



Does the Survey Mode Affect the Association Between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode

Giorgio Piccitto¹ · Aart C. Liefbroer^{2,3,4} · Tom Emery⁵

Accepted: 19 June 2022 / Published online: 14 July 2022
© The Author(s) 2022

Abstract

Subjective well-being research increasingly uses web surveys to understand how subjective well-being indicators are related to other concepts of interest. Although we know that mean scores on these indicators may differ between modes, we know little about whether a move to web will influence the conclusions we draw about our conceptual models. This study uses data from a unique mixed-mode survey collected in Croatia and Germany as part of the Generations and Gender Programme to examine whether the relationships between a range of subjective well-being indicators and a set of objective and subjective determinants differ between respondents answering these questions in face-to-face or web mode. Although respondents report lower subjective well-being in web than in face-to-face mode, the relationships between these variables and a range of objective and subjective indicators are relatively stable across modes. This suggests that substantive conclusions about antecedents of subjective well-being do not depend on whether data are collected via a face-to-face interview or through web survey.

Keywords Mode effect · Subjective well-being · Web survey · Face-to-face survey · Social desirability

✉ Giorgio Piccitto
giorgio.piccitto@unibocconi.it

¹ Dondena Centre for Research on Social Dynamics and Public Policies, Bocconi University, Via Sarfatti, 25, 20136 Milan, Italy

² Netherlands Interdisciplinary Demographic Institute, The Hague, The Netherlands

³ Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands

⁴ Department of Sociology, Vrije Universiteit, Amsterdam, The Netherlands

⁵ Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands

1 Introduction

Much of our knowledge about the determinants of subjective well-being is based on survey research. However, surveys differ in their data collection mode. For instance, in volume 22 (issue 1) of *Journal of Happiness Studies* (Jan 2021), 10 papers used data based on face-to-face interviews (F2F), 3 used data from mixed sources and 9 used data collected by internet-methods (WEB). Although F2F has been considered the 'gold standard', in recent years WEB is becoming increasingly common. The response rates of F2F surveys have shown a clear, though not uniform, decrease over time (Koen et al., 2018). Simultaneously, the costs of F2F have surged and act as a clear impediment to continued use, particularly in countries where labour costs are high. Concomitantly, internet penetration across the globe has risen sharply (Pandita, 2017; Daikeler et al., 2020), making the use of WEB surveys increasingly attractive, as people can answer questions whenever they feel ready to and with minimal interference. As a result, WEB has become more and more popular, either in combination with F2F (e.g. to boost response rates), leading to a kind of mixed-mode survey design, or as a replacement of F2F (Manfreda et al., 2008; Garbarski et al., 2019).

It is well-known, though, that the mode of data collection can influence the univariate distribution of variables (Epstein et al., 2001; Kreuter et al., 2008; Dillman et al., 2014; Toepoel, 2015). This is particularly true for many of the variables of key interest to subjective well-being researchers (Dolan & Kavetsos, 2016; Sarracino et al., 2017; Sanchez Tome, 2018). However, subjective well-being research is not only interested in univariate distributions of single variables, but also in testing theoretical models about the relationship between subjective well-being indicators and their antecedents. It is therefore surprising that only a few studies have examined what potential consequences the use of different modes of data collection have for the association between variables in multivariate models (see Martin and Lynn (2011); Dolan and Kavetsos (2016); Sarracino et al. (2017)). Whether our substantive conclusions about determinants of subjective well-being indicators depend on the survey mode with which the data have been collected is an issue we rarely reflect upon, let alone examine. This paper aims to start filling this gap in our knowledge by exploring the existence and magnitude of mode effects on the multivariate associations between subjective well-being indicators on the one hand, and a set of objective and subjective indicators on the other.

To examine mode effects on multivariate analyses, we use data from a pilot study conducted in 2018 in Germany and Croatia within the Generations and Gender Project (GGP), a cross-national social science data infrastructure which collects survey data in more than 20 countries (Gauthier et al., 2018). Respondents were randomly assigned to two different modes: (a) F2F; or (b) WEB. We compare to what extent multivariate models on the relationship between subjective well-being indicators and a set of determinants were affected by the mode that respondents were assigned to.

2 F2F Versus WEB: What Do We Know?

A rich literature exists identifying differences between on-line survey responses and those collected with other modes (telephone, face-to-face, self-administered) (Epstein et al., 2001; Currivan et al., 2004). Such differences are defined as mode effects, a concept which refers to systematic dissimilarity between data collected with different survey modes.

Generally, it has been found that people who answer subjective well-being related questions in WEB mode report less positively on their well-being than people who answer in other modes, like F2F and telephone (Sarracino et al., 2017; Sanchez Tome, 2018). However, this is not universally true (Hendriks et al., 2018). For instance, Martin and Lynn (2011) did not find differences in questions on life satisfaction between F2F and a mixed-mode approach including WEB.

One important mechanism that could drive the mode effect is social desirability bias, by which respondents over-report 'desirable' attitudes or behaviours and under-report 'undesirable' attitudes or behaviours, especially on questions related to sensitive topics (Edwards, 1957; Tourangeau & Yan, 2007). Social desirability bias rests on the idea that there are social norms governing a range of attitudes and behaviours, and that people may consciously or unconsciously misrepresent themselves to comply with these norms (Krumpal, 2013). This mechanism is assumed to be very important in F2F interviews, with the physical presence of an interviewer, while it is assumed to be less present in WEB surveys due to their perceived high level of anonymity and confidentiality (Aquilino, 1994; Kreuter et al., 2008; Schork et al., 2021). As a result of the differential susceptibility to social desirability bias of WEB and F2F, it can be expected that the univariate distributions of subjective well-being related items differ between WEB and F2F, with people generally giving more socially desirable responses in the latter than in the former.

Indeed, such mode effects on univariate distributions have been examined extensively in the literature. Focusing on the comparison between WEB and F2F, it has for instance been found that respondents in WEB mode score lower on a 'happy-healthy index' than respondents in F2F mode (Heerwegh, 2009). Zhang et al. (2017) reported much lower scores on subjective well-being indicators, like life satisfaction, happiness and positive mental health among respondents in WEB mode than in F2F mode. Generally, these studies are in line with the suggestion made above that WEB leads to less social desirability in the answering patterns for indicators than F2F.

However, we identified only three studies and one meta-analysis that explicitly examined the effect of survey mode on associations between subjective well-being indicators and their potential correlates, although none of them directly compared WEB and F2F modes. Sarracino et al. (2017) compared WEB and telephone mode in data from Luxembourg and examined whether the association between life satisfaction on the one hand and gender, age, education, employment status and household income on the other hand differed between both modes. The only statistically significant difference was that income effects on life satisfaction were stronger in WEB than in telephone mode. Sanchez Tome (2018) compared WEB with mail and telephone mode. She examined the relationship between happiness, social trust and job satisfaction on the one hand and a set of demographic and socio-economic determinants on the other. She did find differences in the strength of a number of associations, although no formal statistical tests were performed. Dolan and Kavetsos (2016) also examined the determinants of subjective well-being indicators, but compared F2F and telephone mode using data from the UK Annual Population Survey. They examined how different well-being indicators correlated with gender, age, partner status, level of education, employment status and disability status. They concluded that mode effects in associations exist, but generally are small. However, they did not explicitly test whether differences were statistically significant. Finally, Walter et al. (2019) conducted a meta-analysis in which they compared the association between several work-related subjective well-being variables and independent variables like leadership and personality that were observed in online panels to the same associations in conventionally sourced data, like F2F interviews. They concluded that most associations are more or less comparable in both

designs. A strong feature of this study is that it was able to compare many studies. However, the comparison is between F2F and online panels rather than between F2F and data collected from random WEB samples. Even more important, in their meta-analysis they were not able to make the data collection for the WEB and F2F modes completely comparable. Thus, the authors did not succeed in disentangling the 'real' mode effect from the individual self-selection (into the survey mode) effect.

