

Happy Talk: Mode of Administration Effects on Subjective Well-Being

Paul Dolan¹ · Georgios Kavetsos¹

Published online: 6 May 2015

© Springer Science+Business Media Dordrecht 2015

Abstract There is increasing interest in subjective well-being (SWB) both in academic and policy circles. As a result, considerable research efforts are now being directed at the validity and reliability of SWB measures. This study examines how SWB reports differ by survey mode. Using data from the April 2011 to March 2012 Annual Population Survey in the UK we find that individuals consistently report higher SWB over the phone compared to face-to-face interviews. We also show that the determinants of SWB differ significantly by mode, with life circumstances tending to matter more in face-to-face interviews. These results have substantial implications for research and policy purposes.

Keywords Subjective well-being · Happiness · Survey mode · Valuation

JEL Classification D60 · I30

1 Introduction

Measures of subjective well-being (SWB) are increasingly becoming prominent in academic circles (Frey and Stutzer 2002; Di Tella and MacCulloch 2006; Dolan et al. 2008), but also amongst policy-makers interested in monitoring progress and evaluating interventions (Stiglitz et al. 2009; Fujiwara and Campbell 2011; National Research Council 2013; OECD 2013). This increasing interest originates from the (a) theoretically rigorous, (b) policy relevant, and (c) empirically robust nature of these measures (Dolan and Metcalfe 2012). Much empirical research has focused on the third category—examples include Sutton and Davidson (1997), Bertrand and Mullainathan (2001), Urry et al. (2004), Krueger and Schkade (2008), Conti and Pudney (2011), Kavetsos et al. (2014).

✉ Georgios Kavetsos
g.kavetsos@lse.ac.uk

¹ London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

A largely neglected question within this research area, however, is the degree to which SWB reports are influenced by the mode of survey administration. Do individuals report consistently higher/lower levels of SWB in one mode versus others? Additionally, and equally important, do the determinants of SWB differ by mode of administration? If so, there are obvious implications for the interpretation of, and policy recommendations resulting from, SWB analysis. This study provides some evidence around these issues by comparing SWB responses obtained via telephone and face-to-face surveys.

There is broad evidence suggesting that responses to subjective questions are likely to be affected by the mode of administration (Schwarz et al. 1991; Bowling 2005; Sakshaug et al. 2010). Although it is not feasible to cover the entire breadth of the literature in this area, some notable evidence on health-related quality of life (HRQoL) suggests that telephone surveys lead to higher—i.e. healthier—responses compared to self-completion ones, and F2F yields higher (more positive) responses compared to either telephone or self-completed questionnaires.

Some of these studies rely on cross-sectional comparisons between modes. For example, in a comparative analysis of health ratings between telephone and mail surveys in Australia, Perkins and Sanson-Fisher (1998) find higher health ratings for those interviewed on the telephone. McHorney et al. (1994) randomly allocate respondents to a mail and a telephone survey. They find that individuals interviewed over the phone report more healthy ratings and fewer chronic conditions on average. Buskirk and Stein (2008) study quality of life (QoL) scores of about 300 cancer survivors in the US, who are randomly assigned to mail or telephone interviews—disregarding the data retrieved from a third group of participants who were given the opportunity to self-select one of the two modes. Their results suggest higher reported QoL for those in the telephone mode, especially in respect to vitality, mental health and physical functioning; results which are also consistent across cancer types. Using four large population surveys in the US, Hanmer et al. (2007) find that telephone surveys yield higher HRQoL compared to postal mail and self-administered (under the presence of an interviewer) surveys amongst the elderly.

In a study comparing depression prevalence, Li et al. (2012) study self-reported data using large national surveys in the US. They find that the prevalence of depression is lower for those in the National Health and Nutrition Examination Survey (NHANES) sample—a personal interview—compared to those surveyed via the Behavioral Risk Factor Surveillance System (BRFSS)—a telephone survey.

Outside the realm of HRQoL, Breunig and McKibbin (2011) use data of self-reported financial difficulty from the 2003 to 2004 wave of the Australian Household Expenditure Survey (HES) and the third wave of the Household Income and Labour Dynamics in Australia (HILDA). They find that reported financial difficulty is about 33–36 % lower in F2F compared to self-completion interviews; an effect plausibly caused by the embarrassing association such a confession to another person has.

Other studies are able to offer more convincing evidence, based on a panel of individuals. In summary, the evidence resulting from this set of studies tends to confirm the strong influence of mode effects as already described. For example, Lyons et al. (1999) obtain postal, self-administered, questionnaires assessing respondents' health status before randomly allocating them in either a F2F interview (administered in the clinic) or a similar self-completion mode. About 200 individuals were recruited for this exercise, with the distance between the two surveys being 10 days. Results suggest a significant increase in health status in F2F interviews. In a similar study of a few hundred patients, Hays et al. (2009) find that those entering a heart failure program and cataract surgery report higher HRQoL when the follow-up survey is randomly administered over the telephone compared

to postal mail. Evans et al. (2004) survey a small sample of about 100 visitors to a UK general practice twice within a 2-day window alternating the order of F2F and telephone modes. They generally find limited evidence of mode effects in responses to the General Health Questionnaire and the Clinical Interview Schedule. Finally, Conti and Pudney (2011) study differences in levels of reported job satisfaction in the British Household Panel Survey (BHPS), obtained both via a F2F interview and a self-completed questionnaire. They find higher reports of job satisfaction in F2F interviews.

The evidence therefore suggests that respondents present themselves in a more positive light (e.g. healthier) the ‘closer’ the interviewer is. This is what Conti and Pudney (2011) call “put on a good show for the visitor” effect in F2F interviews. The social interaction of the respondent with the interviewer has been documented to lead to more socially desirable responses (Tourangeau and Smith 1996; Presser and Stinson 1998; Tourangeau and Yan 2007), though others have found limited effects of a social desirability bias (Fowler et al. 1999; Kaplan et al. 2001). On the other hand, an interviewer can “increase response and item response rates, maintain motivation with longer questionnaires, probe for responses, clarify ambiguous questions, help respondents with enlarged show-cards of response choice options, use memory jogging techniques for aiding recall of events and behaviour, and control the order of the questions” (Bowling 2005) and assure respondents on the confidentiality of their data (Nandi and Platt 2011).

Against this background, this study compares F2F and telephone responses obtained from the UK’s Office for National Statistics (ONS) SWB questions. We also examine whether the determinants of SWB differ by mode. An additional contribution of this study is the analysis of different SWB measures, including evaluative, experience and ‘eudemonic’ measures. These dimensions are in line with recent policy recommendations (Dolan and Metcalfe 2012; National Research Council 2013; OECD 2013) due to their fundamental importance, yet have been scarcely examined in the literature.

