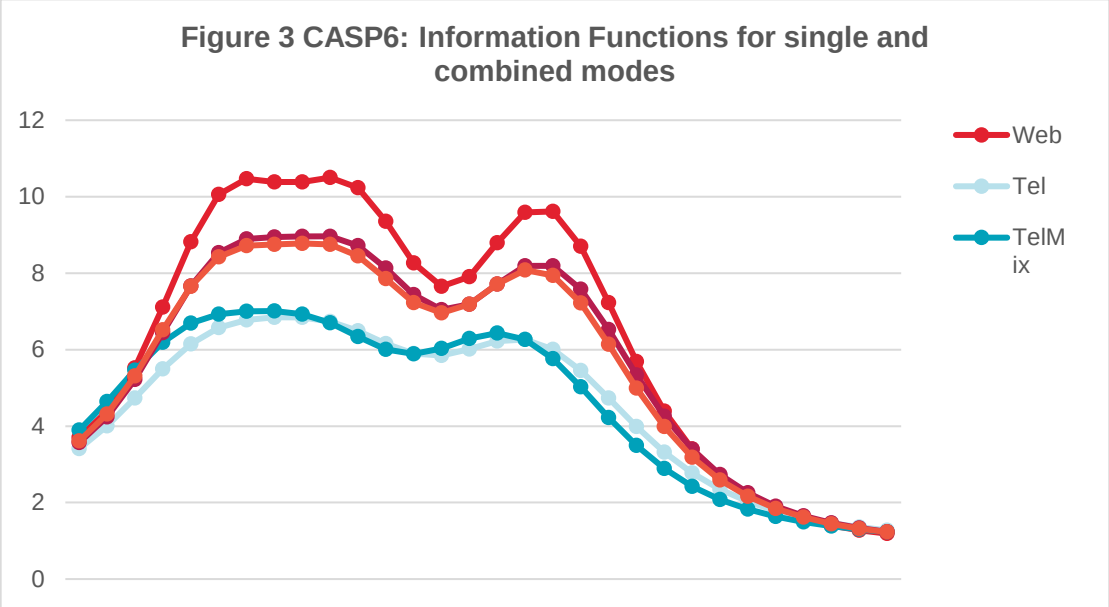
Table 6. Thresholds of the selected measurement model

	Pooled	*Web*	*Tel all*	*Tel (OW)*	*W Vs T Inv*	*Mixed WT*	*Mixed Tel*	*Mixed Inv*
My age prevents me from doing the things I would like to do	-1.794	-2.049	-1.494	-1.425	-1.595	-1.813	-1.628	-1.805
	-0.637	-0.682	-0.553	-0.537	-0.573	-0.642	-0.589	-0.632
	0.148	0.286	-0.114	-0.111	0.114	0.186	-0.13	0.165
I feel what happens to me is out of my control	-1.593	-1.775	-1.342	-1.241	-1.512	-1.589	-1.556	-1.582
	-0.562	-0.646	-0.410	-0.339	-0.561	-0.559	-0.533	-0.542
	0.192	0.261	0.062	0.084	0.143	0.219	0.023	0.216
I feel left out of things	-1.787	-1.94	-1.572	-1.498	-1.75	-1.788	-1.746	-1.767
	-0.871	-0.918	-0.786	-0.727	-0.878	-0.862	-0.908	-0.850
	-0.059	0.065	-0.303	-0.267	-0.092	-0.016	-0.374	-0.039
I feel full if energy these days	-1.424	-1.574	-1.201	-1.119	-1.298	-1.421	-1.355	-1.422
	-0.61	-0.623	-0.587	-0.518	-0.569	-0.585	-0.714	-0.593
	0.601	0.654	0.505	0.533	0.496	0.631	0.44	0.637
I feel that life is full of opportunities	-1.614	-1.779	-1.383	-1.288	-1.552	-1.611	-1.602	-1.595
	-0.761	-0.761	-0.76	-0.697	-0.753	-0.735	-0.896	-0.736
	0.428	0.489	0.316	0.334	0.361	0.469	0.276	0.472
I feel that the future looks good for me	-1.726	-1.839	-1.553	-1.468	-1.703	-1.718	-1.762	-1.703
	-1.034	-1.021	-1.06	-0.997	-1.04	-1.005	-1.208	-1.008
	0.198	0.269	0.064	0.114	0.148	0.243	-0.031	0.235