Two other studies examined mode effects on associations between indicators not linked to subjective well-being. Martin and Lynn (2011) analysed data from an experiment conducted within the European Social Survey in which respondents who had taken part in a F2F interview were compared to respondents who either had a free choice between F2F, WEB and telephone interview or were pushed to WEB first, and if not responding were asked to conduct the interview by telephone, and if they were not willing to do so, by F2F. They compared a series of associations between F2F mode and the two mixed-modes combined. In modelling determinants of political interest, only two out of sixteen determinants showed a statistically significant mode effect in bivariate analyses and only one of them showed a statistically significant mode effect in a multivariate analysis. In modelling determinants of political activism (voter turnout, political involvement) about one in five determinants showed mode effects. Martin and Lynn (2011) concluded that there is evidence for mode effects on the associations of interest, but that it is not overwhelming. In the majority of examined associations no mode effects were detected. Jackle et al. (2010) compared associations between attitudes towards immigration and a set of nine other subjective indicators in a European Social Survey experiment comparing F2F and telephone mode and found only one statistically significant mode effect on associations.

This brief literature review shows that very few studies have examined mode effects on associations within multivariate models. No studies that exclusively compare F2F and WEB (the two most likely types of mode at this time) have been identified at all. This would be unproblematic if mode does not influence univariate distributions of variables. In that instance, it is likely to assume that no mode effect on the strength of the association between these two variables is present either. But, as mentioned earlier, we know that many subjective well-being indicators are influenced by mode effect. This makes it highly relevant to consider whether there is a mode effect on the strength of the association if a mode effect exists in one or both of the univariate distributions of the relevant variables.

3 Choosing a Set of Well-being Indicators and their Determinants

To explore whether associations between well-being and its determinants are subject to mode effects, we decided to focus on multivariate models that correlate a number of well-being indicators with a set of objective and subjective determinants, on the basis both of theoretical considerations and the availability of suitable indicators in the survey.

3.1 Subjective Well-being Indicators

Regarding the well-being indicators, we focus on a number of indicators related to well-being in key life domains. The four domains are: (i) work and employment; (ii) family; (iii) social relationships; (iv) health. Within the domain of work and employment, we examine satisfaction with the current job. In the family domain, we consider satisfaction with the

partner relationship. In the social relationships domain, we use loneliness as our key indicator. Finally, in the health domain, we choose people's subjective health assessment.

3.2 Objective Determinants of Well-being

We select a number of individual characteristics that are key predictors of well-being, namely age, gender and education; interestingly, the extent to which people are prone to provide socially desirable answers to questions on well-being may depend on these characteristics.

Men and women may differ in their tendency to provide social desirable answers (Fisher & Dube, 2005). How this works out for well-being indicators is not clear in advance, and may vary by whether satisfaction is measured in life domains that are traditionally more male-oriented or life domains that are traditionally more female-oriented.

Age is another interesting correlate to consider, as the tendency to provide socially desirable answers has often been found to increase with age (Hitchcott et al., 2020; Soubelet & Salthouse, 2011), which could lead to an inflated positive correlation between age and well-being factors. Given that WEB could reduce social desirability effects compared to F2F, using WEB might attenuate the association between age and well-being indicators.

As for education, Heerwig and McCabe (2009) stated that in general a tendency has been found for the higher educated to be more susceptible to report socially desirable behaviours. In their own research on US presidential candidates, however, they did not observe educational differentials in social desirability in the preference for a black president.

3.3 Subjective Determinants of Well-being

In addition, we decide to include work-life balance (*WLB* from here onwards) as a potential subjective correlate of well-being. Given that this indicator is included less frequently in models of subjective well-being, some additional reflections on this choice are in order. As [451]Pichler (2009) noticed, *WLB* 'is experienced when demands from the domain of (paid) work are compatible with demands from other domains'. Poor *WLB* can have negative consequences for different aspects of people's subjective well-being (Noda, 2020; Schnettler et al., 2021; Sirgy et al., 2020). In their review, Sirgy and Lee (2018) suggested that *WLB* is associated with work-related outcomes, such as job satisfaction, absenteeism and burn-out, with non-work-related outcomes, such as marital satisfaction, leisure satisfaction and family conflict, and with stress-related outcomes, such as depression, illness symptoms and alcohol consumption.

First, there are clear links between *WLB* and work-related outcomes. For instance, Haar et al. (2014) showed that *WLB* is positively related to job satisfaction across a range of different countries (France, Spain, Italy, Malaysia, China and New Zealand), while Uglanova and Dettmers (2018) found a positive association between flexible working time arrangements and job satisfaction both for men and women in Germany.

Second, *WLB* is related to family and social network outcomes. Muurlink et al. (2014) showed that work-life imbalance is not only associated with a lower relationship satisfaction for people experiencing the imbalance, but also with a lower relationship satisfaction reported by their spouse. Ferguson et al. (2012) in a longitudinal study showed that *WLB* is positively associated with both job satisfaction and relationship satisfaction. Given that *WLB* positively influences the time and energy people

have available for catering to their personal relationships, it can be expected that *WLB* is also negatively related to loneliness. Loneliness can be viewed as an imbalance between the quality and quantity of the relationships that one has and the quality and quantity of the relationship one wants (De Jong Gierveld & van Tilburg, 2006).

Third, *WLB* is related to health and other stress-related outcomes. It has been linked to mental health outcomes (Haar et al., 2014; Hagqvist et al., 2017; Lunau et al., 2014) and to physical health outcomes (Lunau et al., 2014). This potential link between *WLB* and well-being in different life domains makes it an ideal candidate to include in an exploration of mode effects on well-being associations.

Thus, this paper will empirically examine the effect of survey mode on the association between (i) indicators of subjective job-related, family-related, social relationship-related and health-related well-being and (ii) gender, age, education and *WLB*. This will be done in a multivariate framework.

4 Data and Methods

4.1 Data

Our data derive from an experiment in the context of the Generations and Gender Programme (GGP), which is a cross-national social science data infrastructure that collects micro- and macro-level data on demographically relevant topics in over 20 countries (Gauthier et al., 2018). In the past, the Generation and Gender Survey (GGS)-the focal survey instrument of the GGP- has been conducted primarily in F2F (CAPI) mode. In 2018, an experimental pilot study was conducted in three countries (Germany, Croatia, Portugal) to study the possibilities of implementing a push-to-web design (Lugtig et al., 2022). Respondents were aged 18 to 49 at the time of the survey. In Portugal, the overall experimental design of the pilot turned out to be problematic, as the fieldwork window was too short to realize an adequate sample and an individual listing of individuals was not available. Therefore, we restricted our analysis to data from Germany and Croatia where a listing of individuals was available for sampling.

In both countries, respondents were randomly assigned to either WEB or F2F mode. A small selection of non-responding WEB invitees was followed up with F2F in both Croatia and Germany. In Croatia, this selection was based on the regional availability of interviewers. In Germany it was based on whether the respondents had received an unconditional incentive when invited to participate in the study. As this selection was not random, we excluded WEB invitees that responded only after having been contacted by an interviewer. This allowed us to make a 'pure' comparison between WEB and F2F modes and to avoid bias due to self-selection of participants into the F2F mode. In Germany, response rates were somewhat lower for WEB than for F2F (23.7% versus 29.5%), whereas the reverse was true for Croatia (49.5% versus 27.7%). Due to the design of the experiment, only a small minority of respondents were assigned to the F2F condition in both countries. In Germany, 1492 (88.6%) respondents reported in WEB mode and 192 (11.4%) in F2F mode. In Croatia, 1296 (89.7%) respondents reported in WEB mode and 149 (10.3%) in F2F.