Our findings suggest that there are large differences by survey mode, with telephone respondents reporting higher levels of SWB. As expected, socio-economic determinants have a varying impact on the dimensions of SWB studied here. More importantly, however, we find a differing importance of the determinants of SWB within dimensions depending on the interview mode, where life circumstances tend to matter more in F2F surveys.

The rest of this study is organised as follows. Section 2 describes the data and methods used in the analysis. Section 3 presents the results of the analysis. Section 4 discusses the implications of the results for research and policy.

2 Data and Methods

We use SWB data gathered in the UK’s Annual Population Survey (APS), spanning over the period April 2011–March 2012. Following recommendations by Dolan and Metcalfe (2012), this survey introduces evaluative, experience and ‘eudemonic’ measures of SWB questions all of which are measured on a 0–10 scale, where 0 denotes ‘not at all’ and 10 denotes ‘completely’:

1. Overall, how satisfied are you with your life nowadays?
2. Overall, to what extent do you feel that the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?
4. Overall, how anxious did you feel yesterday?

The APS is a representative sample of the UK population, drawing data from the Labour Force Survey base and boost sample. In the base sample, a new cohort is selected in every quarter of the year and surveyed for five consecutive quarterly waves; the fifth and final interview thus takes place a year apart from the first. In the boost sample, a new cohort is selected every calendar year and is then surveyed for four consecutive annual waves; interviews are spread equally within each of the four waves and each interview takes place a year apart (LFS User Guide 2011). These data consist of surveys of the same address within a given cohort, hence not necessarily of the same individual.

The APS utilises all four waves of the boost sample, but only waves 1 and 5 of the base sample; this is depicted by the bold cells in Table 1. Respondents are interviewed either F2F or on the phone. In these data more than 165,000 individuals are asked the SWB questions: about 52 % of which via F2F (CAPI) and the remaining 48 % via telephone (CATI) interviews.

Allocation into mode is, however, not entirely clear. It appears that the norm for both the base and boost interviews is for the very first interview in Wave 1 to be conducted F2F. Selected addresses for which phone numbers are readily available though may have the first interview via the telephone. If F2F respondents concur, subsequent waves within the same cohort take place over the phone in order to minimise survey costs; otherwise they continue taking place F2F. The former is indeed what we observe in the data: although based on different cohorts, we find that about 78 % of respondents in the base sample are interviewed F2F in Wave 1 compared to about 25 % in Wave 5.

This non-random allocation into mode leads to issues of self-selection which cannot be overlooked. Therefore, any analysis on the effects of interview mode on SWB reports using the entire APS sample—that is, pooling data from waves 1 and 5 of the base sample and waves 1–4 of the boost sample—might be affected by respondents' self-selection into survey mode following the interview in Wave 1. It might be, for example, that happier individuals opt to have future interviews over the phone. Furthermore, recall that the APS is a survey of the same address within a given cohort, not necessarily of the same

Table 1 APS survey

	2011			2012
	Apr–Jun	Jul–Sep	Oct–Dec	Jan–Mar
Cohort A	W5 Base			
Cohort B	W4 Base	W5 Base		
Cohort C	W3 Base	W4 Base	W5 Base	
Cohort D	W2 Base	W3 Base	W4 Base	W5 Base
Cohort E	W1 Base	W2 Base	W3 Base	W4 Base
Cohort F		W1 Base	W2 Base	W3 Base
Cohort G			W1 Base	W2 Base
Cohort H				W1 Base
Cohort Boost A	2011 Boost W1			2012 Boost W2
Cohort Boost B				2012 Boost W1

'W' denotes wave. Cohort labels (i.e. A, B, ...) are illustrative. Bold cells represent waves in the APS data asking the SWB questions. 'W1' cells represent waves in our analysis

Source: LFS User Guide (2011), p. 12 and 18–19

individual if this person has moved. Hence, mode selection of the property's previous resident determines, to some extent, the interview mode of future residents; especially if the previous resident opts for F2F interviews. In these data we are, however, able to differentiate between the waves each interview was held. Hence, by focussing on Wave 1 respondents only, we remove subsequent biases likely to be caused by self-selection into mode. The data analysed in this study is depicted by the shaded cells in Table 1.

We run an OLS regression model of the following form to estimate the magnitude of the mode effect:

$$SWB_i = a_0 + a_1PHONE_i + a_2DEMO_i + R_s + M_t + D_l + e_i \quad (1)$$

where SWB denotes the score given in a SWB question by individual i ; $PHONE$ is a dummy variable denoting whether the survey was conducted over the telephone rather than F2F; $DEMO$ is a set of socio-demographic characteristics available for the respondent, including age, age squared, gender, marital status, employment status, education level, and ethnicity. As discussed in the previous section, respondent's self-rated health is susceptible to mode effects and is thus excluded from the regressions, which do however control for more objective health indicators, such as disability. R_s is a set of s regional dummy variables. M_t is a set of monthly (time of year) dummy variables and D_l is a set of daily (day of the week) dummy variables, which are shown to affect responses to SWB responses considerably (Kavetsos et al. 2014).¹

Note that household income is not recorded in these data and thus cannot be added as a control variable in our regressions. In a separate set of results we do however include an income measure, gross weekly pay (in main and secondary job), and re-estimate Eq. (1)—though this variable is, by definition, only relevant for respondents who are either employed or are under government employment.

We estimate Eq. (1) using OLS. Although SWB responses are ordinal, assuming cardinality and estimating Eq. (1) using OLS instead of an ordered probit or logit model has been shown not to alter results significantly and also adds to their interpretability (Ferrer-i-Carbonell and Frijters 2004)—we nonetheless also report results using an ordered probit model. In all cases, standard errors are robustly estimated and clustered at the regional level.

3 Results

3.1 Descriptives and t Tests

Table 2 offers some descriptive statistics of the Wave 1 sample of these data by mode. The composition of the sample is in some cases similar between modes—e.g. proportion between genders and education variables—but differs in others. For example, respondents are on average 10 years older on the phone, where the proportion of those being married and widowed is also higher. The proportion of the unemployed is doubled in the F2F sample, and that of being inactive decreases by 11 % points.

Table 3 reports the correlation coefficients between the measures by mode. LS, worthwhile and happiness are positively correlated (Dolan and Metcalfe 2011), and are all negatively correlated with anxiety. Interestingly, correlation coefficients do not change by

¹ Although it is worth noting that omitting calendar controls has only a marginal impact on the remaining coefficients.