6.4 Judging the performance of CASP6 in terms of the information function obtained for various sample sources

Next we examine the information functions across our latent trait, QoL across five sample sources 'W/ALL', 'T/ALL', 'T/E', 'W&T/E' and 'W & T/ALL' or 'pooled' data contained in Figure 3 . The figure contains five lines where the horizontal axis spans -4 to +4 the extent of the latent scores ranging from 'low' to 'high' levels of QoL and the vertical axis maps the height of the information index along the underlying scale for QoL. Each line is to a greater or lesser extent 'bi-modal' quite distinctly visible for 'W/ALL', 'W&T/ALL' and 'W&T/E' whereas the two lines for 'T/ALL' and 'T/E' are flatter and run close together over the range of latent QoL. What we see is that the respondents who furnish their ratings online provide 'maximal' information across low and high levels of QoL and less so for middle positions on our latent scale. Varying combinations of web and telephone data reveal a similar pattern but are less marked. In contrast telephone respondents have a flatter information function across the low-middle-high levels of QoL. The finding that QoL is measured with higher precision in 'W&T/E' compared to the 'T/ALL' and 'T/E' groups indicates that the observed differences in the information function are not due to selection since allocation to the W&T/E' was random. It appears that highest precision in this group is driven by Web

respondents, which is also confirmed by the finding that highest precision is achieved in the ‘W/ALL’ group.



6.5 Illustrating the relative performance of CASP6 and CASP12v2 in predictive models of employment status

Finally, in table 7 we compare the performance of the summated values for CASP6 based on NCDS age 55 years sweep for the pooled (W&T/ALL) data using a probit regression analysis to predict current employment status (being unemployed, sick/disabled or on a government training scheme versus employed) both controlling for mode or not compared to an equivalent analysis using CASP12v2 for the NCDS age 50 years sweep. This is simply for illustration and no assumptions are made about causation or the direction of effects. The regression coefficients (standardised values) are almost identical. For CASP6 mode appears to make very little difference to the estimation at all and a similar inference could be made when comparing the impact of CASP6 to CASP12v2. Key social indicators (shown in table A.2) correlate positively albeit weakly with the exception of self-assessed general health with QoL as measured by the sum of the six CASP items. This would suggest that at age 55 years your general health does influence your QoL or vice versa. Being assigned to a relatively privileged social class, living with a partner, being in work and owning your own home all individually raise quality of life but leave sufficient unexplained variation to other facets of life.

Table 7. Probit coefficients and 95% confidence intervals of the association between various versions of the CASP and employment status (being currently unemployed, sick/disabled/on government training scheme versus employed)

	b	*95% CI*
CASP6 for pooled (W&T/ALL) data controlling for mode	-0.363	-0.392 to -0.333
CASP6 for pooled (W&T/ALL) data not controlling for mode	-0.364	-0.393 to -0.334
CASP12v2 under NCDS age 50 years	-0.354	-0.385 to -0.323

Section 7. Discussion and implications of findings for users

Survey researchers will be aware that switching mode to the telephone or the internet for the administration of the same items will typically imply a decline in the overall levels of response cooperation (Christian et al., (2014); Couper, (2008)) compared to the traditional face to face mode of data collection. Less is known about how robust attitude measures and scale scores will be across different modes of data collection. What this paper has accomplished is an empirical evaluation of a shortened version of a well-known measure of quality of life, CASP19 and its already shortened forms, CASP12v1 and CASP12v2 in order to provide strong statistical evidence for a single birth cohort aged 55 years in NCDS that summated scale scores of quality of life based on six items are robust and useful indicators. Once a second 'wording' factor is included to highlight the separation between negatively and positively worded items we have good evidence that the shortened version has sound measurement properties across different modes of data collection, namely the telephone and the web and it appears to hold up as well as CASP12 v2 as a predictor or associate of key measures like employment status. Indeed, there is additional evidence that use of the online approach produces greater discrimination over varying levels of quality of life compared to the use of the telephone. These differences are not strong enough for the survey designer to preference one mode of data collection over another. Our evidence based on NCDS members aged 55 years does suggest that CASP6 works well in both web and telephone modes of data collection. Similarly, a sequential mixed mode delivery will work for those respondents who prefer to use the telephone over the web.

Interestingly, in a recent paper by Brown et al., 2017, which uses the experimental allocation to mixed mode or telephone only to examine the extent to which there are mode effects on response rates and responses provided to a wide range of observed variables, it is shown that the means for all CASP6 items are lower for mixed mode responders (as a result of the online responders) when compared to the telephone only responders. A similar difference occurs when we contrast the latent means by mode (not shown here) for this evaluation, in a direct comparison between Web and Telephone, but less so when the two experimental conditions are compared. We would argue therefore that based on the available evidence for 55-year olds in NCDS that the measurement properties of the scale remain robust under either mode.

Furthermore, considering that our approach corrects for measurement error, it follows that the between modes difference in the latent mean is largely due to selection. This interpretation is further reinforced by the much less pronounced latent mean difference between the two experimental conditions where selection effects are not possible due to randomisation, a finding that showcases the effectiveness of our approach in correcting for measurement error.