4.2 Measures

For our substantive example, we selected as dependent variables a set of GGS survey items and scales related to subjective well-being. The choice of indicators was driven by substantive arguments (we wanted to include indicators that tap into the important life domains of work, family, social relationship and health) and by the practical availability of subjective well-being indicators in the survey.

First, respondents expressed their evaluation of their satisfaction with their job, answering the question “How satisfied are you with your current job?” (0 “Not at all satisfied”, 10 “Completely satisfied”). Second, respondents were asked to evaluate their satisfaction with their relationship by answering the question “How satisfied are you with your relationship with your partner/spouse?”. Again, the possible answers ranged from 0 “Not at all satisfied” to 10 “Completely satisfied”. Third, we used a loneliness scale based on six items (De Jong Gierveld & van Tilburg, 2006). The items used were: 1 “There are plenty of people that I can lean on in case of trouble”; 2 “I experience a general sense of emptiness”; 3 “I miss having people around”; 4 “There are many people that I can count on completely”; 5 “Often, I feel rejected”; 6 “There are enough people that I feel close to”. Answer categories were “Yes”, “More or less” and “No”. We followed the instructions of the authors in producing scale scores, by counting the number of times that respondents chose the middle or high scores on these six items in terms of being lonely. For instance, if people answered “Yes” or “More or less” to the item “Often, I feel rejected”, this would add a point to their overall loneliness score, ending with a scale ranging from 0 to 6 (0 “Lowest loneliness”, 6 “Highest”). This scale has been shown to be reliable and comparable across contexts and modes (Van Tilburg & de Leeuw, 1991; De Jong Gierveld & van Tilburg, 2010). Our fourth and final measure of well-being, was subjective health, based on the question “How is your health in general?” (range from 1 “Very bad” to 5 “Very good”).

Then, to focus on the potential heterogeneity of the mode effect on subjective well-being across individual characteristics, we explore the moderating role of the three objective determinants we selected: gender (“Female”, “Male”); age classes (“18–25”, “26–35”, “36–45”, “46 or +”);¹ education (“Low”, corresponding to ISCED 0-2; “Medium”, ISCED 3-4; “High”, ISCED 5-8). In order to expand our view, we also check the moderating role of additional socio-demographic variables, namely being married; having children younger than 6 years in the household; total number of children; being a citizen of the surveyed country.

Finally, we built the measure of subjective *WLB* following the International Social Survey Programme (ISSP). The GGS uses the same items and answering scale as in the ISSP (Breyer & Bluemke, 2016), but with answer categories reversed. It reflects the individual’s feeling of not succeeding in combining work tasks and familiar duties. It is built combining the scores (from 1 “Several times a week” to 4 “Never”) the individuals indicated for the following items: 1 “I have come home from work too tired to do the chores that need to be done”; 2 “It has been difficult for me to fulfil my family responsibilities because of the amount of time I spent on my job”; 3 “I have arrived at work too tired to function well because of the household work I had done”; 4 “I have found it difficult to concentrate at

¹ In the models we used age categories instead of a continuous measure of age to account for the potential non-linearity of the association with our subjective well-being indicators.

work because of my family responsibilities". These items show good internal consistency with an ordinal $\alpha = 0.83$ (Zumbo et al., 2007) (see Table A1 in the Online Appendix for the polychoric correlation matrix).

Thus, we calculated factors scores with an iterated principal factors method² (Fabrigar et al., 1999). in order to create a *WLB* index, where a higher score implicates a higher level of *WLB*. Additionally, we examined measurement invariance of *WLB* across modes and countries by running a multi-group confirmatory factor analysis (MGCFAs) where we imposed between-group constraints on factor loadings and used an asymptotically distribution free (ADF) estimator (Bollen, 1989). This check suggests that the association between our four indicators of *WLB* and the latent concept of *WLB* is invariant and does not differ by country or mode (see Table A2 in the Online Appendix for details) (Mellenbergh, 1989).

4.3 Analytical Strategy

Our analytical strategy is divided in two stages. In the first stage, we examined whether the dependent variables of subjective well-being were affected by a mode effect *per se*. We accomplished this by performing a number of statistical analyses. First, we tested whether the means of our variables differed between modes. Next, we tested whether the complete distribution of the dependent variables differed by mode using the Wilcoxon rank-sum test. Third, we used the Blinder-Oaxaca decomposition to examine which part of the difference between modes in the mean scores of the dependent variables could be explained by differences in demographic characteristics of respondents over different modes. This method, which was originally developed to study gender wage discrimination in the labour market (Blinder, 1973; Oaxaca, 1973), has recently been applied in the literature on subjective well-being (Helliwell & Barrington-Leigh, 2010; Sarracino, 2013).

The Blinder-Oaxaca decomposition permits the decomposition of the mode gap in the well-being indicators into two components. The first component, the *explained* part of the gap, comes from the composition of the sample and it is attributable to different observable characteristics of respondents. The second component, the *unexplained* part of the gap, indicates which part is due to differences in the estimated coefficients, namely how people interplay with a specific survey mode.

Formally, the Blinder-Oaxaca decomposition is:

$$\Delta Y = \underbrace{(\bar{X}_a - \bar{X}_b) \cdot \beta'}_{\text{explained}} + \underbrace{\bar{X}_a \cdot (\beta_a - \beta') + \bar{X}_b \cdot (\beta' - \beta_b)}_{\text{unexplained}}$$

ΔY is the gap in the well-being indicators; X_a and X_b are two vectors of explanatory variables of the two groups of respondents (WEB and F2F); β_a and β_b are the coefficients estimated for the two groups of respondents; β' is a vector of non-discriminatory coefficients to evaluate to what extent socio-demographic characteristics of individuals explain the overall difference in the responses by survey mode. This vector comes from the parameter Ω , estimated with a pooled model (Neumark, 1988; Oaxaca & Ransom, 1994).

² We opted for this method to comply with the multivariate non-normality of our measures, as recommended by the literature

We checked the randomization of the two groups controlling for a number of relevant socio-demographic characteristics, included in the vector \bar{X}_a (for F2F respondents) and \bar{X}_b (for WEB respondents): (a) gender; (b) age; (c) level of education; (d) being married; (e) having children younger than 6 years in the household; (f) total number of children; (g) being a citizen of the surveyed country; (h) country of survey. If the differences in responses to the well-being indicators by mode are determined by a “genuine” mode effect and not by a “compositional” effect, we would expect the *explained* part of the gap to be small and the *unexplained* part of the gap to be large.

In the second step, we used OLS regression models to explore whether the subjective well-being determinants differ by survey mode³(Ferrer-i Carbonell & Frijters, 2004). We first included in the model our socio-demographic variables of interest and *WLB* indicator main effects (Model 1), and then we added interaction terms between mode of survey administration and our subjective well-being determinants, to estimate their moderating role on the association between mode and subjective well-being indicators (Model 2). In order to ease the comparability of coefficients, in this step we standardized all the dependent variables and the *WLB_s* indicator. To this aim, we estimated two nested models: the first step (Model 1) is:

$$Y_i = \beta_0 + \beta_1(F2F_i) + \beta_2(DEMO_i) + \beta_3(WLB_i) + \gamma_c + \epsilon_i \quad (1)$$

where Y represents the score for each of the subjective well-being indicators considered; $F2F$ is a dummy variable denoting the survey mode (“WEB”; “F2F”); $DEMO$ is a vector of relevant socio-demographic characteristics (the same as included in the Oaxaca-Blinder decomposition: gender, age, level education, being married, children younger than 6 years in the household, total number of children, being a citizen of the surveyed country); WLB is the measure of work-life balance; γ_c is the country fixed effects.

$$Y_i = \beta_0 + \beta_1(F2F_i) + \beta_2(DEMO_i) + \beta_3(WLB_i) + \beta_4(F2F_i * DEMO_i) + \beta_5(F2F_i * WLB_i) + \gamma_c + \epsilon_i \quad (2)$$

In the second step (Model 2) we added the interaction between the mode of survey administration and both the socio-demographic variables and the *WLB* indicator to estimate to what extent mode moderates the effect of these variables. For the sake of parsimony, we will discuss in the main text only the results for the main predictors of subjective well-being potentially affected by social desirability bias (gender, age, education, *WLB*). Results for the other covariates included in the model are available in Table A3 in the Online Appendix.