Table 2 Summary statistics by mode

	Phone	Face-to-face
Male	41.2 %	44.6 %
Age	59.2	49.7
Married	61.8 %	46.8 %
Civil partnership	0.1 %	0.3 %
Separated	1.7 %	3.8 %
Divorced	8.7 %	11.7 %
Widowed	14.3 %	8.8 %
Degree	20.8 %	23.3 %
Higher education	12.1 %	9.8 %
General certificate of education (GCE)	23.7 %	21.3 %
General certificate of secondary education (GCSE)	21.7 %	21.8 %
Other education	10.4 %	9.9 %
Self-employed	7 %	7.7 %
Government employment and training programmes	0.1 %	0.2 %
Unpaid family worker	0.4 %	0.2 %
Unemployed	2.3 %	5 %
Inactive	51.8 %	40.9 %
Disability: DDA	22.9 %	13.6 %
Disability: work-limiting	2.3 %	3.4 %
Disability: DDA and work-limiting	8.8 %	12.3 %

Table 3 Correlations between SWB measures

	LS	Worthwhile	Happiness	Anxious
Face-to-face				
LS	1			
Worthwhile	0.63	1		
Happiness	0.58	0.51	1	
Anxious	-0.34	-0.25	-0.46	1
Telephone				
LS	1			
Worthwhile	0.61	1		
Happiness	0.57	0.49	1	
Anxious	-0.31	-0.24	-0.45	1

mode; an important finding in its own right, suggesting that the relationship between SWB measures remains consistent irrespective of mode.

Next, we calculate *t* tests in order to compare average SWB scores between modes. The results are presented in Table 4. For LS, worthwhile and happiness, average SWB is higher over the phone than F2F interviews by nearly half a point on a 0–10 scale. Anxiety is also reduced when this question is administered over the phone, by about a third of a point.

Table 4 Average SWB by survey mode

	Telephone		F2F		Mean difference
	<i>N</i>	Mean (SD)	<i>N</i>	Mean (SD)	
LS	17,660	7.796 (1.796)	59,117	7.303 (1.985)	0.493**
Worthwhile	17,572	8.054 (1.662)	58,942	7.603 (1.879)	0.452**
Happiness	17,667	7.683 (2.15)	59,110	7.217 (2.276)	0.465**
Anxious	17,615	2.916 (2.926)	59,053	3.218 (2.917)	-0.302**

N denotes number of observations

** *p* < 0.01

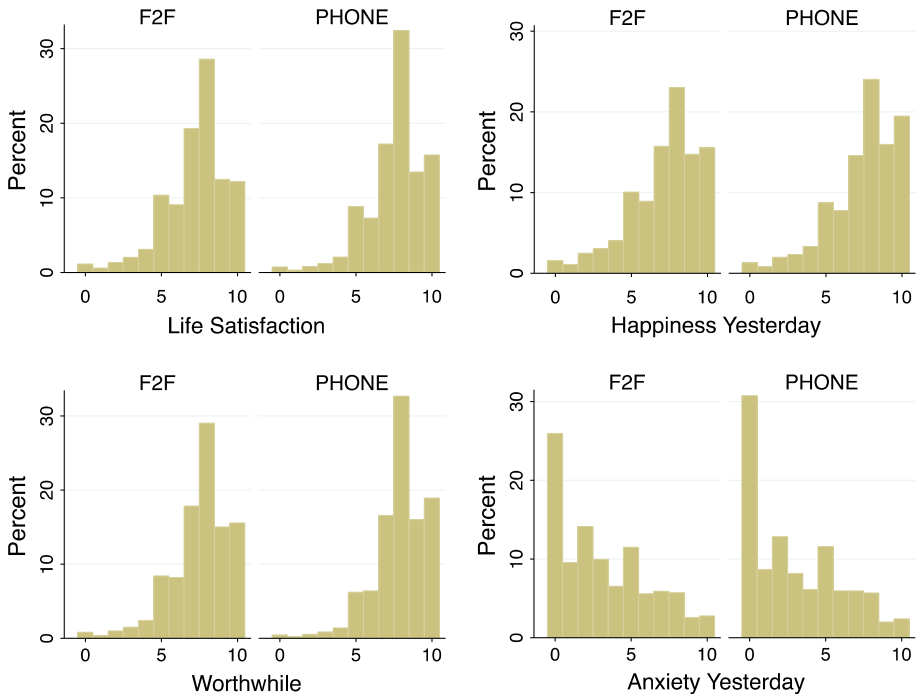


Fig. 1 SWB by survey mode

Rather than focusing on averages, Fig. 1 plots the distribution of SWB responses by measure and mode of administration. The histograms for life satisfaction (LS), worthwhile, and happiness yesterday (henceforth, happiness) corresponding to telephone interviews are right-skewed. That is, there is a larger percentage of respondents scoring 8–10 over the phone. This is also the case for anxious yesterday (henceforth, anxious), where notably a

Table 5 Regression results, OLS

	LS	Worthwhile	Happiness	Anxious
Phone	0.264** (0.029)	0.235** (0.021)	0.191** (0.019)	-0.10* (0.036)
Male	-0.127** (0.014)	-0.326** (0.009)	-0.085** (0.023)	-0.23** (0.023)
Age	-0.123** (0.006)	-0.076** (0.005)	-0.097** (0.005)	0.089** (0.007)
Age ²	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	-0.001** (0.0001)
Marital status (reference category: single)				
Married	0.553** (0.026)	0.467** (0.027)	0.478** (0.032)	-0.17** (0.025)
Civil partnership	0.892** (0.087)	0.677** (0.116)	0.618** (0.109)	-0.299 (0.148)
Separated	-0.317** (0.072)	-0.002 (0.037)	-0.124 (0.06)	0.227* (0.102)
Divorced	-0.116** (0.034)	0.045 (0.032)	-0.034 (0.04)	0.112** (0.031)
Widowed	-0.337** (0.046)	-0.001 (0.052)	-0.262** (0.068)	0.117 (0.069)
Education (reference category: no education)				
Degree	0.274** (0.029)	0.373** (0.024)	0.25** (0.031)	-0.114 (0.054)
Higher education	0.184** (0.032)	0.302** (0.03)	0.186** (0.03)	-0.135* (0.054)
General certificate of education (GCE)	0.149** (0.028)	0.251** (0.02)	0.139** (0.026)	-0.166** (0.044)
General certificate of secondary education (GCSE)	0.108** (0.035)	0.188** (0.024)	0.122** (0.036)	-0.192** (0.045)
Other	0.148** (0.037)	0.203** (0.027)	0.219** (0.041)	-0.185* (0.067)
Employment status (reference category: employed)				
Self-employed	-0.01 (0.022)	0.125** (0.028)	0.029 (0.027)	0.031 (0.027)
Government employment and training programmes	-0.812* (0.285)	-0.51 (0.246)	-0.012 (0.38)	0.349 (0.364)
Unpaid family worker	0.07 (0.133)	0.091 (0.137)	0.032 (0.152)	0.05 (0.274)
Unemployed	-0.969** (0.03)	-0.693** (0.045)	-0.511** (0.038)	0.433** (0.067)
Inactive	-0.151** (0.032)	-0.173** (0.033)	-0.048 (0.035)	0.084 (0.038)
Current disability (reference category: not disabled)				
DDA	-0.413** (0.038)	-0.267** (0.027)	-0.457** (0.035)	0.572** (0.046)
Work-limiting	-0.425** (0.056)	-0.289** (0.057)	-0.363** (0.064)	0.44** (0.065)
DDA and work-limiting	-1.209** (0.02)	-0.901** (0.024)	-1.119** (0.029)	1.322** (0.051)
<i>N</i>	61,720	61,583	61,723	61,660