In a world where time and budgets were more plentiful we might have been able to pilot the short form prior to execution. In reality, our evaluation has had to be pragmatic and risk not being able to adequately capture CMs ‘quality of life’. A fresh and informed approach to selecting a balanced set of items from CASP19 has provided a valuable measure which is quick and easy to administer. There is clearly also potential for CASP6 to be applied in face-to-face applications in the future.

References

Abbott, R.A., Ploubidis, G.B., Huppert, F.A., Kuh, D., Wadsworth, M.E.J., & Croudace, T.J. (2006). Psychometric evaluation of predictive validity of Ryff's psychological well-being items in a UK birth cohort sample of women. *Health and Quality of Life Outcomes*, 4, doi:Artn 76Doi 10.1186/1477-7525-4-76.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable modelsand factor analysis: A unified approach* (Vol. 904): John Wiley & Sons.Bollen, K.A. & Long, Scott, J. (editors). (1993). Testing Structural Equation Models. A SAGE Focus Edition, SAGE Publications Ltd., Newbury, CA.

Bǒrsch-Supan, A., Brugiavini, Jϋrges, H., Makenbach,J., Siegrist, J., & Weber, G. (Eds.). (2005). Health, ageing and retirement in Europe. First results from the survey of health, ageing and retirement in Europe (SHARE). Manheim Research Institute for the Economics of Ageing (MEA).

Brown, M. Calderwood, L., Goodman, A., Harding, C., Howat, N., Pickering, E., Ploubidis, G.B. and Williams, J. (2017). The impact of using the web in a mixed mode follow-up of a longitudinal birth cohort study: evidence from the National Child Development Study. CLS Working Paper, UCL-Institute of Education, University College London.

Campbell, A., Walker, J. & Farrell, G. (2003). Confirmatory factor analysis of the GHQ-12: can I see that again? Australian and New Zealand J of Psychiatry, 37:475-483.

Christian, L.M., Smyth, J.D. & Dillman, D.A. (2014). *Internet, phone, mail and the mixed mode survey: the tailored design method*. Wiley, USA.

Couper, M.J. (2008). *Designing effective Web Surveys*. New York: Cambridge University Press.

DeVellis, R.F. (2012). *Scale development: theory and applications*. Applied Social Research Methods Series No. 26, SAGE. Editors: Bickman, L. and Rog, D.J.

Enders, C.K. (2010). Applied Missing Data Analysis. New York, Guildford Press.

Goldberg, D. And Williams, P. (1988). The user's guide to the general health questionnaire. Windsor: NFER-NELSON.

Brown, M. Calderwood, L., Goodman, A., Harding, C., Howat, N., Pickering, E., Ploubidis, G.B. and Williams, J. (2017). The impact of using the web in a mixed mode follow-up of a longitudinal birth cohort study: evidence from the National Development Study. CLS Working Paper, UCL-Institute of Education, University College London.

Hazlett-Stevens, H., Ullman, J.B., & Craske, M.G. (2004). Factor structure of the Penn State Worry Questionnaire_ Examination of a method factor. *Assessment,* 11(4), 361-370. Doi:10.1177/1073191104269872

Hyde, M., Wiggins, R.D., Higgs, P. and Blane, D. (2003). A measure of Quality of Life in Early Old Age: the theory, development and properties of a needs satisfaction model. Aging and Mental Health 7(3), 186-194.

Hyde, M., Higgs, P., Wiggins, R.D. & Blane, D. (2015). A decade of research using the CASP scale: key findings and future directions. Aging & Mental Health, 19:7, 571-575, DOI: 10.1080/13607863.2015.1018868

Jo, M.S. (2000). Controlling social desirability bias via method factors of direct and indirect questioning in structural equation models. *Psychology & Marketing,* 17(2), 137-148. Doi:10.1002/(sici)1520-6793(2000020)17:2<137::aid-mar5>3.0.co;2-v.

Joseph, S., Linley, A. P., Harwood, J., Lewis, C.A. & McCollam, P. (2004). Rapid Assessment of well-being: the short depression-happiness scale (SDHS). Psychology and Psychotherapy: Research and Practice. The British Psychological Society.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.

Muthén, B. (1984). A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika, 49*(1), 115-132.

Muthén, L.K.., & Muthén, B.O. (2004). Mplus. Statistical Analysis with latent variables. Los Angeles: User's Guide. Muthén and Muthén.