³ Since assuming cardinality or ordinality has little impact on estimation results when studying measures of subjective well-being, we present in the main text the results based on OLS regression for ease of interpretation and to be able to compare the coefficients across models by standardizing the dependent variables. However, to account for the ordinal measurement level of our dependent variables, we performed a robustness check by re-estimating all the models with a set of ordered logit models. Results, presented in Tables A4 and A5 of the Online Appendix, are in line with the OLS regression results.

Table 1 Summary statistics for independent variables by survey mode

	WEB	F2F	<i>p</i> value
<i>Male</i>	45.2%	44.5%	0.822
<i>Age: 18–25</i>	21.8%	24.3%	
26–35	29.7%	26.7%	0.621
36–45	30.3%	30.5%	
46 or +	18.2%	18.5%	
<i>Education: Low sec. or less</i>	9.3%	17.5%	
Upp. sec.	39.5%	39.3%	0.000
Tert. or more	51.1%	43.1%	
<i>Married</i>	60.4%	65.7%	0.103
<i>Child <6 years old</i>	27.8%	38.4%	0.000
<i>Num. of children</i>	0.77 (mean)	0.99 (mean)	– 3.48 (<i>t</i> -value)
<i>Citizenship</i>	95.4%	92.3%	0.040
<i>Croatia</i>	46.5%	43.7%	0.329

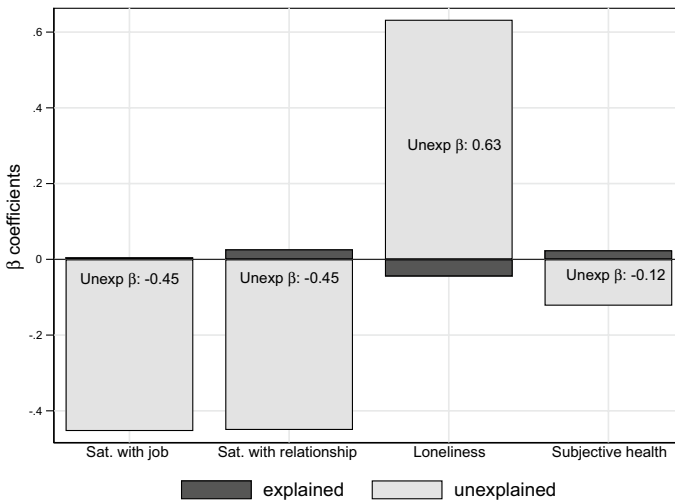


Fig. 1 Blinder-Oaxaca decomposition of subjective well-being indicators gap by survey mode

5 Results

5.1 Mode Differences in Univariate Distributions

Table 1 shows descriptive information on the distribution of our covariates of interest across the two modes. The two samples (WEB and F2F respondents) are similar in terms of some socio-demographic characteristics. At the same time, some differences emerge with respect to (1) level of education, with WEB respondents having slightly higher education than F2F ones; (2) presence of children younger than 6 years in the household and total number of children, with both conditions more often observed

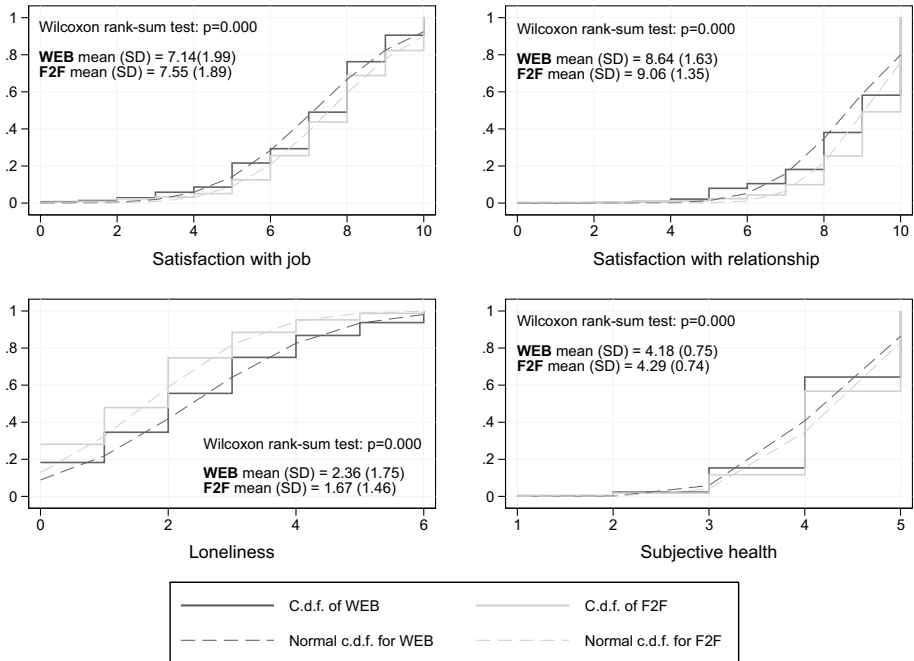


Fig. 2 Cumulative distribution function of subjective well-being indicators by survey mode

among F2F than WEB respondents; (3) being a citizen of the surveyed country, a characteristic somewhat more present among WEB respondents. Thus, the two samples in this study are not fully balanced. This emphasizes the need to include these variables as covariates in our multivariate models.

In this light, to rule out the possibility that our results may be driven not by a “genuine” mode effect but by a “compositional” effect (people with different characteristics may be differently selected into the mode of response, and this may lead us to report a spurious association between subjective well-being indicators and our independent variables), we applied the Blinder-Oaxaca decomposition (see the analytical strategy section for a description). By doing so, we were able to examine whether the differential distribution of respondents across modes influenced the mode effects observed in the dependent variable. The results for our four well-being variables are graphically presented in Fig. 1. It shows that only a very small proportion of the detected mode effect is due to differences in socio-demographics (*explained* part), while the biggest part is *unexplained*. Thus, the Blinder-Oaxaca decomposition demonstrates that it is unlikely that any mode effects we observe are driven by selection bias, and thus increasing the likelihood that they are “genuine” mode effects.

Hence, we move on to examining mode differences in the univariate distributions of our subjective well-being variables. In order to accomplish this analytical goal, we examine the dependent variables’ cumulative distribution functions (*cdf*) by survey mode and test the hypothesis that the two samples (WEB and F2F respondents) are from populations with the same distribution running the Wilcoxon rank-sum test (Wilcoxon, 1945; Mann & Whitney, 1947).

Results of the Wilcoxon rank-sum test, shown in Fig. 2, inform us that all the subjective well-being indicators have a different distribution of scores by survey mode, underscoring the existence of a mode effect in our subjective well-being indicators. In WEB mode, respondents are more likely to give a more negative assessment of their subjective well-being than in F2F mode. The average satisfaction both with the current job and with the relationship are lower in WEB than in F2F (by 4.9% and 5.0%, respectively). The largest difference between WEB and F2F is observed for loneliness: its level is almost 10% higher in WEB than in F2F (0.69 on a scale from 0 to 6). Finally, the mode effect for subjective health is equal to -0.11 points on a scale from 1 to 5 (less than 2% of the total).