Table 5 continued

	LS	Worthwhile	Happiness	Anxious
R^2	0.136	0.097	0.072	0.045

Regressions are OLS. Month of the year effects, day of the week effects, ethnicity and regional controls, and a constant term are included in the regressions. Standard errors, clustered at the regional level, reported in parentheses

** $p < 0.01$; * $p < 0.05$

much larger percentage of respondents report being ‘not at all anxious’ (i.e. a score of zero) on the telephone. A Kolmogorov–Smirnov test—a non-parametric test on the equality of two independent distributions—also suggests that the SWB measures do not have the same distribution function between modes: combined $K-S_{LS} = 0.121$ ($p < 0.001$), combined $K-S_{Worth} = 0.109$ ($p < 0.001$), combined $K-S_{Hap} = 0.096$ ($p < 0.001$), combined $K-S_{Anx} = 0.073$ ($p < 0.001$).

3.2 Regression Results

The results in Table 4 highlight the existence of statistically significant differences between modes of administration, but are not very informative of the relative effects of these differences. An investigation of the effects of survey mode via regression analysis is required for this purpose. The results, reported in Table 5, suggest that the phone mode increases SWB, with the maximum increase being observed for LS (about 0.26 points), followed by worthwhile (about 0.24), happiness (about 0.19), and anxiety (about -0.10).

On the remaining controls we find that the LS, worthwhile and happiness measures share some similarities. In general, men report being less satisfied, lower in worthwhile activities and less happy. SWB decreases with age, unemployment and disability; it increases with marriage, civil partnership, and education. Separated and divorced individuals report lower levels of LS; the divorced report lower LS and happiness. The signs of the estimated coefficients are often reversed when considering anxiety, in the sense that the individual characteristics that generally tend to be related with positive scores of LS and happiness are also related with negative scores of anxiety.

In terms of relative effects, for LS the coefficient of the telephone mode is twice as large as that of being a male, nearly as large as that of degree-level education, and more than alleviates the negative effects associated with divorce. For the worthwhile measure, telephone offsets two thirds of the negative effect associated with being a male, is about as large as having been educated up to GCE level, is twice as large as the level of worthwhile gained by being self-employed, and it alleviates about a third of the effect of unemployment. Similarly, for happiness the positive estimated coefficient for the telephone mode is more than twice as large as the negative effect associated with being male and is sufficient to offset more than half of the effect of widowhood. It is about as large as having a higher education, also offsets about a third of the negative effects of unemployment and half of the negative effect of having a work-limiting disability. For anxiety, it is half the coefficient of being a male, offsets the anxiety stemming from divorce, is nearly as large as having higher education, and it offsets about a fourth of the anxiety caused by work-limiting disability.

Table 6 reports separate regressions by mode to examine whether these determinants of SWB change by mode. We find several differences in the statistical significance of the

Table 6 Regression results by mode

	LS		Worthwhile		Happiness		Anxious	
	Phone	F2F	Phone	F2F	Phone	F2F	Phone	F2F
Male	-0.182** (0.023)	-0.11** (0.014)	-0.333** (0.024)	-0.321** (0.011)	-0.104** (0.029)	-0.078* (0.026)	-0.267** (0.069)	-0.222** (0.029)
Age	-0.121** (0.01)	-0.125** (0.007)	-0.081** (0.007)	-0.073** (0.006)	-0.108** (0.011)	-0.095** (0.005)	0.111** (0.013)	0.088** (0.008)
Age ²	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	-0.001** (0.0001)	-0.001** (0.0001)
Marital status (reference category: single)								
Married	0.594** (0.072)	0.545** (0.029)	0.479** (0.053)	0.464** (0.034)	0.552** (0.052)	0.466** (0.04)	-0.382** (0.082)	-0.138** (0.031)
Civil partnership	0.659* (0.268)	0.912** (0.093)	0.418 (0.293)	0.701** (0.112)	0.895 (0.467)	0.58** (0.129)	0.278 (0.826)	-0.351* (0.157)
Separated	-0.339* (0.153)	-0.308** (0.079)	-0.002 (0.144)	0.001 (0.043)	-0.002 (0.106)	-0.136 (0.07)	-0.122 (0.22)	0.272* (0.107)
Divorced	-0.067 (0.073)	-0.124* (0.043)	0.074 (0.058)	0.041 (0.034)	-0.282** (0.081)	-0.04 (0.05)	-0.233 (0.149)	0.173** (0.043)
Widowed	-0.449** (0.057)	-0.29** (0.054)	-0.022 (0.064)	0.004 (0.055)	0.041 (0.078)	-0.241** (0.068)	-0.176 (0.157)	0.18** (0.055)
Education (reference category: no education)								
Degree	0.03 (0.063)	0.32** (0.032)	0.112 (0.068)	0.421** (0.034)	0.072 (0.084)	0.29** (0.042)	-0.174* (0.078)	-0.094 (0.058)
Higher education	0.027 (0.096)	0.211** (0.028)	0.128 (0.074)	0.336** (0.038)	-0.007 (0.054)	0.205** (0.03)	-0.22 (0.13)	-0.107* (0.045)
General certificate of education (GCE)	-0.011 (0.061)	0.178** (0.033)	0.053 (0.049)	0.288** (0.027)	0.002 (0.082)	0.162** (0.033)	0.144 (0.086)	-0.163** (0.052)
General certificate of secondary education (GCSE)	-0.024 (0.081)	0.129** (0.034)	0.013 (0.069)	0.221** (0.028)	-0.106 (0.083)	0.146** (0.041)	-0.268* (0.10)	-0.174** (0.042)