Ploubidis, G.B. & Grundy, E. (2009). Later-Life Mental Health in Europe: A Country-Level Comparison. *Journals of Gerontology Series B- Psychological Sciences and Social Sciences,* 64(5), 666-676. Doi:10.1093/geronb/gbp026

Rabe-Hesketh, S., & Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research, 17*(1), 5-32. doi: 10.1177/0962280207081236|issn 0962-2802

Wiggins, R.D., Netuveli, G., Hyde, M., Higgs, P. & Blane, D. (2008). The evaluation of a self-enumerated scale of quality of life (CASP-19) in the context of research on ageing: a combination of exploratory and confirmatory approaches. *Social Indicators Research*, 89:61-77. DOI 10.1007/s11205-007-9220-5.

Ware, J.E., Kosinski, M., Turner-Bowker, D.M., & Gnadek, B. (2002). How to score version 2 of the SF-12. Health Survey (with Supplement documenting Version 1). Lincoln: QualityMetric Incorporated.

Appendix

Table A.1: CASP statements for the original 19-item version of CASP together with CASP12v1, CASP12v2 and CASP6.

Note: the main body of the list below contains the 12 statements for CASP12v2 listed by life domains control, autonomy, self-realisation and pleasure where the six *italicised* statements define CASP6 as administered in the age 55 year sweep. The labels for each CASP item contain the number identifier for the original 19-item scale item position. See further notes for item exchange to form CASP12v1 and the remaining items excluded from CASP19.

Q1a: Here is a list of statements that people have used to describe their lives and how they feel. We would like to know how often, if at all, you think this applies to you…

(please tick one box)

1. *Often*
2. *Sometimes*
3. *Rarely*
4. *Never*

CONTROL

CASP1 My age prevents me from doing the things I would like to+

CASP2 I feel that what happens to me is out of my control +

CASP4 I feel left out of things+

AUTONOMY

CASP5 I can do the things that I want to do

CASP7 I feel that I can please myself what I do++

CASP9 Shortage of money stops me from doing the things I want to do +

PLEASURE

CASP10 I look forward to each day

CASP11 I feel that my life has meaning

CASP12 I enjoy the things that I do++

SELFREALISATION

CASP15 I feel full of energy these days

CASP18 I feel that life is full of opportunities

CASP19 I feel that the future looks good for me

+ reverse coded

++ CASP7 replaced by CASP6 ‘family responsibilities prevent me from doing what I want to do’ and CASP12 replaced by CASP14 ‘On balance, I look back on my life with a sense of happiness’ to form CASP12v1.

Items contained in CASP19 by life domain excluded from all version above

Control: CASP3 ‘I feel free to plan for the future’

Autonomy: CASP8 ‘My health stops me from doing the things I want to’

Pleasure: CASP13 ‘I enjoy being in the company of others’

Self-Realisation: CASP16 ‘I choose to do the things that I have never done before’, CASP17, ‘I feel satisfied with the way my life has turned out’.

Table A.2: inter correlations of CASP6 items* with totals score and selected characteristics age 55 years

	My age prevents me from doing the things I would like to do	I feel what happens to me is out of my control	I feel left out of things	I feel full of energy these days	I feel that life is full of opportunities	I feel that the future looks good for me	**Age 55 - CASP6 Score**	Age 55 - general health*	Age 55 - whether lives with a partner**	Age 55 - work status***	Age 55 - whether higher social class (NS-SEC cats 1-3)****	Age 55 - Owner Occupier*****
My age prevents me from doing the things I would like to do		.357**	.346**	.293**	.248**	.253**	.605**	.323**	.041**	.096**	.056**	.086**
I feel what happens to me is out of my control			.555**	.306**	.326**	.375**	.710**	.351**	.116**	.191**	.085**	.101**
I feel left out of things				.309**	.334**	.384**	.701**	.313**	.157**	.195**	.076**	.106**
I feel full of energy these days					.503**	.502**	.695**	.498**	.081**	.195**	.132**	.140**
I feel that life is full of opportunities						.656**	.724**	.356**	.083**	.158**	.147**	.121**
I feel that the future looks good for me							.745**	.410**	.165**	.191**	.149**	.171**
Age 55 - CASP6 Score								.539**	.152**	.242**	.152**	.172**
Age 55 - general health									.128**	.287**	.221**	.215**
Age 55 - whether lives with a partner										.104**	.072**	.264**
Age 55 - work status											.451**	.185**
Age 55 - whether higher social class (NS-SEC cats 1-3)												.205**

* General health recorded using 5 point scale - Excellent to Poor - here re-coded so that high scores = better health
** 1 = lives with a partner (married, civil or cohabiting), 0 = no co-residential partner, *** 1 = in work, 0 = not in work,
**** 1 = NS-SEC cats 1-3, 0 = NS-SEC cats 4-8,
***** Recoded 1 = Owner occupier, 0 = Not owner occupier (note that tenure missing for a substantial proportion becaus

***Cronbach's Alpha for 6 –item version of CASP = 0.79**