5.2 Mode Differences in Multivariate Distributions

Next, we examined whether any evidence of mode effects on subjective well-being could be found using an OLS framework. In Model 1, we examined whether there were mode effects on the mean score of well-being, controlling for a set of socio-demographic variables and *WLB*. In Model 2, we focused on whether the effects of the socio-demographic variables and *WLB* on subjective well-being varied by survey mode by including interaction terms between survey mode and these variables in the model. Regression coefficients of these models are presented in Table 2.

Results of Model 1 confirm the existence of a mode effect on our subjective well-being variables. In Fig. 3, dots represent the point estimates of the linear regression coefficients and the lines represent the width of the 95% confidence intervals. Respondents in F2F mode report higher satisfaction with their job, with their relationship and subjective health, and lower levels of loneliness. Regarding the individual characteristics, a negative gradient between age and satisfaction with relationship and subjective health is evident, with older individuals reporting lower scores on these two outcomes. Job satisfaction shows no significant association with any individual characteristic.

With respect to the additional socio-demographic variables (see Table A3 in the Online Appendix) it turns out that those who are married are more satisfied with their relationship than people who are not married. Furthermore, the married report lower loneliness than the non-married. Respondents who have children under the age of 6 report lower satisfaction with their relationship and slightly higher loneliness levels than respondents without children under the age of 6. Finally, being a citizen of the surveyed country is negatively associated with loneliness.

To analyse if the effect of substantive variables on individual well-being is moderated by survey mode, we then interacted the mode of survey administration with the determinants of subjective well-being (Model 2). To complement the results presented in Table 2, we plotted the average predicted values (AP), so as to illustrate the relationship between our socio-demographic variables of interest and the indicators of subjective well-being over survey mode (Figs. 4, 5, 6 and 7). In these Figures, dots are the average predicted values and vertical lines are the width of the 95% confidence intervals. Hence, black dots show, for each of our three selected socio-demographic characteristics, the predicted values of our standardised subjective well-being indicators in WEB mode; analogously, grey dots show the predicted values in F2F mode.

Results for gender in Table 2 do not reveal any statistically significant interaction with the mode of survey administration for any of the four subjective well-being indicators

Table 2 Association between mode of survey, socio-demographic variables, *WLB* and subjective well-being indicators. Models 1 and 2. OLS regression.

Variables	Sat. with job		Sat. with relationship		Loneliness		Subj. health	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
F2F	0.14*	1.07***	0.23***	0.15	0.34***	-1.64***	0.14**	0.79**
	(0.08)	(0.46)	(0.08)	(0.46)	(0.07)	(0.43)	(0.06)	(0.31)
Male	-0.02	-0.03	0.14***	0.14***	0.09**	0.10*	-0.01	-0.01
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)
<i>Age</i>								
26–35 [ref.: 18–25]	-0.03	0.01	-0.28**	-0.30	-0.07	-0.14	.10	0.10
	(0.11)	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)	(0.10)	(0.09)
36–45	-0.02	0.07	-0.47***	-0.51***	0.16	0.07	-0.17	-0.12
	(0.12)	(0.13)	(0.13)	(0.13)	(0.12)	(0.13)	(0.11)	(0.09)
45 or +	-0.12	-0.04	-0.55***	-0.58**	0.19	0.10	-0.25**	-0.15
	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.12)	(0.10)
<i>Education</i>								
Mid edu [ref.: low edu]	-0.07	-0.01	-0.07	-0.11	-0.11	-0.12	0.04	0.04
	(0.09)	(0.10)	(0.10)	(0.11)	(0.09)	(0.10)	(0.08)	(0.07)
High edu	0.04	0.10	-0.03	-0.06	-0.24***	-0.30***	0.27***	0.22**
	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)	(0.08)	(0.07)
Croatia	0.02	0.02	0.13**	0.12**	0.29***	0.30***	0.25***	0.18***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)
<i>WLB</i>	0.30***	0.30***	0.17***	0.20***	-0.19***	-0.20***	0.27***	0.22***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.07)
F2F*Male	-	0.11	-	-0.02	-	-0.11	-	0.02
		(0.16)		(0.16)		(0.15)		(0.11)
F2F*26–35	-	0.01	-	0.23	-	0.45	-	-0.12
		(0.13)		(0.37)		(0.35)		(0.25)
F2F*36–45	-	0.07	-	0.34	-	0.52	-	-0.01
		(0.13)		(0.39)		(0.37)		(0.26)
F2F*46 or +	-	-0.04	-	0.18	-	0.64*	-	-0.28
		(0.14)		(0.40)		(0.38)		(0.27)
F2F*Mid edu	-	-0.41**	-	0.20	-	-0.05	-	-0.07
		(0.23)		(0.24)		(0.23)		(0.16)
F2F*High edu	-	-0.17	-	0.12	-	0.30*	-	-0.14
		(0.23)		(0.25)		(0.23)		(0.16)
F2F* <i>WLB</i>	-	-0.03	-	-0.11	-	0.10	-	-0.14**
		(0.08)		(0.08)		(0.08)		(0.05)
Observations	1236	1236	1281	1281	1216	1216	1309	1309
R-squared	0.10	0.11	0.08	0.08	0.12	0.14	0.12	0.13

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

considered. This suggests that gender differences in well-being are equally large in WEB and F2F mode.

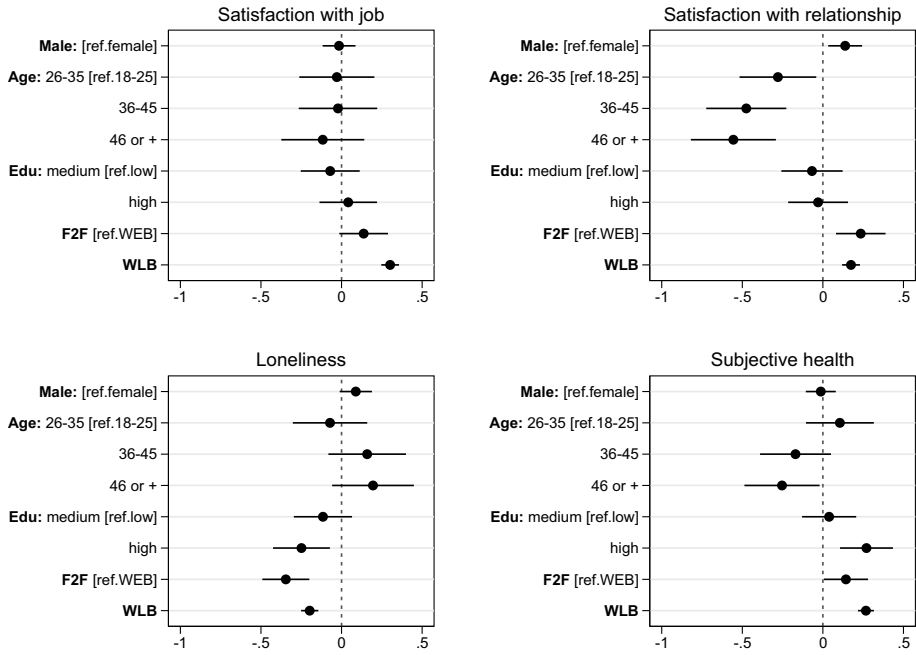


Fig. 3 OLS regression estimates with 95% CI of the effect of socio-demographic characteristics, survey mode and *WLB* on subjective well-being indicators (based on Model 1 in Table 2)