Table 6 continued

	LS		Worthwhile		Happiness		Anxious	
	Phone	F2F	Phone	F2F	Phone	F2F	Phone	F2F
Other	0.044 (0.063)	0.166** (0.04)	0.16 (0.079)	0.204** (0.034)	0.158 (0.101)	0.23** (0.047)	-0.154 (0.116)	-0.195** (0.067)
Employment status (reference category: employed)								
Self-employed	0.018 (0.038)	-0.012 (0.029)	0.122** (0.039)	0.129** (0.03)	0.089 (0.05)	0.017 (0.034)	-0.031 (0.07)	0.047 (0.043)
Government employment and training programmes	-1.144 (0.702)	-0.775* (0.297)	-0.959 (0.691)	-0.449 (0.275)	-0.779 (0.73)	0.082 (0.417)	2.356* (0.766)	0.085 (0.435)
Unpaid family worker	-0.141 (0.134)	0.25 (0.205)	-0.107 (0.234)	0.223 (0.155)	-0.246 (0.151)	0.251 (0.194)	0.291 (0.427)	-0.169 (0.251)
Unemployed	-0.984** (0.07)	-0.966** (0.033)	-0.637** (0.104)	-0.696** (0.048)	-0.469** (0.103)	-0.515** (0.041)	0.79** (0.152)	0.394** (0.081)
Inactive	-0.041 (0.054)	-0.174** (0.033)	-0.095** (0.03)	-0.188** (0.039)	0.093 (0.051)	0.08 (0.043)	-0.163** (0.045)	0.142* (0.056)
Current disability (reference category: not disabled)								
DDA	-0.336** (0.058)	-0.449** (0.055)	-0.195** (0.038)	-0.297** (0.039)	-0.395** (0.047)	-0.479** (0.046)	0.57** (0.109)	0.566** (0.07)
Work-limiting	-0.506** (0.093)	-0.404** (0.062)	-0.30** (0.09)	-0.286** (0.059)	-0.389** (0.126)	-0.356** (0.072)	0.368** (0.166)	0.454** (0.068)
DDA and work-limiting	-1.074** (0.07)	-1.236** (0.023)	-0.716** (0.061)	-0.939** (0.033)	-0.891** (0.053)	-1.166** (0.029)	1.14** (0.11)	1.359** (0.05)
N	12,477	49,243	12,458	49,125	12,482	49,241	12,461	49,199
R ²	0.117	0.132	0.075	0.093	0.065	0.07	0.042	0.046

Regressions are OLS. Month of the year effects, day of the week effects, ethnicity and regional controls, and a constant term are included in the regressions. Standard errors, clustered at the regional level, reported in parentheses

** $p < 0.01$; * $p < 0.05$

demographic controls. Of equal importance, we find that the size of the estimated coefficients differs by mode. In most cases this difference is quite small, but there are cases where it is quite substantial. We present a few examples in Table 7.

3.3 Robustness and Additional Results

As previously mentioned the reliability of our estimates might be questioned due to the fact that we are missing an income variable. Table 8 offers estimates, similar to those of Table 5, including ‘gross weekly pay’. As mentioned in Sect. 2, this measure of income is not necessarily equivalent to household income, but can trivially be considered as a major component of it, and is only relevant to certain employment categories—these results are then equivalent to simply estimating the model for a sub-group of the sample which is under employment and as such might not be deemed to be very informative in our understanding around the exclusion of income. Table 8 thus also includes estimates for the same sub-group of individuals excluding this income measure in order to assess the robustness of the estimates once income is controlled for.

For brevity we present only some of the estimates of our model. We observe that the coefficient of the phone dummy is quite robust—for example, the estimate in LS regressions excluding income is 0.228, marginally increasing to 0.241 once income is included—and hold the same interpretation. For this sub-set of respondents, the coefficient of mode is not statistically significant for anxiety; irrespective of the inclusion of income. The estimates of the remaining controls are also quite robust to the inclusion of the income measure. Overall, this exercise relaxes our concerns about the implications the missing income variable might have on the size and statistical significance of the mode variable.

Finally, Table 9 presents estimates of Eq. (1), estimated via an ordered probit. The statistical significance of the main coefficient of interest, *phone*, is high. Its relative significance is also large—as was the case of the OLS estimates. For example, for LS the

Table 7 Differences in determinants of SWB by mode

	Employment	Education
LS	Economic inactivity reduces LS in F2F interviews, but not over the phone	All education variables are statistically insignificant on the phone, but positively significant in F2F interviews
Worthwhile	Economic inactivity reduces worthwhile by twice as much in F2F, compared to phone interviews	All education variables are statistically insignificant on the phone, but positively significant in F2F interviews
Happiness		All education variables are statistically insignificant on the phone, but positively significant in F2F interviews
Anxiety	Unemployment increases anxiety twice as much in phone compared to F2F interviews. A government employment/training programme increases anxiety by 2.4 points on a 0–10 scale in phone interviews, but has a statistically insignificant effect in F2F ones. (Note, however, that for phone interviews this estimate is based on 21 respondents only)	Similar to cases above, where in addition having a degree reduces anxiety in the telephone but has insignificant effect on F2F interviews

Table 8 Regression results including income

	LS		Worthwhile		Happiness		Anxious	
	No income	Income	No income	Income	No income	Income	No income	Income
	Phone	0.228** (0.037)	0.241** (0.038)	0.175** (0.028)	0.169** (0.028)	0.134** (0.026)	0.14** (0.033)	-0.07 (0.051)
Male	-0.047* (0.018)	-0.144** (0.023)	-0.239** (0.015)	-0.291** (0.019)	-0.032 (0.026)	-0.072* (0.028)	-0.273** (0.028)	-0.243** (0.026)
Age	-0.103** (0.006)	-0.119** (0.006)	-0.062** (0.005)	-0.071** (0.005)	-0.085** (0.005)	-0.094** (0.005)	0.077** (0.008)	0.084** (0.01)
Age ²	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	-0.001** (0.0001)	-0.001** (0.0001)
Gross Weekly Pay	0.001** (0.0001)	0.001** (0.0001)	0.0004** (0.0001)	0.0004** (0.0001)	0.0003** (0.0001)	0.0003** (0.0001)		-0.0002 (0.0001)
Current disability (reference category: not disabled)								
DDA	-0.259** (0.054)	-0.267** (0.044)	-0.153** (0.041)	-0.17** (0.038)	-0.321** (0.049)	-0.336** (0.042)	0.47** (0.076)	0.477** (0.078)
Work-limiting	-0.466** (0.058)	-0.477** (0.067)	-0.324** (0.079)	-0.337** (0.089)	-0.409** (0.077)	-0.427** (0.085)	0.531** (0.06)	0.513** (0.07)
DDA and work-limiting	-0.857** (0.03)	-0.813** (0.035)	-0.535** (0.027)	-0.521** (0.028)	-0.863** (0.056)	-0.849** (0.059)	1.061** (0.088)	1.013** (0.086)
N	34,287	30,053	34,236	30,016	34,282	30,049	34,259	30,033
R ²	0.075	0.08	0.047	0.048	0.035	0.0378	0.024	0.024