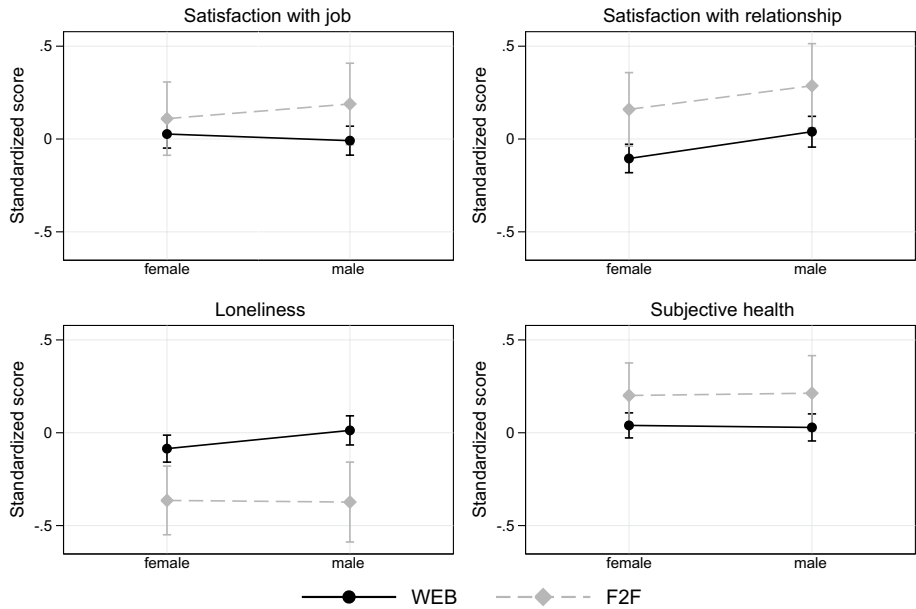


Fig. 4 Average predicted values of subjective well-being indicators for gender by survey mode (based on Model 2 in Table 2)

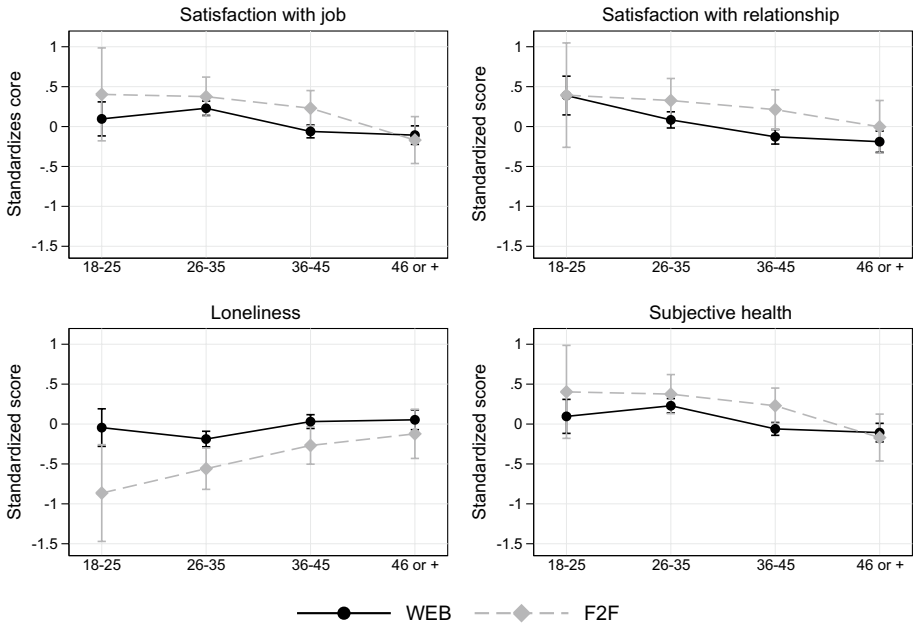


Fig. 5 Average predicted values of subjective well-being indicators for different age groups by survey mode (based on Model 2 in Table 2)

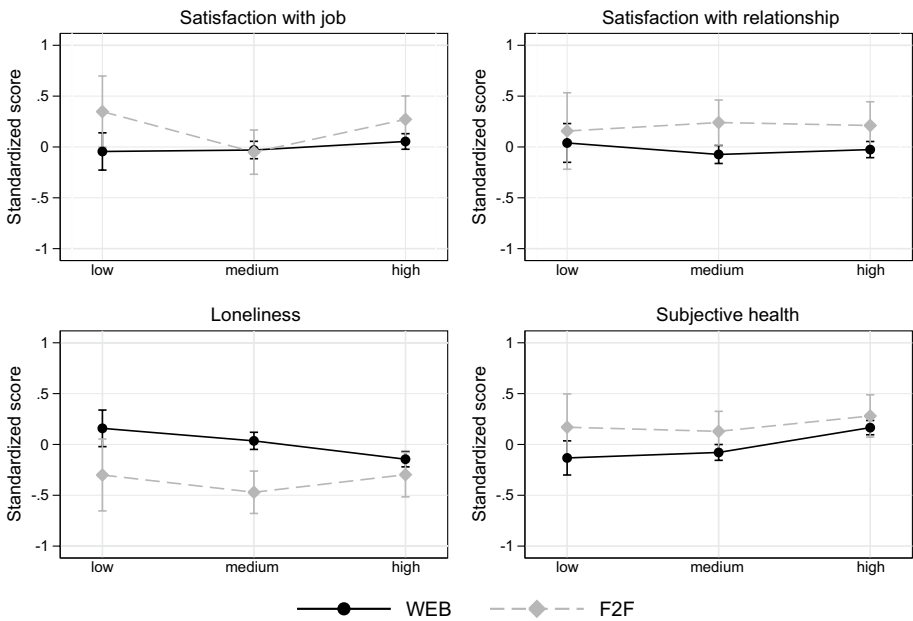


Fig. 6 Average predicted values of subjective well-being indicators for different educational levels by survey mode (based on Model 2 in Table 2)

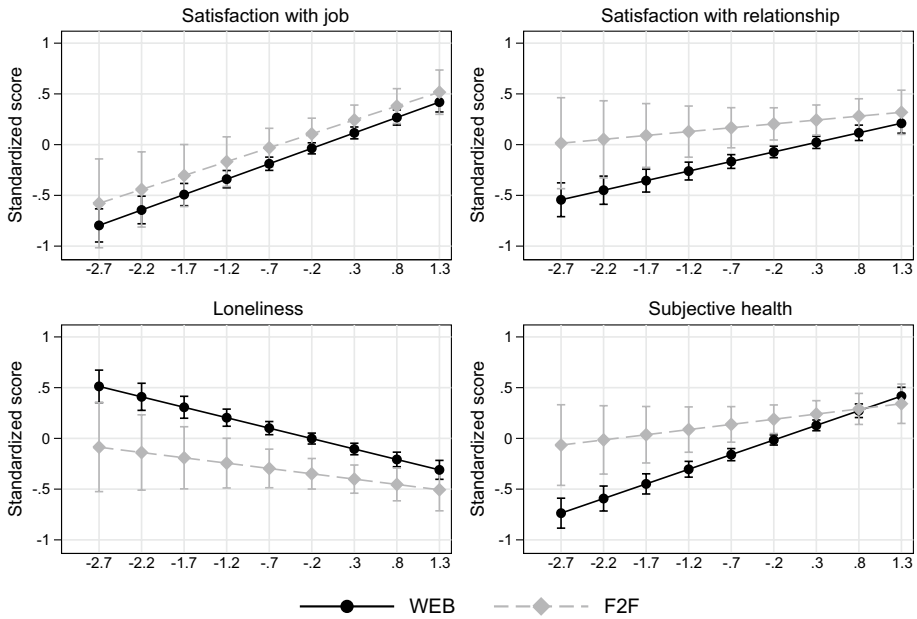


Fig. 7 Average predicted values of subjective well-being indicators for different *WLB* levels by survey mode (based on Model 2 in Table 2)