Regressions are OLS for those under employment. Controls include those under Eq. (1). Standard errors, clustered at the regional level, reported in parentheses

** $p < 0.01$; * $p < 0.05$

Table 9 Regression results, ordered probit

	LS	Worthwhile	Happiness	Anxious
Phone	0.157*** (0.016)	0.139** (0.013)	0.093** (0.009)	-0.047** (0.013)
Male	-0.082*** (0.009)	-0.206*** (0.007)	-0.052*** (0.011)	-0.071** (0.009)
Age	-0.07** (0.004)	-0.045** (0.003)	-0.048** (0.002)	0.033** (0.003)
Age ²	0.001** (0.0001)	0.001** (0.0001)	0.001** (0.0001)	-0.001** (0.0001)
Marital status (reference category: single)				
Married	0.321** (0.016)	0.275** (0.017)	0.225** (0.014)	-0.061** (0.009)
Civil partnership	0.513** (0.065)	0.401** (0.084)	0.287** (0.061)	-0.114 (0.059)
Separated	-0.157** (0.04)	0.012 (0.02)	-0.03 (0.027)	0.081* (0.036)
Divorced	-0.058** (0.018)	0.034 (0.019)	-0.004 (0.017)	0.039** (0.01)
Widowed	-0.192** (0.025)	0.001 (0.034)	-0.126** (0.031)	0.04 (0.024)
Education (reference category: no education)				
Degree	0.095** (0.019)	0.175** (0.016)	0.053** (0.016)	-0.007 (0.02)
Higher education	0.062** (0.018)	0.137** (0.018)	0.047** (0.017)	-0.035 (0.019)
General certificate of education (GCE)	0.045* (0.018)	0.107** (0.013)	0.029* (0.014)	-0.052** (0.016)
General certificate of secondary education (GCSE)	0.025 (0.021)	0.074** (0.017)	0.03 (0.018)	-0.067** (0.016)
Other	0.071** (0.021)	0.102** (0.018)	0.098** (0.018)	-0.072** (0.025)
Employment status (reference category: employed)				
Self-employed	0.006 (0.013)	0.095** (0.017)	0.023 (0.012)	0.011 (0.011)
Government employment and training programmes	-0.368* (0.147)	-0.237 (0.138)	0.054 (0.179)	0.093 (0.133)
Unpaid family worker	0.088 (0.082)	0.112 (0.086)	0.033 (0.077)	0.007 (0.097)
Unemployed	-0.496** (0.013)	-0.358** (0.022)	-0.21** (0.018)	0.146** (0.023)
Inactive	-0.043* (0.017)	-0.056** (0.017)	0.006 (0.017)	0.025 (0.013)
Current disability (reference category: not disabled)				
DDA	-0.239** (0.021)	-0.148** (0.015)	-0.208** (0.014)	0.202** (0.016)
Work-limiting	-0.252** (0.031)	-0.168** (0.033)	-0.173** (0.03)	0.156** (0.024)
DDA and work-limiting	-0.628 (0.013)	-0.467** (0.012)	-0.48** (0.011)	0.464** (0.017)

Table 9 continued

	LS	Worthwhile	Happiness	Anxious
<i>N</i>	61,720	61,583	61,723	61,660
<i>Pseudo-R</i> ²	0.034	0.023	0.016	0.01

Regressions are ordered probits. Month of the year effects, day of the week effects, ethnicity and regional controls, and a constant term are included in the regressions. Standard errors, clustered at the regional level, reported in parentheses

** $p < 0.01$; * $p < 0.05$

coefficient telephone is about twice as large (in absolute terms) as that of being a male and alleviates the negative effects associated with divorce separation; for worthwhile, it is about as large as higher education and it alleviates more than a third of the coefficient of unemployment; for happiness it is nearly twice as large as the negative effect associated with being male and offsets about half the coefficient of unemployment; and for anxiety, it more than the coefficient of divorce and offsets about a third of the coefficient of work-limiting disability.²

4 Discussion

On-going debate about the suitability of different measures of wellbeing for policy purposes has mostly concentrated on the differences between accounts, but there is now more of a focus on measurement issues *within* each account. This study highlights another dimension in the measurement realm by examining the impact of survey modes on reports of SWB. We use data from the UK Annual Population Survey between April 2011 and March 2012, which is the first to introduce the four SWB questions used by the ONS to measure SWB in the UK: life satisfaction, worthwhile, happiness yesterday, and anxiety yesterday. This survey was administered over the phone and by face-to-face interviews.

We find that the correlation coefficients do not change as a function of survey mode, implying that the relationship between the measures is stable irrespective of mode. Importantly, we do find that phone interviews are associated with significantly and substantially higher reports of SWB. We also report important differences in the statistical significance of different determinants of SWB by mode, where we find that the impact of these characteristics—notably in relation to education—is much less important in phone compared to F2F surveys.

Policy-makers have recently shown considerable interest in measuring SWB in order to monitor progress (e.g. Stiglitz et al. 2009)—the inclusion of the four SWB questions in the APS reflects David Cameron's, the UK's Prime Minister, pledge to measure SWB in the UK. Treating the entire sample uniformly by disregarding mode effects—see for example the ONS (2012) and NEF (2012) analyses of the same data—could result in misleading conclusions about the average levels of SWB in different regions, especially as some are interviewed using one mode only. For example, all interviews in Scotland, north of the Caledonian Canal, are currently administered by telephone only; in contrast, those in Northern Ireland are administered via F2F. As the collection of such SWB data increases in time, a uniform analysis will result in misleading time trends.

² Further ordered probit estimations by mode support the evidence based on OLS, presented in Table 6. These results can be made available upon request.

Such omission is not only relevant for SWB in the UK. Just as objective lists are sensitive to the choice of an indicator measuring a certain outcome, country rankings based on SWB scores might be sensitive to the mode of interview. Gallup have long been measuring SWB around the world, having employed both telephone and F2F interviews; with F2F being mostly used in developing nations. If the data in the Gallup World Poll follow the same pattern as that in the UK—with higher SWB reported over the phone—then appropriately controlling for survey mode might suggest a different gradient for the GDP-happiness relationship.