When considering age, Table 2 shows that one interaction term regarding the senior individuals (aged 46 or more) on loneliness is statistically significant at $\alpha = 10\%$ ($\beta = 0.64$), meaning that the loneliness difference between individuals aged 18–25 and those aged 46 or more varies by survey mode. Indeed, as shown by Figure 5, while in WEB mode individuals aged 18–25 have an AP on loneliness of -0.04 (CI: $-0.28, 0.19$) and those aged 46 or more of 0.05 (CI: $-0.07, 0.18$), the same AP in F2F mode is equal to -0.86 (CI: $-1.47, -0.26$) for people aged 18–25 and to -0.12 (CI: $-0.43, 0.19$) for those aged 46 or more. With respect to the other subjective well-being indicators, Fig. 5 shows that in WEB mode, no age gradient is visible in the loneliness score, whereas in F2F mode, loneliness scores are higher for the oldest age groups. Similarly, while there is no age gradient in WEB mode for subjective health, a negative gradient emerges in F2F, with individuals aged 46 or more reporting significantly lower health (AP = -0.17 ; CI: $-0.46, 0.12$) than those aged 26–35 (AP = 0.38 ; CI: $0.13, 0.62$). As a robustness check, the model was also run with age as either a continuous or a dichotomous variable (see Tables A6 and A7 in the Online Appendix). Results were consistent with those reported in the main text. The results for additional socio-demographic variables in Table A3 in the Online Appendix show that there is also a positive interaction between citizenship and survey mode, implying that citizens only report lower loneliness if they respond to a WEB survey, but not to a F2F survey.

When looking at the relationship between education and mode is negative ($\beta = -0.41$) and statistically significant at $\alpha = 5\%$. The results in Figure 6 suggest that those with medium education report lower job satisfaction (AP = -0.05 ; CI: $-0.27, .17$) than those with low (AP = $.35$; CI: $-0.00, .70$) or high (AP = $.27$; CI: $.04, .50$) education in F2F mode, whereas no educational differences in job satisfaction are visible in WEB mode. Table 2 shows also that for loneliness, a positive ($\beta = .30$) and statistically significant interaction at $\alpha = 10\%$ is

visible between high education and mode. Figure 6 points out that among highly-educated no statistically significant difference in loneliness emerges across survey mode, with an AP of -0.14 (CI: $-0.22, -0.07$) in WEB mode and of -0.30 (CI: $-0.52, -0.08$) in F2F mode. Both the low- and the mid-educated individuals report higher loneliness in WEB than in F2F mode, with low-educated having an AP of $.16$ (CI: $-0.02, .34$) in WEB mode and of -0.30 (CI: $-0.65, .05$) in F2F mode, and mid-educated with an AP of $.04$ (CI: $-0.05, .12$) in WEB mode and of -0.47 (CI: $-0.68, -0.26$) in F2F mode. As a robustness test, the model was re-estimated with education specified as a dichotomous variable and the results, presented in Table A8 in the Online Appendix, were consistent with the main results.

As our last variable of interest, we focused on the relation between *WLB* and each of the four well-being indicators. From Table 2, showing the relationship between *WLB* and our dependent variables over survey mode, it emerges that the only interaction that is significant at $\alpha = 10\%$ is between *WLB* and subjective health ($\beta = -0.14$). Figure 7 shows that for this indicator the *WLB* gradient is less steep for F2F than for WEB mode, implying that the relationship between *WLB* and health is stronger in WEB than in F2F. Figure 7 also shows that respondents who experience high *WLB* report the same level of health irrespective of whether they answered in WEB or F2F mode: for instance, individuals with a value of 1.3 in our *WLB* indicator have an AP of 0.42 (CI: 0.33, 0.50) if they responded in WEB mode, and of 0.34 (CI: 0.15, 0.54) if they responded in F2F mode. Conversely, respondents who experience low *WLB* are more likely to report low subjective health in WEB than in F2F: as an example, people with a value of -2.7 in the *WLB* indicator have an AP of -0.74 (CI: $-0.88, -0.59$) in WEB mode and of -0.07 (CI: $-0.46, 0.33$) in F2F mode. For the three other subjective well-being indicators, no statistically significant interaction is observed in Table 2.

6 Discussion

Different types of survey data are used to test models about subjective well-being indicators and their determinants. It is well-known that survey mode (whether a survey is conducted in face-to-face mode, by telephone, via the web, via mail etc.) strongly affects the average scores on subjective well-being variables. However, little is known about whether survey mode affects the association between subjective well-being indicators and their determinants. This is unfortunate, as web surveys are rapidly gaining in popularity and starting to replace face-to-face interviews as the most common mode of data collection. Therefore, the aim of this study is to examine whether survey mode affects the statistical relationships within subjective well-being models, by comparing the relationship between a set of objective and subjective measures and subjective well-being indicators in WEB and F2F modes. We use representative data from a sophisticated mixed-mode design in a pilot survey of the Generations and Gender Programme conducted in Germany and Croatia.

As expected, we observe large differences in the mean scores on subjective well-being indicators. In WEB mode, respondents report a much lower satisfaction with their job, satisfaction with their partner relationship and subjective health, and higher levels of loneliness. These findings stress that, if one is interested in monitoring trends in subjective well-being, potential changes in modes over time or cross-national differences should be considered. For instance, if one moves from F2F to WEB for cost-related reasons, one should try to get an inkling of the differences in mode effects in subjective well-being

indicators in order to be able to separate 'genuine' change in subjective well-being over time from change resulting from a different survey mode administration.

We observe little evidence of a mode effect in the association between well-being indicators and basic socio-demographic variables. Where differences are shown to be statistically significant, the effects suggest a higher degree of social desirability amongst younger and more educated respondents. Such results suggest a degree of caution when describing socio-economic differentials in well-being through data collected using multiple modes.

The relationship between work-life balance (*WLB*) and subjective well-being indicators is strong. Those respondents who report a better *WLB* are more satisfied with their job and their partner relationship, less lonely and report better health. The good news is that these relationships are generally found irrespective of the mode in which the data are collected. Only in one of four comparisons do we observe a slight difference by mode: the relationship between *WLB* and subjective health is somewhat stronger in WEB than in F2F mode. This leads us to conclude that generally mode does not affect the substantive conclusions we draw about how work-life balance and subjective well-being are interrelated, even if a weak effect seems to emerge when considering self-assessed health. Further research is needed to shed light on this finding.

In general, change from F2F to WEB mode will not lead to a need to rethink our causal or associational models. However, associations generally were slightly larger in the WEB version. A potential reason for this difference could be that the variation in the answering patterns is larger in WEB, as answering patterns are less influenced by social desirability bias than in F2F mode. This might warrant further investigation.

Clearly, this study also has some limitations. First, we only studied a subset of all potential associations that could be analyzed. We opted to focus on a substantively interesting example rather than to embark on a 'fishing expedition' into all potential associations, as science generally develops by focusing on relationships and models of theoretical interest. But it could be, of course, that a focus on other indicators could lead to different results. However, we think that the fact that WEB generally leads to answers that are less susceptible to social desirability bias than F2F, will make the variation in answering patterns generally larger in WEB, and make it easier to observe statistically significant effects.

Second, our sample sizes were not particularly large making it hard to detect mode differences. With larger sample sizes, differences could have turned out to be statistically significant. Finally, we focused on just two country contexts. It could be that country-context matters. Although we did not focus on country differences in mode effects in this study, it could be that mode effects are stronger in countries where people are more susceptible to social desirability bias.

Overall, our conclusion is that mode effects on mean scores of subjective well-being indicators are much larger than on associations within models in a multivariate framework. Thus, these results suggest that our conclusions about substantive models studying subjective well-being are relatively robust against a change in survey mode. Still, further research is warranted to corroborate these findings and their application to specific contexts.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10902-022-00553-y>.

Acknowledgements The GGP Pilot was funded through the Horizon 2020-project 'Generations & Gender Programme: Evaluate, Plan and Initiate (GGP EPI)' (Grant Agreement No 739511). We greatly benefitted from the comments by the GGP Pilot team, Martijn Hendriks and the three anonymous reviewers.

Funding Open access funding provided by Università Commerciale Luigi Bocconi within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly*, 58(2), 210–240.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.
- Breyer, B., & Bluemke, M. (2016). *Work-family conflict scale (ISSP)*. GESIS - Leibniz-Institut für Mannheim.
- Currihan, D. B., Nyman, A. L., Turner, C. F., & Biener, L. (2004). Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? evidence from the umass tobacco study. *Public Opinion Quarterly*, 68(4), 542–564.
- Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539.
- De Jong Gierveld, J., & van Tilburg, T. (2006). A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Research on Aging*, 28(5), 582–598.
- De Jong Gierveld, J., & van Tilburg, T. (2010). The de jong gierveld short scales for emotional and social loneliness: Tested on data from 7 countries in the un generations and gender surveys. *European Journal of Ageing*, 7(2), 121–130.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons
- Dolan, P., & Kavetsos, G. (2016). Happy talk: Mode of administration effects on subjective well-being. *Journal of Happiness Studies*, 17(3), 1273–1291.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.
- Epstein, J. F., Barker, P. R., & Kroutil, L. A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly*, 65(4), 529–549.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
- Ferguson, M., Carlson, D., Zivnuska, S., & Whitten, D. (2012). Support at work and home: The path to satisfaction through balance. *Journal of Vocational Behavior*, 80(2), 299–307.
- Ferrer-i Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497), 641–659.
- Fisher, R. J., & Dube, L. (2005). Gender differences in responses to emotional advertising: A social desirability perspective. *Journal of Consumer Research*, 31(4), 850–858.
- Garbarski, D., Schaeffer, N. C., & Dykema, J. (2019). The effects of features of survey measurement on self-rated health: Response option order and scale orientation. *Applied Research in Quality of Life*, 14(2), 545–560.
- Gauthier, A. H., Cabaço, S. L. F., & Emery, T. (2018). Generations and gender survey study profile. *Longitudinal and Life Course Studies*, 9(4), 456–465.
- Haar, J. M., Russo, M., Suñe, A., & Ollier-Malaterre, A. (2014). Outcomes of work-life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior*, 85(3), 361–373.
- Hagqvist, E., Gädin, K. G., & Nordenmark, M. (2017). Work-family conflict and well-being across europe: The role of gender context. *Social Indicators Research*, 132(2), 785–797.

- Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1), 111–121.
- Heerwig, J. A., & McCabe, B. J. (2009). Education and social desirability bias: The case of a black presidential candidate. *Social Science Quarterly*, 90(3), 674–686.
- Helliwell, J. F., & Barrington-Leigh, C. P. (2010). Measuring and understanding subjective well-being. *Canadian Journal of Economics/Revue canadienne d'économique*, 43(3), 729–753.
- Hendriks, M., Burger, M., Ray, J., & Esipova, N. (2018). Do international migrants increase their happiness and that of their families by migrating. In J. F. Helliwell, R. Layard & J. Sachs (Eds.), *World Happiness Report 2018* (pp. 44–65). UN SDSN.
- Hitchcott, P. K., Penna, M. P., & Fastame, M. C. (2020). Age trends in well-being and depressive symptoms: The role of social desirability. *Psychiatric Quarterly*, 91(2), 463–473.
- Jackle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1), 3–20.
- Koen, B., Loosveldt, G., Vandeplass, C., & Stoop, I. (2018). Response rates in the european social survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*, Retrieved from <https://surveyinsights.org/?p=9673>, 1–12.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047.
- Lugtig, P., Toepoel, V., Emery, T., Cabaço, S. L. F., Bujard, M., Naderi, R., Schumann, A., Lück, D., et al. (2022). Can we successfully move a cross-national survey online? results from a large three-country experiment in the gender and generations programme survey. *SocArXiv. February*, 13.
- Lunau, T., Bambra, C., Eikemo, T. A., van Der Wel, K. A., & Dragano, N. (2014). A balancing act? work-life balance, health and well-being in european welfare states. *The European Journal of Public Health*, 24(3), 422–427.
- Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79–104.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(19), 50–60.
- Martin, P. & Lynn, P. (2011). The effects of mixed mode survey designs on simple and complex analyses. Technical report, ISER Working Paper Series.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127–143.
- Muurlink, O., Peetz, D., & Murray, G. (2014). Work-related influences on marital satisfaction amongst shift-workers and their partners: A large, matched-pairs study. *Community, Work & Family*, 17(3), 288–307.
- Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human resources*, 23(3), 279–295.
- Noda, H. (2020). Work-life balance and life satisfaction in OECD countries: A cross-sectional analysis. *Journal of Happiness Studies*, 21(4), 1325–1348.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693–709.
- Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5–21.
- Pandita, R. (2017). Internet: A change agent an overview of internet penetration & growth across the world. *International Journal of Information Dissemination and Technology*, 7(2), 83–91.
- Pichler, F. (2009). Determinants of work-life balance: Shortcomings in the contemporary measurement of WLB in large-scale surveys. *Social Indicators Research*, 92(3), 449–469.
- Sanchez Tome, R. (2018). *The impact of mode of data collection on measures of subjective wellbeing*. Ph.D. thesis, Université de Lausanne, Faculté des sciences sociales et politiques.
- Sarracino, F. (2013). Determinants of subjective well-being in high and low income countries: Do happiness equations differ across countries. *The Journal of Socio-Economics*, 42(C), 51–66.
- Sarracino, F., Riillo, C. F. A., & Mikucka, M. (2017). Comparability of web and telephone survey modes for the measurement of subjective well-being. *Survey Research Methods*, 11(2), 141–169. <https://doi.org/10.18148/srm/2017.v11i2.6740>.
- Schnettler, B., Miranda-Zapata, E., Grunert, K. G., Lobos, G., Lapo, M., & Hueche, C. (2021). Testing the spillover-crossover model between work-life balance and satisfaction in different domains of life in dual-earner households. *Applied Research in Quality of Life*, 16, 1475–1501.

- Schork, J., Riillo, C. A., & Neumayr, J. (2021). Survey mode effects on objective and subjective questions: Evidence from the labour force survey. *Journal of Official Statistics (JOS)*, *37*(1), 213–237.
- Sirgy, M. J., & Lee, D.-J. (2018). Work-life balance: An integrative review. *Applied Research in Quality of Life*, *13*(1), 229–254.
- Sirgy, M. J., Lee, D.-J., Park, S., Joshanloo, M., & Kim, M. (2020). Work-family spillover and subjective well-being: The moderating role of coping strategies. *Journal of Happiness Studies*, *21*(8), 2909–2929.
- Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of personality*, *79*(4), 741–762.
- Toepoel, V. (2015). *Doing surveys online*. California.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.
- Uglanova, E., & Dettmers, J. (2018). Sustained effects of flexible working time arrangements on subjective well-being. *Journal of Happiness Studies*, *19*(6), 1727–1748.
- Van Tilburg, T., & de Leeuw, E. (1991). Stability of scale quality under various data collection procedures: A mode comparison on the De Jong-Gierveld loneliness scale. *International Journal of Public Opinion Research*, *3*(1), 69–85.
- Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, *34*(4), 425–452.
- Wilcoxon, F. (1945). Some uses of statistics in plant pathology. *Biometrics Bulletin*, *1*(4), 41–45.
- Zhang, X., Kuchinke, L., Woud, M. L., Velten, J., & Margraf, J. (2017). Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior*, *71*, 172–180.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.