The importance of differing determinants of SWB is also of significant interest, as researchers and policy-makers need to understand the aetiology of SWB in order to ‘compensate’ for major life events (Blanchflower and Oswald 2004) and target population groups accordingly. The most interesting differences for policy purposes found across interview modes here relate to respondents’ employment status and education level. Economic inactivity reduces life satisfaction in F2F interviews—as has been found many times before (e.g. Dolan et al. 2008)—but it does not in phone surveys. For education, we confirm previous findings of a significantly positive association with life satisfaction in the F2F surveys, increasing with the level of education attained (Dolan et al. 2008), but the association disappears in the phone surveys.

These differences in the determinants of SWB by mode are as strong as differences found by measure. For example, unemployment seems to matter about twice as much for evaluative measures of SWB (life satisfaction) than for more experience-based measures (happiness yesterday)—supporting evidence in Knabe et al. (2010) from DRM data suggesting that the unemployed compensate losses in life satisfaction by being able to devote more time in enjoyable activities.

In addition, our results have substantial implications for policy evaluations using the SWB approach (Fujiwara and Campbell 2011; HM Treasury 2011). This has recently gained popularity in valuing a range of intangibles and non-market goods, including airport noise (van Praag and Baarsma 2005), air quality (Levinson 2012), natural disasters (Luechinger and Raschky 2009), health (Oswald and Powdthavee 2008; Dolan 2011), sports events (Kavetsos and Szymanski 2010), as well as SWB population patterns (Frijters et al. 2004). The results presented in this study question the robustness of compensation schemes based on the SWB approach, as these are likely to differ depending on the interview mode used to make these evaluations.

Note that we cannot offer any prescriptions on the superiority of phone versus F2F surveys and we are also not considering differences in the time and cost for gathering survey data—these are important aspects of data collection that are best evaluated elsewhere (e.g. Dillman 2000; Groves 2004; McMorris et al. 2009). What we highlight here is that the results cannot be generated by social desirability bias alone, since interviewers are ‘closer’ to respondents in F2F surveys, yet people are happier in phone surveys. Because of the further ‘distance’ between the interviewer and the respondent in telephone surveys, the latter might not be allocating as much attention to the interview process as in F2F surveys, leading to inaccurate responses (Holbrook et al. 2003).

In a related way, it is also plausible that higher average SWB over the phone might be the result of gravitation of scores to the top end of the response scale because questions are hard to interpret and which, arguably, an interviewer could have clarified. This could be especially true for lower educated respondents (Dolan and Metcalfe 2011), which is a pattern we also observe here. For example, about 22.5 % of respondents with no educational qualification give a life satisfaction score of ‘ten out of ten’, compared to about 9 % of those with degree-level education.

This study does not come without limitations. First, there is the issue of the lack of a household income variable. In our robustness analysis we do however estimate models including gross weekly pay, available for those who are employed. Separate estimates for this specific sub-group based on the exclusion and inclusion of this variable offer fairly robust estimates of the impact of mode effects. Second, although our focus on Wave 1 respondents resolves much of the issues related to subsequent self-selection into interview mode in future waves, some degree of selection in the data still remains; for example, from those not listing their telephone in the directory (ex-directory), or those households who only have mobile phones. Third, we have no information on the number of telephone attempts made to respondents, which has been recently shown to affect happiness scores (Heffetz and Rabin 2013).

Given the importance of mode effects, future research should compare SWB scores between additional survey modes, such as online surveys and text messaging—which offer cost-efficient methods of data collection—as well as evidence between countries. Recent research also examines the relationship between SWB and capabilities by using self-reported indicators of freedoms and capabilities in various domains as explanatory variables of life satisfaction (Anand et al. 2011)—empathy, life autonomy, and safety are positively correlated to life satisfaction; worthlessness, stress, and the possibility of future discrimination are negatively correlated to life satisfaction. Our findings have potential implications on the differing impact such explanatory variables might have depending on the mode of administration used to elicit responses.

Mode effects have also been shown to influence preference satisfaction accounts as well, especially in eliciting willingness to pay (WTP)—the maximum amount an individual is willing to pay to receive or avoid a good or service—(Marta-Pedroso et al. 2007; Maguire 2009; Olsen 2009; Lindhjem and Navrud 2011). Leggett et al. (2003), for example, estimate a 23–29 % larger WTP in F2F interviews compared to self-administered surveys. Further research gathering evidence across a broader range of interview modes should be promoted and compared to other valuation methods, such as the SWB approach.

These are all avenues that future research could significantly contribute upon. But as things stand, the most cost-effective way to increase reported SWB in the UK is to conduct all interviews over the phone.

Acknowledgments We thank Femke De Keulenaer, Daniel Fujiwara, Stephen Hicks, Danny Kahneman, Laura Kudrna, Kate Laffan, Robert Metcalfe, Michael Norton, Tessa Peasgood and participants at the “New scholarship on happiness” conference held at Duke University, two anonymous referees and the editor for useful comments and suggestions that improved this paper considerably. Further thanks go to the UK Data Archive for provision of the APS data, and Necla Acik-Toprak of the ESDS Government, and Simon Woodsford and Dawn Snape of the ONS Social Survey Division for providing clarifications on the APS survey design.

References

- Anand, P., Krishnakumar, J., & Tran, N. B. (2011). Measuring welfare: Latent variable models for happiness and capabilities in the presence of unobservable heterogeneity. *Journal of Public Economics*, 95, 205–215.
- Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, 91, 67–72.
- Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88, 1359–1386.

- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27, 281–291.
- Breunig, R., & McKibbin, R. (2011). The effect of survey design on household reporting of financial difficulty. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 991–1005.
- Buskirk, T. D., & Stein, K. D. (2008). Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. *Journal of Clinical Epidemiology*, 61, 1049–1055.
- Conti, G., & Pudney, S. (2011). Survey design and the analysis of satisfaction. *The Review of Economics and Statistics*, 93, 1087–1093.
- Di Tella, R., & MacCulloch, R. (2006). Some uses of happiness data in economics. *Journal of Economic Perspectives*, 20, 25–46.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. New York: Wiley.
- Dolan, P. (2011). *Using happiness to value health*. UK: Office of Health Economics.
- Dolan, P., & Metcalfe, R. (2011). *Comparing measures of subjective well-being and views about the role they should play in policy*. Newport: Office for National Statistics.
- Dolan, P., & Metcalfe, R. (2012). Measuring subjective wellbeing: Recommendations on measures for use by National Governments. *Journal of Social Policy*, 41, 409–427.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29, 94–122.
- Evans, M., Kessler, D., Lewis, G., Peters, T. J., & Sharp, D. (2004). Assessing mental health in primary care research using standardized scales: Can it be carried out over the phone? *Psychological Medicine*, 34, 157–162.
- Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, 114, 641–659.
- Fowler, F. J., Gallagher, P. M., & Nederend, S. (1999). Comparing telephone and mail responses to the CAHPS survey instrument: Consumer assessment of health plans study. *Medical Care*, 37, 41–49.
- Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, 40, 402–435.
- Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Investigating the patterns and determinants of life satisfaction in Germany following reunification. *Journal of Human Resources*, 39, 649–674.
- Fujiwara, D., & Campbell, R. (2011). *Valuation techniques for social cost-benefit analysis: Stated preference, revealed preference and subjective well-being approaches*. A discussion of the current issues, HM Treasury and Department for Work and Pensions.
- Groves, R. M. (2004). *Survey errors and survey costs*. Hoboken, NJ: Wiley.
- Hanmer, J., Hays, R. D., & Fryback, D. G. (2007). Mode of administration is important in US national estimates of health-related quality of life. *Medical Care*, 45, 1171–1179.
- Hays, R. D., Kim, S., Spritzer, K. L., Kaplan, R. M., Tally, S., Feeny, D., et al. (2009). Effects of mode and order of administration on generic health-related quality of life scores. *Value in Health*, 12, 1035–1039.
- Heffetz, O., & Rabin, M. (2013). Conclusions regarding cross-group differences in happiness depend on difficulty of reaching respondents. *American Economic Review*, 103, 3001–3021.
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparison of respondent satisficing and social desirability bias. *Public Opinion Quarterly*, 67, 79–125.
- Kaplan, C. P., Hilton, J. F., Park-Tanjasiri, S., & Perez-Stable, E. J. (2001). The effect of data collection mode on smoking attitudes and behaviour in young African American and Latina women. Face-to-face interview versus self-administered questionnaires. *Evaluation Review*, 25, 454–473.
- Kavetsos, G., Dimitriadou, M., & Dolan, P. (2014). Measuring happiness: Context matters. *Applied Economics Letters*, 21, 308–311.
- Kavetsos, G., & Szymanski, S. (2010). National wellbeing and international sports events. *Journal of Economic Psychology*, 31, 158–171.
- Knabe, A., Rätzl, S., Schöb, R., & Weimann, J. (2010). Dissatisfied with life but having a good day: Time-use and well-being of the unemployed. *Economic Journal*, 120, 867–889.
- Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, 92, 1833–1845.
- Leggett, C. G., Kleckner, N. S., Boyle, K. J., Duffield, J. W., & Mitchell, R. C. (2003). Social desirability bias in contingent valuation surveys administered through in-person interviews. *Land Economics*, 79, 561–575.
- Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics*, 96, 869–880.

- LFS User Guide. (2011). *LFS background and methodology* (Vol. 1). Newport: Office for National Statistics.
- Li, C., Ford, E. S., Zhao, G., Tsai, J., & Balluz, L. S. (2012). A comparison of depression prevalence estimates measured by the Patient Health Questionnaire with two administration modes: Computer-assisted telephone interviewing versus computer-assisted personal interviewing. *International Journal of Public Health, 57*, 225–233.
- Lindhjem, H., & Navrud, S. (2011). Are internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics, 70*, 1628–1637.
- Luechinger, S., & Raschky, P. A. (2009). Valuing flood disasters using the life satisfaction approach. *Journal of Public Economics, 93*, 620–633.
- Lyons, R. A., Wareham, K., Lucas, M., Price, D., Williams, J., & Hutchings, H. A. (1999). SF-36 scores vary by method of administration: Implications for study design. *Journal of Public Health Medicine, 21*, 41–45.
- Maguire, K. B. (2009). Does mode matter? A comparison of telephone, mail, and in-person treatments in contingent valuation surveys. *Journal of Environmental Management, 90*, 3528–3533.
- Marta-Pedroso, C., Freitas, H., & Domingos, T. (2007). Testing for survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics, 62*, 388–398.
- McHorney, C. A., Kosinski, M., & Ware, J. E. (1994). Comparisons of the costs and quality of norms of the SF-36 health survey collected by mail versus telephone interview: Results from a national survey. *Medical Care, 32*, 551–567.
- McMorris, B. J., Petrie, R. S., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2009). Use of web and in-person survey modes to gather data from young adults on sex and drug use: An evaluation of cost, time, and survey error based on a randomized mixed-mode design. *Evaluation Review, 33*, 138–158.
- Nandi, A., & Platt, L. (2011). *Effect of interview modes on measurement of identity*. Understanding Society Working Paper Series. No. 2011-02.
- National Research Council. (2013). *Subjective well-being: Measuring happiness, suffering, and other dimensions of experience*. Washington D.C: The National Academies Press.
- NEF. (2012). *Well-being patterns uncovered: An analysis of UK data*. London: New Economics Foundation.
- OECD. (2013). *OECD guidelines on measuring subjective well-being*. Paris: OECD Publishing.
- Olsen, S. B. (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental & Resource Economics, 44*, 591–610.
- ONS. (2012). *First ONS annual experimental subjective well-being results*. Newport: Office for National Statistics.
- Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics, 92*, 1061–1077.
- Perkins, J. J., & Sanson-Fisher, R. W. (1998). An examination of self- and telephone- administered modes of administration for the Australian SF-36. *Journal of Clinical Epidemiology, 51*, 969–973.
- Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review, 63*, 137–145.
- Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly, 74*, 907–933.
- Schwarz, N., Strack, F., Hippler, H.-J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*, 193–212.
- Stiglitz, J., Sen, A., & Fitoussi, J. (2009). *Report by the commission on the measurement of economic performance and social progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress.
- Sutton, S., & Davidson, R. (1997). Prefrontal brain symmetry: A biological substrate of the behavioural approach and inhibition systems. *Psychological Science, 8*, 204–210.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.
- Treasury, H. M. (2011). *The Green Book: Appraisal and evaluation in central government*. London: Treasury Guidance.
- Urry, H., Nitschke, J., Dolski, I., Jackson, D., Dalton, K., Mueller, C., et al. (2004). Making a life worth living. *Psychological Science, 15*, 367–372.
- Van Praag, B. M. S., & Baarsma, B. E. (2005). Using happiness surveys to value intangibles: The case of airport noise. *Economic Journal, 115*, 224–246.